Siberian Branch of the Russian Academy of Sciences
Federal Agency for Scientific Organizations
Federal Research Center Institute of Cytology and Genetics

THE TENTH INTERNATIONAL CONFERENCE ON BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE\SYSTEMS BIOLOGY

Abstracts

BGRS\SB-2016
Novosibirsk, Russia
29 August – 2 September, 2016

Novosibirsk
2016

INTERNATIONAL PROGRAM COMMITTEE

Nikolay Kolchanov, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Ralf Hofestädt, University of Bielefeld, Germany

Dmitriy Afonnikov, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Lubomir Aftanas, State Scientific-Research Institute of Physiology & Basic Medicine, Novosibirsk, Russia
Konstantin Anokhin, SIC "Kurchatov Institute", Moscow, Russia
Tatyana Chernigovskaya, St. Petersburg State University, St. Petersburg, Russia
Ivo Grosse, Halle-Wittenberg University, Halle, Germany
Vladimir Ivanisenko, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Sergey Kabanikhin, Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk,
Alexander Kel, geneXplain GmbH i.G., Wolfenbüttel, Germany
Elsa Khusnutdinova, Institute of Biochemistry and Genetics, Russian Academy of Sciences, Ufa, Russia
Alexey Kochetov, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Natalya Kolosova, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Fyodor Kondrashov, Centre for Genomic Regulation, Barcelona, Spain
Inna Lavrik, Otto von Guericke University, Magdeburg, Germany
Andrey Lisitsa, IBMC, Moscow, Russia
Yuri Matushkin, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Mikhail Moshkin, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Alexey Moskalev, Institute of Biology, Komi Science Centre
Evgeny Nikolaev, Institute of Energy Problems of Chemical Physics, Russian Academy of Sciences, Moscow, Russia
Yuriy Orlov, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Ivan Paponov, Norwegian Research Institute for Agriculture and the Environment, Norway
Sergey Peltek, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Aleksey Penin, Moscow State University, Lomonosov Moscow State University, Moscow, Russia
Nikolay Podkolodny, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Vladimir Poroikov, Institute of Biomedical Chemistry, Moscow, Russia
Egor Prokhortchouk, Center of Bioengineering RAS, Moscow, Russia
Valery Puzyrev, Research Institute of Medical Genetics of the Academy of Medical Sciences, Tomsk, Russia
Alexander Ratushny, Institute for Systems Biology and the Institute of Biomedical Research, Seattle, USA
Evgeny Rogaev, University of Massachusetts, USA
Andrey Rzhetsky, University of Chicago, USA
Elena Salina, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Maria Samsonova, St. Petersburg State Polytechnic University, St. Petersburg, Russia
Elena Schwartz, Ariadne Diagnostics LLC, Rockville, USA
Andrey Seluyanov, University of Rochester, USA
Konstantin Skryabin, Center of Bioengineering RAS, Moscow, Russia
Mikhail Voevoda, Institute of Internal Medicine, Novosibirsk, Russia

LOCAL ORGANIZING COMMITTEE

Tatyana Chalkova, ICG SB RAS
Nadezhda Dmitrieva, NSU, ICG SB RAS
Nadezhda Glebova, ICG SB RAS
Tatyana Karamysheva, ICG SB RAS
Andrey Kharkevich, ICG SB RAS
Galina Kiseleva, ICG SB RAS
Evgenia Oshchepkova, ICG SB RAS
Olga Petrovskaya, ICG SB RAS
Erlan Tokpanov, ICG SB RAS
Svetlana Zubova, ICG SB RAS (Chairperson)

CONTACTS

The Federal Research Center
Institute of Cytology and Genetics
The Siberian Branch of the Russian Academy of Sciences

Russian Federation
630090 Novosibirsk, Lavrentyeva, 10

Tel: +7(383) 363-49-80
Fax: +7(383) 333-12-78

URL - ICG SB RAS: www.bionet.nsc.ru
URL - BGRS\SB-2016: http://conf.bionet.nsc.ru/bgrssb2016/
Organizing committee: bgrs2016@icg.sbras.ru

ISBN 978-5-91291-026-5

Organizers

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences

Chair of Information Biology, Novosibirsk State University

Siberian Branch of the Russian Academy of Sciences

Company “Scientific service” Ltd.

The Vavilov Society of Geneticists and Breeders

The program “Genomics, proteomics and bioinformatics” of the SB RAS

Vavilov Journal of Genetics and Breeding

Novosibirsk State University

Sponsors

Russian Foundation for Basic Research

Siberian Branch of the Russian Academy of Sciences

Federal Agency for Scientific Organizations (FASO Russia)

SILVER SPONSORS

Khimexpert Ltd.

DIA-M Ltd.

Eppendorf Russia Ltd.

Roche Diagnostics Rus Ltd.

Kazan Federal Universiti

Agilent Technologies Russia

Bio-Rad

Qvadros-Bio

Albiogen

BRONZE SPONSORS

Genotek Ltd.

ИЦиГ СО РАН

Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук» (ИЦиГ СО РАН)

директор – академик РАН Николай Александрович Колчанов

ИЦиГ СО РАН – это мультидисциплинарный, многопрофильный биологический институт с 60-летней историей, проводящий исследования по ключевым направлениям, связанным с феноменом наследственности (на уровне генов, геномов, клеток, организмов и популяций). ИЦиГ СО РАН по праву считается одним из ведущих биологических институтов в России, обладающим высокопрофессиональным кадровым составом.
В 2015 году образовался Федеральный исследовательский центр путем присоединения к Институту цитологии и генетики СО РАН в форме филиала Сибирского научно-исследовательского института растениеводства и селекции СО РАСХН (СибНИИРС).

Миссия ИЦиГ СО РАН: Решение приоритетных задач развития научно-технологического комплекса РФ в области генетики и селекции растений, генетики и селекции животных, генетики человека и биотехнологии на основе методов молекулярной генетики, клеточной биологии и биоинформатики.

Стратегическая задача: Проведение полных циклов исследований от генерации фундаментальных знаний до прикладных разработок в области генетики и селекции растений, генетики и селекции животных, генетики человека и биотехнологии на основе методов молекулярной генетики, клеточной биологии и биоинформатики, имеющих приоритетное значение для решения задач агропромышленного, биотехнологического, биологического и фармацевтического комплексов России.

Кадровый состав: в составе ИЦиГ СО РАН насчитывается 63 научных подразделения, в которых работает 990 человек, в том числе 417 научных работников, из которых 175 сотрудников в возрасте до 39 лет, 2 советника РАН, 4 академика РАН, 3 члена-корреспондента РАН, 58 докторов наук, 247 кандидатов наук (по данным на конец 2015 года).

Аспирантура: На конец 2015 года в аспирантуре ИЦиГ СО РАН обучалось 69 аспирантов.

Публикации: Институт активно печатается в российских и зарубежных журналах и является в российской биологии одним из признанных лидеров. В 2015 году опубликовано 202 статьи, зарегистрированных в международной базе Web of Knowledge (WoS), общее количество статей в рецензируемых журналах более 370. Статьи Института цитировались в WoS в 2015 г. более 4 000 раз.

Инфраструктура ИЦиГ СО РАН: УНУ «Центр генетических ресурсов лабораторных животных (ЦГР)» и 7 центров коллективного пользования (http://www.bionet.nsc.ru/uslugi/).

Имущественный комплекс: земельный участок площадью 35 тыс. га, закрепленный на праве постоянного пользования; 85 тыс. кв. м рабочих площадей, расположенных на территориях Советского района г. Новосибирска, Барышевского сельского совета НСО, в Искитимском и Черепановском районах НСО, в пос. Краснообск НСО.

Федеральный исследовательский центр ИЦиГ СО РАН приглашает к сотрудничеству научные и коммерческие организации!

Адрес: 630090, Россия, Новосибирск, проспект Академика Лаврентьева, 10, ИЦиГ СО РАН
Тел./факс: +7(383) 363-49-80 / +7(383) 333-12-78
URL: www.bionet.nsc.ru
E-mail: icg-adm@bionet.nsc.ru

CONTENTS

SMALL MOLECULE AGONISTS OF RELAXIN RECEPTOR

A.I. Agoulnik[1]*, I.U. Agoulnik[1], X. Hu[2], C. Myhr[1], Z. Huang[1], B.A. Ho[1], E. Barnaeva[2], J. Xiao[2], M. Ferrer[2], N.T. Southall[2], J.J. Marugan[2]

[1] Herbert Wertheim College of Medicine, Florida International University, Miami, FL, USA; [2]NIH Chemical Genomics Center, National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, USA

** Corresponding author: aagoulni@fiu.edu*

Key words: *relaxin peptide, RXFP1 GPCR, small molecule agonist*

Aim: The anti-fibrotic, vasodilatory, and angiogenic therapeutic properties of relaxin peptide have been shown in several animal models of human diseases and in clinical trials. The aim of the study was to identify stable and bioavailable small molecule agonists of relaxin G protein-coupled receptor, RXFP1.

Methods and Algorithms: High-throughput screening (HTS) of small molecule library, structure-activity relationship (SAR) studies, ligand-receptor interaction modeling, site-specific mutagenesis, transgenic animal studies.

R*esults:* We discovered the first series of small molecule agonists of RXFP1 using HTS followed by SAR optimization. The compounds are selective RXFP1 agonists with low cytotoxicity, excellent *in vitro* ADME, and *in vivo* pharmacokinetic properties. These molecules display efficacy similar to the natural hormone in several functional assays *in vitro* [1]. The agonists did not activate rodent receptors but produced response in cells transfected with human, macaque, pig, and rabbit RXFP1 [2]. Computational modeling in combination with site-directed mutagenesis studies indicated that the small molecules activated relaxin receptor through an allosteric site and did not compete with the relaxin binding to RXFP1 [3]. Analysis showed that the small molecule agonist binds to RXFP1 in a manner similar to the agonist binding complex of β2AR. Unique to RXFP1, however, is the involvement of extracellular loop 3 (ECL3) of 7TM domain whose extensive H-bonding is crucial to the observed selectivity of the compounds for RXFP1 over other related receptors. To test agonists' activity *in vivo* we produced mice with human RXFP1 receptor.

Conclusion: The first potent, selective, and bioavailable small molecule agonists for relaxin receptor RXFP1 were identified. Mutagenic and modeling studies indicate that the agonists use allosteric binding to the transmembrane region of the receptor. Understanding the structural basis and molecular mechanism of activation and selectivity of the RXFP1 agonists provides useful information for future optimization of the lead series.

Acknowledgements: Supported by NIH (R03MH085705, 1U01CA177711), Florida Department of Health (3KFO1) grants, and NIH Roadmap for Medical Research.

References:

1 Xiao et al. (2013) Identification and optimization of small-molecule agonists of the human relaxin hormone receptor RXFP1. *Nature Comm*, 4:1953.

2. Huang et al. (2015). Activation of Relaxin Family Receptor 1 from different mammalian species by relaxin peptide and small molecule agonist ML290. *Front. Endocrinol*. 2015, doi: 10.3389/fendo.2015.00128.

3. Hu et al. Structural insights into the activation of human relaxin family peptide receptor 1 by small molecule agonists. *Biochemistry*, 55:1772-83.

THE ROLE OF FUNCTIONAL DOMAINS OF DROSOPHILA SEPTIN PNUT

K.A. Akhmetova[1, 2, 3]*, N.V. Dorogova[1], M.L. Balasov[3], S.A. Fedorova[1, 2], I.N. Chesnokov[3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *University of Alabama at Birmingham, Birmingham, USA*
* *Corresponding author: katarina@bionet.nsc.ru*

Key words: *septin, Drosophila, peanut*

Motivation and Aim: Septins are highly conserved GTP-binding filament-forming proteins found in majority of eukaryotic organisms. They localize primarily to the cell membranes and participate in many cellular processes including cytokinesis, cell movement and polarity, secretion and cytoskeletal dynamic [1]. The functions of conserved septin domains are not well understood. The goal of this study was to analyze the role of C-terminal and GTP-binding domains of *Drosophila* septin Peanut (Pnut) in somatic cells as well as gonadogenesis and gametogenesis.
Methods and Algorithms: C-terminal and GTP-binding domain mutants were created using site-directed mutagenesis and standard molecular biology techniques. The ability of Pnut mutants to form septin complexes was assessed using baculovirus expression system. Filament formation was observed using negative stain electron microscopy. Constructs carrying wild type and mutant *pnut* transgenes were injected into w^{1118} *Drosophila* embryos. Obtained transgenes were analyzed on *pnut*-null background using $pnut^{XP}$ deletion (Flybase ID: FBal0035461). Cytological analysis was performed as described earlier [2].
Results: Deletion of Pnut C-terminal domain prevents the formation of septin complexes and filaments *in vitro*. *In vivo*, in *Drosophila* tissues, truncated Pnut protein forms aggregates. Most of mutants die as third-instar larvae; however, they develop imaginal disks and do not show polyploidy in neural ganglia. Only 2-3% of mutants survive to imago stage. Eggs, produced by such females are shorter compare to wild type and have abnormal dorsal appendages. Mutant males are sterile, their testes are underdeveloped, often have round shape and do not contact with seminal vesicles. Mutations in GTP-binding domain affect the formation of septin filaments *in vitro*. *In vivo*, mutant Pnut protein often fails to localize at cell membranes and is found in cytoplasm. Most severe mutation, Pnut(G1,G3,G4), results in third-instar lethality, with no imaginal disks and polyploid cells in larval neural ganglia. About 2% of mutants survive to imago stage. Such females produce egg chambers with abnormally low number of nurse cells. In spermatogenesis, mutants show cyst polarization defects.
Conclusion: We showed that both C-terminal and GTP-binding domain are important for *Drosophila* survival. Both domains are necessary for the formation of complexes and filaments – biologically active septin structures. Phenotypes observed in mutants suggest the important role of C-terminal and GTP-ase domains of Pnut in the processes of gonadogenesis and suggest the multifunctionality of this septin.
References:

1. S. Mostowy, P. Cossart. (2012) Septins: the fourth component of the cytoskeleton Nat. Rev. Mol. Cell Biol.,13:183-194.
2. N. Dorogova et al. (2008) The role of Drosophila Merlin in spermatogenesis, BMC Cell Biol.,9:1-15.

A FUNCTIONAL ANALYSIS OF SEPTIN PROTEINS IN *DROSOPHILA MELANOGASTER* S2 CELLS

A.L. Alekseeva[1, 2]*, E.N. Andreyeva[1], L.A. Yarinich[1], A.V. Pindyurin[1, 3], S.A. Fedorova[3]
[1] *Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russia*
[2] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
[3] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
* *Corresponding author: aalekseeva@mcb.nsc.ru*

Key words: *septins, cytokinesis, GTPase, cytoskeleton, S2, Drosophila, RNAi*

Motivation and Aim: Septins are conserved filament-forming GTP-binding proteins found in all eukaryotic organisms except plants, but their functions are not fully understood. *Drosophila melanogaster* has only 5 septins (Sep1, Sep2, Pnut, Sep4 and Sep5) making it a convenient model organism to study septin functions. The Pnut, Sep1 and Sep2 proteins are known to form a heteromeric complex. Such complexes interact with each other forming filaments, which particularly participate in the formation of the cleavage furrow. The function of Sep4 and Sep5 remains largely unknown. Here, we studied the role of all 5 septins in cultured *Drosophila* S2 cells by using RNAi.
Methods and Algorithms: D. melanogaster S2 cell line was grown in Schneider's medium (Sigma S0146) supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco, 10270-106). At the end of RNAi treatments (5 days), cells were collected, fixed in 3,7% formaldehyde and stained first with mouse anti-α-tubulin (Sigma T5168) and rabbit anti-DSpd2 [1] primary antibodies and then with anti-mouse-FITC (Sigma F8264) and anti-rabbit-Alexa568 (Invitrogen A11036) secondary antibodies. Efficiency of RNAi was checked by Western-blot analysis and RT-qPCR. Mitotic index was calculated as a percentage ratio of dividing to total number of cells.
Results: RNAi knockdown of either *pnut*, *Sep1* or *Sep2* genes decreased amounts of proteins encoded by all the three genes. This indicates that none of these proteins can be substituted in the six-subunit heteromeric complex, which leads to its breakdown and degradation of the components. We also observed that depletion either Sep1 or Pnut leads to significantly decreased amount of transcripts of both *Sep1* and *pnut* genes suggesting the existence of a mechanism of their interdependent regulation. The amount of *Sep2* gene transcripts was significantly affected (decreased) only after RNAi against *Sep2* and *Sep5* genes. While the former is trivial, the latter suggests another mechanism of cross-regulation between septin genes. On the other hand, RNAi against *Sep5* gene affected (besides *Sep5* gene itself) only transcript level of *Sep2*, but not the other septin genes. Importantly, *Sep5* is a retrogene copy of *Sep2* and we observed similar mitotic abnormalities in both Sep2- and Sep5-depleted cells suggesting the interaction between Sep2 and Sep5 at the protein or/and genetic levels. Surprisingly, we found that amount of *Sep4* gene transcripts is substantially increased upon depletion of Sep2 or Pnut. Despite the fact that depletion of some septins resulted in cytokinesis defects, we found no effect on mitotic index in *Drosophila* S2 cells.
References:
1. M.G. Giansanti et al. (2008) Drosophila SPD-2 is an essential centriole component required for PCM recruitment and astral-microtubule nucleation. Curr. Biol. 18:303-309.

APPROACH TO PREDICTING THE SOLUBILITY/ INSOLUBILITY OF *E. COLI* PROTEINS BASED ON THEIR PRIMARY STRUCTURE USING SEQUENCE NORMALIZATION AND MACHINE LEARNING TECHNIQUES

N.A. Alemasov[1]*, N.V. Ivanisenko[1], K.S. Antonets[2,3], A.A. Nizhnikov[2,3], V.A. Ivanisenko[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *St. Petersburg State University, St. Petersburg, Russia*
[3] *Vavilov Institute of General Genetics SPB RAS, St. Petersburg, Russia*
* *Corresponding author: alemasov@bionet.nsc.ru*

Key words: *E. coli*, *protein solubility, aggregation, amino acid sequence, machine learning*

Motivation and Aim: Many human diseases arise from aggregation of different proteins involved [1]. All *Escherichia coli* proteins are found to fall into two distinct groups: soluble and aggregation-prone [2]. Recently proteomic analysis of *E. coli* was carried out discovering several dozens of proteins in fractions resistant to solubilization by ionic detergents [3]. Latter analysis showed correlation between experimentally demonstrated detergent-resistance of proteins and their predicted amyloidogenicity. Thus, a computational approach was required to learn from these experiments and to allow further computer-guided analysis of proteins' aggregation propensity.
Methods and Algorithms: A range of machine learning methods was applied to construct the solubility classifiers and regression models. New approach was used to normalize sequences to uniform length based on [4, 5]. Solubilities of more than 2000 *E. coli* proteins were taken from the experiments [2, 3] and were used to build training/test sets.
Results: Solubility estimations of *E. coli* proteins were made. Additionally proteins from the test set were classified as the soluble/insoluble. R^2 of the best performing random forest regression was about 0.86. AUC of the best performing random forest classifier was 0.81.
Conclusion: Regression models and classifiers constructed allow predicting solubility of a protein by its sequence. The approach is about to be compared with present rivals.
Availability: Software is freely available as a Python-script.
Acknowledgements: The approach development was supported by RSF grant 14-24-00123, and preparing training/test sets – by budget project 0324-2015-0003.
References:
1. F. Chiti, C. M. Dobson. (2006) Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, **75**: 333-366.
2. T. Niwa et al. (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *PNAS*, **106**(11): 4201-4206.
3. K. S. Antonets et al. (2016). Proteomic Analysis of Escherichia coli Protein Fractions Resistant to Solubilization by Ionic Detergents. *Biochemistry. Biokhimiia*, **81**(1): 34-46.
4. D. Heider, J. Verheyen, D. Hoffmann. (2011) Machine learning on normalized protein sequences. *BMC Research Notes*, **4**(1): 94.
5. M. Z. Tien et al. (2013) Maximum allowed solvent accessibilites of residues in proteins. *PloS One*, **8**(11): e80635.

THE FUNCTIONAL INTERACTIONS OF PLEIOTROPIC PROTEIN YB-1 WITH KEY BASE EXCISION REPAIR FACTORS

E.E. Alemasova[1], N.A. Moor[1], K.N. Naumenko[1, 2], P.E. Pestryakov[1], O.I. Lavrik[1, 2]*
[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: lavrik@niboch.nsc.ru*

Key words: *Y-box binding protein 1 (YB-1), base excision repair (BER), PARP1(2), poly(ADP-ribose) (PAR), APE1, NEIL1, pol β*

Motivation and aim: Base excision repair (BER) is a flagship DNA repair system responsible for maintaining genome integrity. Alongside with basal enzymes, this system involves several accessory proteins essential for coordination and regulation of DNA processing during consecutive repair steps. Y-box-binding protein 1 (YB-1) is a multifunctional factor that can interact with DNA, RNA, poly(ADP-ribose) and plenty of proteins including DNA repair enzymes. Its distinctive feature is accumulation in the nucleus upon genotoxic stress conditions. The aim of present research was to investigate YB-1 potential to participate in BER pathway as an accessory regulatory protein.
Methods: Fluorescence titration method, gel-mobility shift analysis, gel electrophoresis, Western blot analysis.
Results: We detected and characterized quantitatively YB-1 physical interactions with key BER proteins: apurinic/apyrimidinic endonuclease 1 (APE1), DNA glycosylase NEIL1, DNA polymerase β (pol β), poly(ADP-ribose) polymerases 1 and 2 (PARP1, PARP2) and DNA-binding fragment of PARP1 – p24. Functional coupling of YB-1 and these DNA repair enzymes was also established. YB-1 was shown to modulate AP endonuclease activity of APE1, AP lyase activity of NEIL1 and dRP lyase activity of pol β. Interestingly, we found that YB-1 can significantly contribute to poly(ADP-ribosyl)ation signaling – the main regulatory system of BER – by covalent and non-covalent interactions with PAR. We demonstrated for the first time poly(ADP-ribosyl)ation by PARP1 and PARP2 as a new posttranslational modification of YB-1. It was shown that covalent attachment of PAR polymer to YB-1 resulted in dramatically decrease in YB-1 affinity for DNA. Non-covalent binding of YB-1 to PAR was proposed to underlie strong stimulation of PARP1 autopoly(ADP-ribosyl)ation and inhibition of poly(ADP-ribose) degradation by poly(ADP-ribose) glycohydrolase (PARG) observed in the presence of YB-1.
Conclusion: The results obtained not only reveal YB-1 potential to facilitate BER, but also offer a challenge for future research as discovered involvement of YB-1 into poly(ADP-ribosyl)ation can contribute to multiple facets of cellular response to genotoxic stress.
Acknowledgements: This work was supported by fellowship from President of RF and by grant from RSF (14-24-00038).

THE P53 FAMILY IN CANCER BIOLOGY

I. Amelio[1], F. Bernassola[2], T.W. Mak[2], and G. Melino[1, 3]
[1] *MRC Toxicology Unit, LeicesterLE1 9HN, United Kingdom*
[2] *The Campbell Family Cancer Research Institute, Toronto, Ontario M5G 2M9, Canada*
[3] *University of Rome Tor Vergata, Rome, Italy*

Key words: *Cancer, transcription, tumor suppression*

p73 and p63 are a members of the p53 family, transcribed as two distinct isoforms TA-isoforms and DN-isoforms, containing or not the N-terminal transactivation domain. Both p63 and p73 are involved in female infertility maternal reproduction (Nature Rev Mol Cell Biol 2011;12,4:259-65) and as well as in cancer formation (TiBS 2014;39(4):191-8). We identified their activation during DNA damage, several transcriptional targets, the mechanisms of regulation of cell death, and the protein degradation pathway.
TAp73 knockout mice show high tumor incidence with hippocampal dysegensis. Conversely, ΔNp73 knockout mice show a very low incidence of cancer, with sign of moderate neurodegeneration with a significant loss of cellularity in the cortex. This indicate a tumor suppressor role for TAp73 and an oncogenic role for ΔNp73. Here, we demonstrate that the transcription factor TAp73 opposes HIF-1 activity through a non-transcriptional mechanism, thus affecting tumour angiogenesis. TAp73-deficient mice have an increased incidence of spontaneous and chemically induced tumours that also display enhanced vascularisation. Mechanistically, TAp73 interacts with HIF-1a, promoting HIF-1a polyubiquitination and consequent proteasomal degradation. In human lung cancer, TAp73 strongly predicts good patient prognosis, and its expression is associated with low HIF-1 activation and angiogenesis. These findings demonstrate a novel mechanism for HIF-1 regulation and provide an additional explanation for the molecular basis of the growth, progression, and invasiveness of human cancers. (PNAS-USA 2015. 112,1:226-31. PMID: 25535359) (TiBS 2015. 40,8:425-34. PMID: 26032560)
P63 is a determinant of skin development. Using a MMTV-ErbB2 murine model, we found that ΔNp63 regulates mammary Cancer Stem Cells self-renewal and breast tumorigenesis via the direct transactivation of Sonic Hedgehog (Shh), GLI family zinc finger 2 (Gli2), and Patched1 (Ptch1) genes. (PNAS-USA 2015. 112,11: 3499-504. PMID: 25739959). At least in part, this seems to be exerted by regulation of the metabolism via Hexokinase II (PNAS-USA 2015. 112,37: 11577-82. PMID: 26324887).

CHANGE OF THE SCENARIO OF THE TRP-CAGE MINIPROTEIN FOLDING WITH TEMPERATURE

V.A. Andryushchenko*, S.F. Chekmarev
Institute of Themophysics SB RAS and Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: vladimir.andryushchenko@gmail.com*

Key words: *protein folding, hydrodynamic approach, folding scenario, folding pathways, melting temperature*

Motivation and Aim: The Trp-cage miniprotein is a very popular system to study protein folding because it folds very fast and contains secondary structure elements typical of globular proteins. Most studies agree that there are two characteristic folding pathways. In one pathway (I), the hydrophobic collapse precedes the formation of the alpha-helix, and in the other pathway (II), the events occur in the reverse order. However, there is no agreement about the efficiency of these pathways. To get a closer insight into the folding mechanisms of Trp-cage, we perform a systematic study of Trp-cage folding at different temperatures using hydrodynamic approach [1].
Methods and Algorithms: The simulation of folding trajectories was performed with the CHARMM program [2]. The Principal Component Analysis method was used to transform the multi-dimensional conformation space of the protein to two-dimensional space of collective variables. Representative points of the protein states were clustered using the MCLUST method [3]. Following the hydrodynamic approach [1], the probability fluxes of probability transitions were calculated and the streamlines of the folding flow were determined, which allowed us to separate different folding pathways.
Results: It has been found that as the temperature increases, the pathway I gradually transforms into pathway II. At T = 285K, approximately 90% of the total flow follow pathway I. At T = 315K, the fraction of the flow through pathway I decreases to 50%. Finally, at T = 325K, the pathway II was found to be dominant (90%).
Conclusion: It would be tempting to connect folding dynamics to thermodynamics, in particular, to relate "pathway switch" temperature (T=315K) to melting temperature, especially as it coincides with the experimental melting temperature. However, the calculated melting temperature was found much higher than the experimental value, similar to some previous works. The calculated heat capacity curve also showed that the melting is gradual, with no pronounced premelting effects. This suggests that the Trp-cage folding mechanism is determined by kinetic factors rather than thermodynamics.
This work was performed under a grant from the Russian Science Foundation (No. 14-14-00325).

References:

1. S.F. Chekmarev, A.Yu. Palyanov, M. Karplus. (2008) Hydrodynamic Description of Protein Folding, *Phys. Rev. Lett.,* **100:** 018107 (1-4).
2. B.R. Brooks, et al. (2009) CHARMM: The biomolecular simulation program, *J. Comput. Chem.* **30:** 1545-1614.
3. C. Fraley, A.E. Raftery. (2002) Model-based clustering, discriminant analysis, and density estimation, *J. Am. Stat. Assoc.* **97:** 611-631.

AGING AND CANCER: STATE-OF-ART AND PROSPECTS FOR PREVENTION

V.N. Anisimov
Department of Carcinogenesis and Oncogerontology, N.N. Petrov Research Institute of Oncology
** Corresponding author: aging@mail.ru*

Key words: *Aging, cancer, international expertize*

The incidence of cancer increases with age in humans and in laboratory animals. A clear understanding of the causes of the age-related increase in cancer incidence is needed to develop a strategy for primary cancer prevention. Carcinogenesis is a multistage process: neoplastic transformation implies the engagement of a cell through sequential stages, and different agents may affect the transition between stages. Multistage carcinogenesis is accompanied by disturbances in tissue homeostasis and perturbations in nervous, hormonal, immune and metabolic systems which may affect antitumor resistance. The development of these changes depends on the susceptibility of various systems to a carcinogen and on the dose of the carcinogen. Changes in the microenvironment may modify key carcinogenic events and determine the duration of each carcinogenic stage, and sometimes they may even reverse the process of carcinogenesis. These microenvironmental changes influence the proliferation rate of transformed cells together, the total duration of carcinogenesis and, consequently, the latent period of tumor development. Aging may increase or decrease the susceptibility of various tissues to initiation of carcinogenesis and usually facilitates stages promotion and progression of carcinogenesis. Aging may predispose to cancer by two mechanisms: tissue accumulation of cells in late stages of carcinogenesis and alterations in internal homeostasis, in particular, alterations in immune and endocrine system. Increased susceptibility to the effects of tumor promoters is found in both aged animals and aged humans, as predicted by the multistage model of carcinogenesis. Aging is associated with number of events at molecular, cellular and physiological levels that influence carcinogenesis and subsequent cancer growth. There are a huge amount of new facts and concepts in the field. However today we are not more close to understanding real relationships between aging and carcinogenesis. A significant increase of the elderly in populations of developed countries is followed by increase morbidity and mortality from main age-related diseases – cardiovascular and neuro-degenerative, cancer, diabetes mellitus, declining in a resistance to infections. Obviously the development of means of the prevention of the premature ageing and these diseases in humans is the crucial at present. However data on such kind means rather scarce, contradictory and often are not reliable from the points of view of the adequacy of the experiments to current scientific requirements as well as the interpretation of the results and safety. Available data on the life span extension and adverse effects of chemical compounds and drugs suggested as geroprotectors are critically analysed, mainly focused on antidiabetic biguanises and melatonin. Most of the results could not convincingly evidence the life span extension and safety of the suggested geroprotectors. We believe that it is necessary to establish an international program for the expert evaluation of the life span extension potential of pharmacological interventions for humans. The scope of the program should be to evaluate chemical, immunological, dietary and behavioural interventions that may lead to life span extension, and the objective – preparation of critical reviews and evaluations on evidence of the life span extending properties of a wide range of potential geroprotectors and strategies by international groups of working experts.

VLINCRNA DATABASE: TOOL FOR VERY LONG INTERGENIC NON-CODING RNA FUNCTIONAL ANNOTATION

D. Antonets[1, 2, 4]*, Y. Vyatkin[2, 3], D. Luppov[2, 3], P. Kapranov[3, 5], M. Ri[2, 3], O. Saik[2, 3, 6], D. Shtokalo[1, 2, 3]

[1] *A.P. Ershov Institute of informatics systems, Novosibirsk, Russia*
[2] *AcademGene LLC, Novosibirsk, Russia*
[3] *St. Laurent Institute, 317 New Boston St., Woburn, MA 01801, USA*
[4] *State Research Center of Virology and Biotechnology 'Vector', Novosibirsk, Russia*
[5] *Institute of Genomics, School of Biomedical Sciences, Huaqiao University, 668 Jimei Road, Xiamen, China*
[6] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: antonec@yandex.ru*

Key words: *very long intergenic non-coding RNA, systems biology, function annotation, gene ontology, database, web-service*

Motivation and Aim: Vast amount of lncRNA species have been discovered recent years in mammalian genomes and stored on the web. However there is an impressive misbalance between portion of at least 10% of human genome occupied by non-coding RNA and relatively few facts known about its function. Our results suggest that vlincRNAs represent a hitherto hidden layer of regulation involved in critical biological processes and diseases. Many of them are highly expressed in cancers, and some are expressed in stem cells and appear to be regulated by transcription factors involved in stem cell differentiation [1]. Here we present our vlincDB – a web-tool for functional annotation and analysis of human very long intergenic non-coding RNAs.
Methods and Algorithms: vlincDB web-site was created with PHP (v.5) and bootstrap framework (v.3.3.6), all annotation tables are stored in MySQL database (v.14.14).
Results and Conclusion: VlincRNA database, provides both the list of 5151 putative very long intergenic non-coding RNA transcripts and the list of separate 1542 vlinc RNA genes. The integrated analysis of genomic features leveraged by transcription level data measured in 833 tissues and cell lines by FANTOM5 consortium allowed us to predict possible functions of vlincRNA genes [1]. The database provides annotation and a tool of search by gene ontology categories, SNP traits, chromatin modification states, ChIP-seq signals in vlincRNA genes promoter regions, by overlapping known lncRNA, nearby genes and other. Advanced tools implement scenario of overlapping vlincRNA genes with user-defined genomic intervals (provided in BED format), GO terms enrichment analysis and classification into cancer/normal/stem-cell categories according to FANTOM5 gene expression data. All annotation data can be downloaded.
Availability: http://office.nprog.ru:8081/table.php

References:

1. G. St. Laurent et al. (2016) Functional annotation of the vlinc class of non-coding RNAs using systems//*Nucl. Acid Res., doi: 10.1093/nar/gkw162.*

POXVIRAL CHEMOKINE-BINDING PROTEINS: THEORETICAL STUDY OF STRUCTURE AND FUNCTION EVOLUTION

D.V. Antonets*[1], K.V. Gunbin[2], T.S. Nepomnyashchikh[1]
[1] *State Research Center of Virology and Biotechnology "Vector", Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: antonec@yandex.ru*

Key words: *poxviruses, immunomodulatory proteins, chemokines, GIF – GM-CSF/IL-2-binding protein, glycosaminoglycans, molecular modeling, phylogenetic analysis*

Motivation and Aim: Poxviruses are large enveloped dsDNA viruses with complex genomes containing about 200 genes, a half of which codes for immunomodulatory proteins subverting antiviral responses of the host. One of the most interesting groups of poxviral immunomodulatory proteins are chemokine-binding proteins. Despite sharing low sequence identity the members of this protein family posses remarkably similar tertiary structures. With the help of phylogenetic analysis and molecular modeling here we tried to get some insights about the evolution of the proteins of this family and about changes of their molecular functions.
Methods and Algorithms: Multiple protein alignment was made by PROMALS3D. Secondary structures were predicted by SCRATCH-1D. We generated 300 sub-alignments using 50% alignment jackknife. Then for each alignment the phylogenetic tree was reconstructed by RaxML 8 and GTR model. Phylogenetic analysis for each sub-alignment was based on 3 data sources: (1) amino acid sequences and protein secondary structures using (2) 3 or (3) 8 structure types. Consensus phylogenetic trees were reconstructed by Dendroscope 3.4.0. The reconstruction of the last common ancestors (LCA) of 12 clusters of proteins (corresponding to statistically significant phylogenetic clades) were made using RaxML 8 and GTR model. Protein 3D structures were reconstructed by I-TASSER and RaptorX web-services. Structure refinement was done using FG-MD, ModRefiner and GalaxyRefine. Electrostatic surface potentials were calculated with DelPhi software.
Results and Conclusion: For each protein cluster at least one 3D protein structure was modelled. Using these modelled structures, we distinguished two structural types of investigated proteins: (A) composed of one chemokine-binding domain (1CQ3/2VGA/4P5I) only, and (B) composed of two domains: chemokine-binding SECRET domain and TNFR2-like domain (CrmB and CrmD proteins – 3ON9). These two protein types form two statistically significant subtrees composed of 7 and 5 clades, respectively. Similarity analysis of GTR matrices of amino acid substitution rates for 12 clusters as well as the analysis of secondary structure evolution shown that subtree A contains LCA of poxviral chemokine-binding proteins. There are two clusters located closely to LCA: the cluster containing secreted chemokine-binding proteins from Vaccinia and Cowpox viruses, and the cluster of such proteins from Orf virus. The more ancestral subtree of GM-CSF/IL2-binding (GIF) proteins was predicted not only to share high structural similarity with A41 but also to bear prominent positive electrostatic charge at surface formed with second β-sheet as it was shown for A41. Thus GIF proteins might bind glycosaminoglycans interfering with chemokines binding to GAGs and chemotactic gradient formation.
Acknowledgements: This work was supported with RFBR grant #15-04-08956-a.

TRANSCRIPTOME WIDE PREDICTION OF LNCRNA-RNA INTERACTIONS BY A THERMODYNAMICS ALGORITHM

I.V. Antonov*, M.A. Zamkova, A.V. Marakhonov, M.Y. Skoblov, Y.A. Medvedeva
Research Center of Biotechnology RAS, Moscow, Russia
** Corresponding author: ivan.antonov@gatech.edu*

Key words: *antisense interaction, long non-coding RNA, post-transcriptional regulation*

Motivation and Aim: Long noncoding RNAs (lncRNAs) are a large and diverse class of transcribed RNA molecules with a length of more than 200 nucleotides that do not encode proteins. The discovery of thousands of lncRNAs in mammals raised a question about their functionality. Due to functional diversity the role and/or molecular mechanism of only few hundred lncRNAs have been determined by the date. Particularly, it has been shown that some of them function post-transcriptionally via formation of intermolecular RNA-RNA duplexes. The primary aim of this study is to bioinformatically address novel lncRNA functions by predicting RNA-RNA interactions transcriptome-wide.
Methods and Algorithms: To search for potential antisense partners for a given noncoding RNA, existing large-scale studies utilized sequence alignment tools (such as BLASTn) without taking into account RNA secondary structure and interaction energy, crucial for RNA-binding. To compensate for this disadvantage co-folding of two RNAs (the query lncRNA versus each of the RNAs in the transcriptome) into minimal free energy structure using thermodynamics-based methods (e.g. *bifold*) can be used. Unfortunately, this task is not computationally feasible on the transcriptome-wide level.
In this work we developed a new pipeline, called ASSA (''AntiSense Search Approach''), which reduces running time by fast identification of putative antisense sites by a sequence alignment tool BLASTn followed by verification of each potential interaction by *bifold*. In our pipeline we automated selection of the initial set of putative antisense sites (i.e. optimized thresholds for BLASTn search), estimated statistical significance (E-value) of antisense interaction energy and the length of the flanking sequences to putative site for validation by *bifold*.
Results: ASSA was capable of predicting 26 out of the 29 known functional RNA-RNA interactions (both *cis* and *trans*) in human and mouse transcriptomes. Comparison of ASSA with other tools showed that it produces one of the strongest predictions in terms of Sensitivity, Accuracy and AUC. We have also applied ASSA to publicly available data from knockdown experiments of 49 murine lncRNAs. We identified four lncRNAs with statistically significant overlap between the ASSA predictions and the differentially expressed genes observed in the experiment, suggesting possible molecular mechanism for these long noncoding RNAs.
Conclusion: We have developed a new computational approach for transcriptome-wide prediction of lncRNA-RNA interactions. We believe that ASSA will be a useful tool to both bioinformatics and wet-lab researches to study lncRNA mechanisms and to select potential antisense partners for the RNA of interest.

DIFFERENTIAL ALTERNATIVE SPLICING IN RATS BRAIN TISSUES SELECTED BY AGGRESSIVE BEHAVIOUR

V.N. Babenko*, A.O. Bragin, I.V. Chadaeva, Y.L. Orlov
Novosibirsk State University, Novosibirsk, Russia
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: bob@bionet.nsc.ru*

Key words: *alternative splicing, RNA profiling, aggressive behavior, synaptic conduction genes, glutamate receptor NMDA, Grin1 gene*

Motivation and Aim: Alternative splicing is important basis of gene functioning and differentiation in neuronal tissues of higher eukaryotes [1]. The process of cell specialization is multi-level one that includes processes of replication, transcription and splicing, as well as miRNA regulation of splicing factors. Earlier several neurospecific splicing enhancers regulating mRNA structure of large number of gene-targets were revealed: NOVA1/2, FOX1/2, nSR100/SRM4 and silencers PTB1/2 [1, 2]. Analysis of such molecular mechanisms has a great fundamental importance in biomedicine and neurosciences. In this paper we considered differential alternative splicing of genes by analysis of RNA-Seq data of aggressive and tame rat lines selected at ICG SB RAS [3].
Methods and Algorithms: We analyzed tissue samples from several brain areas of laboratory animals including hypothalamus. Previously it was shown that these brain tissues are associated with aggressive behavior in rats.
Results: By using RNA profiling the main class of neuronal genes with alternative splicing such as genes of synaptic specializations, among which are highlighted profiles differentially spliced isoforms was identified. We studied in details difference in *Grin1* isoforms. It was a significant difference between the aggressive and tame rats in proportions of alternative transcripts in a number of the synapse gene. Such difference may determine the relevant specific behavior.
Conclusion: The deviations of proportions of synapse transcripts may be due to a change of the expression of neurospecific RNA-binding splicing proteins such as SLM1, NOVA, PTB2 and others. Overall, we present alternative splicing as molecular mechanisms affecting gene expression isoforms and behavior patterns laboratory rats.
Availability: Software is available from the author upon request.

References:

1. Venables J.P. *et al*. (2012) Tissue-specific alternative splicing is conserved in deuterostomes. Mol Biol Evol. 29 (1): 261-9.
2. Raj B., Blencowe B.J. (2015) Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. Neuron. 87 (1): 14-27.
3. Spitsina A.M. *et al*. (2015) Supercomputer analysis of genomics and transcriptomics data revealed by high-throughput DNA sequencing. Program systems: theory and applications. 6:1(23): 157–174. (In Russian)

MOLECULAR MODELING OF INFLUENZA VIRUS H1N1 HEMAGGLUTININ INHIBITION BY CAMPHOR IMINES

D.S. Baev*, A.S. Sokolova, O.I. Yarovaya, T.G. Tolstikova, V.V. Zarubaev
N.N. Vorozhtsov Novosibirsk Institute of Organic Chemistry SB RAS, Novosibirsk, Russia
Influenza Research Institute, Petersburg, Russia
** Corresponding author: baev@nioch.nsc.ru*

Key words: *influenza, H1N1, hemagglutinin, camphor imines*

Motivation and Aim: Clinical use of antiviral M2 and neuraminidase inhibitors is limited due widely distributed drug resistance [1]. This fact drives the research to identify new anti-influenza drugs with novel targets and mechanisms of activity for treatment of influenza. The influenza surface glycoprotein hemagglutinin (HA) is a potential target for antiviral drugs because of its key roles in the initial stages of infection: receptor binding and the fusion of virus and cell membranes [2].
Methods and Algorithms: The docking analysis of molecules was carried out using Autodock Vina. The structural coordinates of HA from the 2009 human pandemic influenza virus (A/California/04/2009, PDB ID 3UBE) were obtained from the protein databank. 3UBE model was superimposed with 3EYK HA model by sequence alignment for detection of binding site of the inhibitor of membrane fusion, tert-butylhydroquinone (TBHQ).
Results: Spatial characteristics of HA A/California (3UBE) structure lead to the formation of two possible cavities in TBHQ-friendly region. In the case of molecular docking grid captures both of these sites, the aliphatic imines trying to embed in specific hydrophobic site while the conservative site binds TBHQ forming hydrogen bonds.
Conclusion: Computer simulation of camphor imines interaction with viral HA suggests that the probable mechanism of their action is inhibition of HA activity by binding to hydrophobic site on its molecular surface.

References

1. F.G. Hayden et al. (2011) Emerging influenza antiviral resistance threats, *J. Infect. Dis.* 203: 6-10.
2. R.J. Russell et al. (2008). Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion, *Proc. Natl Acad. Sci. USA.* 105(46): 17736-41.

THE USE OF DISCIMINANT ANALYSIS AND ARTIFICIAL NEURONAL NETWORK IN BREAST CANCER DETECTION

U.S. Bagina*, L.V. Shchegoleva, T.O. Volkova
Petrozavodsk State University, Petrozavodsk, Russia
** Corresponding author: uliana.bagina@gmail.com*

Key words: *artificial neuronal networks, discriminant analysis, diagnosis, breast cancer*

Motivation and Aim: Tumorigenesis is accompanied with the changes in different systems of the organism. There are many factors involved that affect each other. The combined effect from two factors often exceeds the sum of the effects therefore increasing the individual effect. Important requirement of tumor development is ability of cancer cells to cause death of the lymphocytes. Therefore, in establishing the algorithm of the diagnostics of the breast cancer, we used several parameters approach – specifically, activity of caspases and the ratio of lymphocytes subsets.
Methods and Algorithms: 108 peripheral blood samples were obtained from the Oncologic clinic of Republic of Karelia. 15 of them were collected from patients with breast benign disease, 20 – stage I of breast cancer, 30 – stage II breast cancer, 43 – stage III of breast cancer. As control samples, group of 30 healthy controls was similarly studied to establish normal ranges and means. Caspase-3, -6, -8, -9 activity assay was carried out in peripheral blood lymphocytes. The ratio of T-cell subsets such as CD3, CD4, CD8, CD16, CD20, CD25 и CD95 was estimated. Statistical analysis was performed using Statgraphics Plus 5.0 software. Artificial neural network (multilayer perceptron) was developed with a 3-layer design, with one hidden layer.
Results: The 11 criteria such as activity of caspase-3, -6, -8, -9 and rate of CD3, CD4, CD8, CD16, CD20, CD25 and CD95 T-cell subsets in peripheral blood were tested as a diagnostic biomarkers. Based on discriminant analysis for 138 cases, it was found that the best separation into groups should use all 11 biomarkers. Discriminant analysis allows for a correct attributing to the group 98.9% of patients. The second approach in classifying the cases was based on the artificial neural network which included the 11 neurons on the input layer, the 5 neurons on the output layer and 6 on the single hidden layer. The network was tested on 138 observations. All of them were correctly classified and error rate was zero. Based on both discriminant analysis and neural networks using free R-statistics, we developed software that automatically with certain probability (% - discriminant analysis) and in a binary system “yes / no” (1/0) differentiates the blood samples into groups of benign breast disease, stage I, stage II or stage III of breast cancer and group without pathology.
Conclusion: Using discriminant analysis and neural network allows the development of high-precision computational tool in noninvasive differential diagnosis of breast pathologies based on peripheral blood biomarkers. Some of them – T-cells – already widely used in laboratory and clinical practice, other one – the caspase activity – does not require laborious methods of analysis and significant costs, making them available for routine determination.
Acknowledgements: Project was support by the grant No 2014/154 of the Ministry of Education and Science of Russia, scientific research number – 1713 and by the grant of Petrozavodsk State University program of strategic development.

DE NOVO SEQUENCING AND COMPARATIVE ANALYSIS OF CHLOROPLAST GENOMES FOR FOUR FERNS OF DRYOPTERIS AND ADIANTUM GENERA

M.S. Belenikin*[1, 2], A.A. Krinitsina[1], S.V. Kuptsov[1], M.D. Logacheva[1], A.S. Speranskaya[1]
[1] *Lomonosov Moscow State University, Moscow, Russia*
[2] *Pirogov Russian National Research Medical University, Moscow, Russia*
** Corresponding author: genetics.npcmpd@gmail.com*

Key words: *fern, NGS, Illumina, de novo assembly, chloroplast genome*

Motivation and Aim: Chloroplast (cp) genomes provide various genetic information for evolutionary and functional studies in plants. However, there are limited number of fern cp-genomes, for some genera date on cp-genomes are absent. Here, we present next-generation sequencing, de novo assembly and comparative analysis of cp-genomes for species of *Dryopteris* genus (*D. villarii*, *D. filix-mas*, *D. blanfordii*) and *Adiantum* genus (*A. hispidulum*).
Methods and Algorithms: Paired-end library was constructed using TruSeq or Nextera protocols. The sequencing was performed by MiSeq (Illumina) producing PE (2x300bp) read datasets. Then, after read trimming, at the first stage we have performed target pair read filtering using the fern relatives with known sequences of chloroplast genomes . At the second stage we have performed de novo assembly using a number of de-novo assemblers (velvet, mira, spades, newbler) and in-house scripts. Cp-genomes were circulated, a few gaps were closed by Sanger sequencing. Protein-coding genes were annotated by DOGMA (Wyman et al. 2004).
Results: Comparative analysis of independently de novo assembled entire cp-genomes showed high identity and gene order both *Dryopteris* and *Adiantum*. Inside genera the main differences between cp-genomes of species are short indels that located at intergenic or intranic regions. However, comparison of cp-genomes of two species of *Adiantum* (*A. hispidulum* and *A. capillus-veneris* [AY178864]) and three species of *Dryopteris* (*D.villarii*, *D.filix-mas*, *D.blanfordii*) have showed the loss of tRNA coding gene in inverted repeat regions of *Dryopteris*.
Conclusion: For *Dryopteris* genus the complete sequences of cp-genomes were obtained for the first time; comparative analysis of fern genera *Dryopteris* and *Adiantum* showed that *D. villarii*, *D. filix-mas*, *D. blanfordii* loss of the the gene, coding one of tRNA. A similar loss of the genes encoding the tRNA was found in cp-genome of tree fern *Alsophila spinulosa* (Gao et al., 2009).
Acknowledgements: This work was partially supported by RFBR grant 14-04-01852a and RSF grant 14-50-00029.
References:
1. S.K.Wyman et al. (2004) Automatic annotation of organellar genomes with DOGMA, Bioinformatics, 20(17): 3252-5.
2. L.Gao et al. (2009) Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes, *BMC Evolutionary Biology,* 9: 130.

EVOLUTION OF RESTRICTION-MODIFICATION SYSTEMS IN LARGE SCALE

O.I. Bezsudnova[1], I.S. Rusinov[1, 2], A.S. Ershova[2, 3, 4], A.S. Karyagina[2, 3, 4] S.A. Spirin[1, 2, 5], A.V. Alexeevski[1, 2, 5]*

[1] *Faculty of Bioengineering and Bioinformatics*

[2] *Belozersky Institute of Physico-Chemical Biology, Moscow State University, Russia*

[3] *Gamaleya Center of Epidemiology and Microbiology, Moscow, Russia*

[4] *Institute of Agricultural Biotechnology RAS, Moscow, Russia*

[5] *Scientific Research Institute for System Studies, RAS, Moscow, Russia*

* *Corresponding author: aba@belozersky.msu.ru*

Key words: *restriction-modification system, horizontal transfer, evolution*

Motivation and Aim: Restriction-modification (R-M) systems defend bacteria and archaea from bacteriophages and play variety of roles in populations of prokaryots [1]. They can be considered as mobile elements despite absence of own mechanism of translocation, making them interesting model for evolutionary study [2]. We studied R-M systems evolution in a large scale.

Methods and Algorithms: We downloaded data about 17327 Types I - IV R-M systems encoded in 4258 completely sequenced genomes from REBASE (http://rebase.neb.com). We divided restriction endonucleases (REases) and DNA methyltransferases (MTases) into groups with homologous catalytic domains, 39 for REases and 7 for MTases. Domains were defined according to Pfam annotations. Two R-M systems were considered homologous if catalytic domains of REases as well as catalytic domains of MTases are homologous. We found 61 class of homologous R-M systems.

Results: We considered putative evolutionary scenarios for each homology class. In first scenario an R-M system of last common ancestor was inherited mainly vertically accomplished with R-M system loss in some descendants. Putative example is N6_Mtase#ResIII homology class, which includes 3142 Type I R-M systems. Indeed, systems of this class are presented in 32% of archaeal and 18% of bacterial genomes. Moreover, they were found in all bacterial and archaeal phyla with 10 or more sequenced genomes. In second scenario, an R-M system originated in one taxon and then expanded by intensive horizontal transfer. Putative example is N6_N4_Mtase#RE_TdeIII class, which includes 21 Type II R-M systems. Seven of them are in *Firmicutes,* possible origin of the class. Other class members belong to three phyla of Archaea and six phyla of Bacteria.

Conclusion: Different classes of homologous R-M systems vary in evolutionary scenarios.

Acknowledgements. The work was partially supported by RNF grant 16-14-10319.

References:

1. A.S.Ershova et al., (2015) Restriction-Modification Systems in Prokaryotic Evolution and Ecology, *Biochemistry (Mosc)*, 80:1373-86
2. Makarova KS et al. (2013) Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res*. 41:4360-77.

HOW TO ACCOMPLISH A RAPID DEFENSE AGAINST FOREIGN DNA – RESTRICTION-MODIFICATION SYSTEMS AND IMPLICATIONS FOR SYNTHETIC GENE CIRCUITS

B. Blagojevic[1]*, M. Djordjevic[1], M. Djordjevic[2]
[1] *Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia*
[2] *Institute of Physiology and Biochemistry, Faculty of Biology, University of Belgrade, Belgrade, Serbia*
** Corresponding author: bojanab@ipb.ac.rs*

Key words: *R-M systems, transcript processing, synthetic gene circuits, CRISPR/Cas*

Motivation and Aim: Restriction-modification (RM) systems consist of restriction endonuclease (R), which cuts specific DNA sequences, and methyltransverase (M), which methylates and protects the same sequences from cleavage. It is considered that R must make a fast transition from "OFF" to "ON" state during RM establishment, so that the host cell becomes rapidly protected from foreign DNA. On the other hand, to prevent the host genome being cut by R, methylation must precede R expression. The relationship between these constraints on the enzyme expression dynamics, and the features controlling RM expression, is however unclear. To this end, we here develop a biophysical model of gene expression regulation in RM systems, to analyze dynamics of RM establishment in a naïve host. We then use this quantitative understanding to propose a synthetic gene circuits that can control how rapidly a potentially toxic molecule is expressed.
Methods and Algorithms: We develop a model of the enzyme synthesis in RM systems, based on statistical thermodynamics. We apply it to EcoRV, which is RM system with divergent CR and M promoters [1], that is under control of specialized control (C) protein. Overlapping RC and M promoters is the main feature of EcoRV, which we show is enough to ensure: *i*) the time-delayed expression of R with respect to M *ii*) the fast transition of the toxic molecule (R) from "OFF" to "ON" state *iii*) the increased stability of the steady-state of R [1]. Furthermore, we consider a novel synthetic gene circuit, which is capable of achieving a rapid cell defense against foreign DNA [2]. To this end, we combine transcription control inherent to RM systems, with the transcript processing inherent to CRISPR/Cas system [3].
Results and Conclusion: This, to our knowledge, represents the first quantitative model of expression regulation for a divergent RM system architecture. We show that EcoRV satisfies the proposed dynamical constraints, while any perturbation of system features makes these constraints less optimal. Combining transcription control of RM systems with transcript processing in CRISPR/Cas systems, allows a significantly faster transition from OFF to ON state.

References:

1. A. Rodic, B. Blagojevic, E. Zdobnov, M. Djordjevic, M. Djordjevic. (2016), submitted.
2. A. Rodic, B. Blagojevic, E. Zdobnov, M. Djordjevic, M. Djordjevic. (2016) to be submitted.
3. M. Djordjevic, M. Djordjevic, K. Severinov. (2012) CRISPR transcript processing: a mechanism for generating a large number of small interfering RNAs, *Biology Direct*, **7**:24.

STATISTICS OF INTERVALS BETWEEN SIMILAR MONOMERS: A COMPLEMENTARY WAY TO ASSESS THE STRUCTURAL PROPERTIES OF BIOLOGICAL POLYMERS

M.I. Bogachev[1]*, A.R. Kayumov[2], O.A. Markelov[1]
[1] *St. Petersburg Electrotechnical University, St. Petersburg, Russia*
[2] *Kazan (Volga region) Federal University, Kazan, Russia*
** Corresponding author: rogex@yandex.com*

Key words: *DNA structural organization, long-range correlations, interval distributions*

Motivation and Aim: Understanding the laws governing the arrangement of monomers in the primary structure of biological polymers over long scales is essential to reveal their structural organization. Since the discovery of long-range correlations in DNA sequences, it has been assessed rather in an indirect way via artificial DNA walk representations [1].
Methods and Algorithms: We focus on the distributions and correlation properties of the intervals between the occurrences of similar nucleotides in the primary DNA sequences.
Results: We have recently discovered that the distribution of intervals between all four identical nucleotides in the primary DNA sequence decays by a universal q-exponential. Analysis of 130 complete genomes of organisms at different evolutionary positions from *Archaea* and *Bacteria* to *Human* revealed that this distribution is independent of the evolutionary position of the organism. In prokaryotes, the shape parameter q of the distribution exhibits moderate variations with the changes in the GC content of the genome and in the optimal living temperature of organism [2]. Here we show that the q-exponential distribution can be explicitly reproduced by a superstatistical model that takes into account the local variations of the GC content and thus also the cumulative bending angle among short DNA segments. Our computer simulations reveal that another essential requirement is the presence of long-range correlations in the local GC concentrations and thus also their cumulative bending angles in short DNA segments. Further extensions of our model additionally account for the interval distribution variations for different GC content and optimal living temperature in prokaryotes. Our findings are in a good agreement with the state-of-the-art multiscale models of the DNA elasticity, structure and packaging [3].
Conclusion: Interval statistics between similar monomers explicitly reflect the structural properties of biological polymers and may act as powerful testbed for biomolecular models.
Availability: List of sources available at http://dx.doi.org/10.1371/journal.pone.0112534
Acknowledgements: We like to thank the Ministry of Education and Science of the Russian Federation for the financial support of this work.
References:

1. Li, W., Kaneko, K. (1992) Long-Range Correlation and Partial 1/f Spectrum in a Noncoding DNA Sequence, Europhys. Lett. 17: 655-660; Peng, C. et al. (1992) Long-range correlations in nucleotide sequences, Nature 356: 168-170.
2. Bogachev, M. I., Kayumov, A.R., Bunde, A. (2014) Universal internucleotide statistics in full genomes: A footprint of the DNA structure and packaging? PLoS One, 9: e112534.
3. Arneodo, A. et al. (2011) Multi-scale coding of genomic information: From DNA sequence to genome structure and function, Physics Reports 498: 45-188.

HUMAN AUTHENTICATION USING ELECTROCARDIOGRAM

M.R. Bogdanov*
M. Akmullah named after Bashkir State Pedagogical University, Ufa, Russia
** Corresponding author: bogdanov_marat@mail.ru*

Key words: *electrocardiogram, biometric authentication, wavelet analysis*

Motivation and Aim: Biometric methods of human identification are discussed. The mathematical aspects of human identification based on electrocardiogram are considering.
Results of electrocardiogram recognition with wavelet analysis are presented.
Methods and Algorithms: wavelet analysis.
Results: 97% accuracy while biometric identification was achieved.
Conclusion: Biometric methods of identification are all wider application in our lives. Such methods of identification like fingerprint recognition, iris, voice, or palms recognition gradually enter a phase of maturity and are increasingly beginning to be used in a variety of mobile, web and other applications, however, the evidence suggests that the traditional methods of identification inherent vulnerability . Yield is seen in the development of multi-modal identification systems using biometrics such as, for example, an electrocardiogram, an electroencephalogram and DNA. The most preferred option in our opinion, is the use of the ECG.
Availability: The technology may be used in various security applications.
Acknowledgements (if necessary)
References:
1. Bogdanov M.R. Using of wavelet analysis under Electrocardiogram Recognition. ITIDS'2015. The 3 rd International Conference on Information Technologies for Intelligent Decision Making Support. Ufa, Russia. May 18 - 21, 2015.
2. Bogdanov M.R. Unconventional encryption algorithm [Netradicionnyj algoritm shifrovaniya]. Electronic scientific journal "Computer facilities and software" [Elektronnyj nauchnyj zhurnal "Vychislitelnaya tehnika i programmnoe obespechenie"]. 2014, Release 1(1) January-April, pp. 17-23. [Online]. Available at: http://www.computer-facilities-andsoftware.ingnpublishing.com/archive/2014/release_1_1_january-april/bogdanov_m_r_netradicionnyj_algoritm_shifrovaniya/Received: 2014-03-12 Accepted: 2014-04-23 Published on-line: 2014-04-30.
3. Bogdanov M.R., Zakharov A.V., Gabidullin Ju.Z., Dumchikov A.A.Unconventional encryption algorithms.. ISSN: 2319. 5967. ISO 9001:2008 Certified. International Journal of Engineering Science and Innovative Technology (IJESIT). Volume 3, Issue 5, September 2014.
4. Bogdanov M.R., Gabidullin Ju.Z., Dumchikov A.A., Zakharov A.V., Gorbunova V.Ju. Method for recognition of bird's voices with wavelet analysis. Russian Patent #2014611697. 07.02.2014.
5. Bogdanov M.R. The way of researching of biological diversity of ornitophauna with wavelet analysis. Proceedings of Samara Scientific Centre of RAS. #3(4). 2013. P. 1232-1236.

INTERSPERSED REPETITIVE SEQUENCES DISTRIBUTION IN HUMAN CHROMOSOMES ANALYZED BY IN SITU HYBRIDIZATION AND IN SILICO ANALYSIS

A.G. Bogomolov[1, 2]*, T.V. Karamysheva[1], N.B. Rubtsov[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *Corresponding author: mantis_anton@bionet.nsc.ru*

Key words: *interspersed repetitive sequences, fluorescence in situ hybridization (FISH), visualization special signal in silico (VISSIS), image analysis*

Motivation and Aim: In 2012, we developed method VISSIS [1] based on comparative FISH of pairs of whole chromosomes painting probes (WCPs) performed without suppression of repetitive DNA sequences hybridization. The obtained indicated to the influence of the paint sets on the efficiency of method VISSIS application. In case of paints derived from chromosomes, contrasting in content of LINE/ SINE repeats identification of C-negative chromosome regions appeared to be complicated. We suggest this problem could be solved by normalization of the FISH signal intensity based on data of LINE/ SINE repeats extracted from draft human genome sequence. The aim of this work is to analyze the results of FISH image analysis obtained with WCPs derived from different chromosome pairs, taking into account the contents of LINE/ SINE repeats of original chromosomes and painting of the chromosome regions bearing heavy isochors.
Methods and Algorithms: DNA probes of individual chromosomes were generated by microdissection following DOP-PCR amplification [2]. FISH of microdissection DNA probes generated from human chromosomes 2, 3, 7, 10, 15, 18, 19 and X with human chromosomes was performed without suppression of repetitive DNA hybridization. The chromosomes were characterized using data of GC-content and repetitive DNA sequences in chromosomes (SINE, LINE, Alu, L1, L2). The average intensity profile and histogram of intensities were used to compare distribution intensities in chromosomes in processed images.
Results: Currently a set of microscopic images obtained, which will be involved in further research. The software for analyzing and comparison of FISH signal intensities in chromosome regions after computer processing of FISH images by VISSIS was developed and applied. Normalization of the FISH signal intensity based on data of LINE/ SINE repeats extracted from draft human genome sequence was performed and its efficiency was estimated
Acknowledgements: The reported study was partially supported by the Ministry of Education and Science of the Russian Federation and RFBR, research projects No. 16-34-00498, No. 14-04-00086.

References:

1. A.G. Bogomolov et al. (2012) Visualization of chromosome-specific DNA sequences by FISH of microdissection DNA Probes with metaphase Chromosomes, *Russian Journal of Genetics,* 2(6): 413-420
2. N.B. Rubtsov et al. (2000) Chromosome microdissection using NIR Femtosecond laser pulses and generation of band specific DNA-libraries with DOP-PCR, *Cell Mol. Biol,* 46.

THE SOFTWARE FOR ESTIMATION OF TELOMERE LENGTH ON INDIVIDUAL CHROMOSOME ARMS IN IMMUNOPATHOLOGY

A.G. Bogomolov[1, 3]*, M.S. Barkovskaya[2], N.B. Rubtsov[1, 3], V.A. Kozlov[2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Research Institute of Fundamental and Clinical Immunology, Novosibirsk, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
* *Corresponding author: mantis_anton@bionet.nsc.ru*

Key words: *quantitative fluorescence in situ hybridization (Q-FISH), telomere length measurements, integrated fluorescence intensity(IFI),image analysis, immunopathology*

Motivation and Aim: The telomeres are important to maintain chromosome stability, and the decrease of telomere length may lead to immunopathology. This paper presents the enhanced Q-FISH protocol and software for estimation of telomere length on individual chromosome arms in immune-associated diseases.
Methods and Algorithms: Two sets of images were obtained according to the standard [1] and modified (change of temperature of hybridization, concentration Cy3-PNA $(CCCTAA)_3$-probes in hybridization mixture, fixation and washing conditions) Q-FISH protocols. Bead images were obtained for each image capture session of metaphase chromosomes.
The developed software can be divided into two functional blocks: bead image analysis and estimation of telomere length on individual chromosome arms.
Bead image analysis includes: finding beads (iterative selection method), division images into blocks, calculation of average fluorescence value of the beads in the whole image and in each block. Calculated average values are used not only for intensity calibration and comparing fluorescence measurements between experiments, but also to correction of irregular light effects.
Image analysis of metaphase chromosomes means: segmentation of chromosomes and telomeres (methods of thresholding segmentation and methods based on average difference filter were used), determination of telomere belonging to the chromosomes, correction of irregular light effects and fluorescence intensity calibration using beads, calculation integrated fluorescence intensity (the background level is estimated and subtracted from measured value).
Results: The MeTeLen software for estimation of telomere length in images of metaphase chromosomes was developed. A series of validation experiments was carried out and confirmed the correctness of the obtained results. The optimum parameters for the signal normalization were selected and different estimations of background level were compared. It was also shown that a modified Q-FISH protocol can be used to calculate telomere length on individual chromosome arms. Moreover chromosome morphology was better preserved improving the image quality and quality of subsequent cytogenetic analysis.
Acknowledgements: This study was supported by RNF grant № 14-15-00346.
References:
1. S.S. Poon, P.M. Lansdorp (2001) Quantitative fluorescence in situ hybridization, In: Current Protocols in Cell Biology, p 18.4.1 – 18.4.21 (John Wiley & Sons)

RIBOSOMAL GENES AS PHYLOGENETIC MARKERS FOR STUDING EVOLUTION OF BLUE-FLOWERED FLAXES

N.L. Bolsheva[1]*, N.V. Melnikova[1], A.A. Dmitriev[1], M.S. Belenikin[1,2], A.S. Speranskay[2], A.A. Krinitsina[2], G.S. Krasnov[1], V.A. Lakunina[1], A.V. Snezhkina[1], A.F. Sadritdinova[1], T.A. Rozhmina[3], A.V. Amosova[1], T.E. Samatadze[1], O.Yu. Yurkevich[1], N.G. Shostak[1], S.A. Zoshchuk[1], A.V. Kudryavtseva[1], O.V. Muravenko[1]

[1] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

[2] Department of Higher Plants, Lomonosov Moscow State University, Moscow, Russia

[3] All-Russian Research Institute for Flax, Torzhok, Russia

** Corresponding author: nlbolsheva@mail.ru*

Key words: *flax, phylogeny, rRNA genes, high-throughput sequencing, karyotype, FISH*

Motivation and Aim: The species relationships within the genus *Linum* have already been studied several times by means of different molecular and phylogenetic approaches. Nevertheless, a number of ambiguities in phylogeny of *Linum* still remain unresolved. In particular, the species relationships within the sections *Stellerolinum* and *Dasylinum* need further clarification. Also, the question of independence of the species of the section *Adenolinum* still remains unanswered. Moreover, the relationships of *L. narbonense* and other species of the section *Linum* require further clarification. Additionally, the origin of tetraploid species of the section *Linum* (2n=30) including the cultivated species *L. usitatissimum* has not been explored. The present study examines the phylogeny of blue-flowered species of *Linum* by comparisons of 5S rRNA gene sequences as well as ITS1 and ITS2 sequences of 35S rRNA genes.

Methods and Algorithms: High-throughput sequencing has been used for analysis of multicopy rRNA genes families. In addition to the molecular phylogenetic analysis, the number and chromosomal localization of 5S and 35S rDNA sites has been determined by FISH.

Results: Our findings confirm that *L. stelleroides* forms a basal branch of the clade of blue-flowered flax species which is independent from the branch formed by the species of the section *Dasylinum*. The current data as well as the results of genomic DNA fingerprinting and cytogenetic investigations described previously could not discriminate separate species within the section *Adenolinum.* The allotetraploid cultivated species *L. usitatissimum* and its wild ancestor *L. angustifolium* (2n=30) could originate either as the result of hybridization of two diploid species (2n=16) related to the modern *L. gandiflorum* and *L. decumbens*, or hybridization of a diploid species (2n=16) and a diploid ancestor of modern *L. narbonense* (2n=14).

Conclusion: High-throughput sequencing of multicopy rRNA gene families allowed us to make several adjustments to the phylogeny of blue-flowered flax species and also estimate the levels of intraspecific rRNA gene sequence polymorphism and interspecific differences in these sequences.

Acknowledgements: This work was financially supported by the RAS Presidium Program “Biodiversity of natural systems” (subprogram “Gene pools of nature and their conservation”) and the Russian Foundation for Basic Research, grant 16-04-01239.

SIBERIAN LARCH CHLOROPLAST GENOME ANALYSIS OVER TRIPLET FREQUENCY DISTRIBUTION

E.I. Bondar[1]*, Y.A. Putintseva[1, 2], K.V. Krutovsky[2, 3, 4]
[1] *Siberian Federal university, Krasnoyarsk, Russia*
[2] *Institute of forest of SD RAS, Krasnoyarsk, Russia*
[3] *University of Göttingen, Göttingen, Germany*
[4] *Texas A&M University, College Station, Texas, USA*
* *Corresponding author: bone-post@ya.ru*

Key words: *frequency; triplet; order; cluster; elastic map*

Motivation and Aim: Structures observed within a DNA sequence reveal an order and provide further understanding of functional roles of a sequence or its fragments. A new function (or a connection between function and structure) may manifest through new patterns found in symbol sequences corresponding to DNA molecule. Here we report the seven-cluster structure observed over a chloroplast genome of *L. sibirica* Ledeb.
Methods and Algorithms: The chloroplast genome sequence Siberian larch (*L. sibirica* Ledeb.) has been sequenced using the Illumina HiSeq2000 sequencer at the Laboratory of Forest Genomics of the Siberian Federal University [1]. The length of the genome sequence is 122 561 bp. Genome contains 121 coding regions. GC content is 38.75%. Analysis of the *L. Sibirica* chloroplast genome takes several steps. Firstly, the genome sequence was mapped into a set of equal length overlapping fragments. Numbers of different non-overlapping triplets in each fragment were counted. Then, a frequency dictionary, which contains the set of all the triplets counted within the fragments, was developed. The clusters in 63-dimensional space were identified with elastic map technique, where the objects to be clusterized are the fragments identified in the sequence. The 63-dimentional space is provided with frequencies of non-overlapping triplets found in the different fragments of the genome. The fragments comprising the clusters were checked against the annotation of the genome. Thus, clusters where coding fragments prevail and those where non-coding fragments prevail had been identified.
Results: Unlike the patterns described in [2], the structure observed here differs: the nodes comprising coding and non-coding regions are located in the basically different manner. Among the clusters three gene-coding, three non-coding clusters are found and one central cluster has both of coding and non-coding fragments in equal proportions. Also, for coding clusters their genes were identified.
Conclusion: Seven cluster structure in chloroplast genome for *L. sibirica* was found. Unlike nuclear genome of bacteria, the chloroplast genome yields more complex structure. Further studies may bring new understanding of a fine structure details, or of relations between structure and function of chloroplast genome.

References:

1. Krutovsky K. et al. (2014) Some preliminary results of a full genome de novo sequencing of *Larix sibirica* Ledeb. and *Pinus sibirica* Du Tour. *Siberian Forest Journal*. 1(4): 79 – 83.
2. Gorban A. N. et al. (2005) Four basic symmetry types in the universal 7-cluster structure of microbial genomic sequences. *In Silico Biology* 5: 25-37.

THE OPPOSING EFFECTS OF SHORT- AND LONG-TERM SOCIAL STRESS ON PREFRONTAL CORTEX TRANSCRIPTOME

N.P. Bondar[1]*, L.O. Bryzgalov[1], N.E. Ershov[1, 3], F.E. Gusev[3, 5], V.V. Reshetnikov[1], D.F. Avgustinovich[2], M.V. Tenditnik[4], E.I. Rogaev[3, 5], T.I. Merkulova[1]

[1] Laboratory of Gene Expression Regulation, Institute Cytology and Genetics SB RAS, Russia

[2] Laboratory of Molecular Mechanisms of Pathological Processes, Institute Cytology and Genetics SB RAS, Russia

[3] The Center of Brain Neurobiology and Neurogenetics, Institute Cytology and Genetics SB RAS, Russia

[4] Laboratory of Experimental Models of Emotional Pathologies, Institute of Physiology and Fundamental Medicine, SB RAMS, Russia

[5] University of Massachusetts Medical School, USA

** Corresponding author: nbondar@bionet.nsc.ru*

Key words: *social defeat stress, depression, RNA-seq, prefrontal cortex*

Motivation and Aim: Chronic social defeat stress is a well-validated murine model of depression. However, little is known about the gene activity dynamics during the development of a depression-like state.

Methods and Algorithms: We analyzed the effects of social defeat stress of varying duration (10 and 30 days) on the behavioral patterns and prefrontal-cortex transcriptome in C57BL/6 mice.

Results: Commonly used 10-day exposure to social defeat stress resulted in a high level of social avoidance with no sign of depression-associated behavior. Contrariwise, most animals exposed to 30-day stress demonstrated clear hallmarks of depression, including higher level of social avoidance, increased immobility in the forced swim test, and anhedonic behavior. The monitoring of transcriptome changes revealed massive alterations in gene expression on the 10th day. Surprisingly, expression of a few genes was only affected on the 30th day of stress, apparently, due to a reversal of the majority of the early stress-induced changes to the original basal state. Moreover, we have found that glucocorticoid-sensitive genes are clearly enriched targets on the 10th day of stress, but these genes stop responding to the elevated corticosterone level after the 30th day of stress. The majority of genes altered by 30-day stress were downregulated, with the most relevant ones participating in chromatin-modifications and neuroplasticity (e,g, guanine nucleotide exchange factors (GEFs) of Rho-family GTPases).

Conclusion: Taken together, our data suggest that depression may be caused by weakening of the response to the stressful environmental factors in terms of both behavior and gene expression. Our results also support the hypothesis that major depressive disorder is associated with defective cell adhesion and impaired neuronal plasticity.

Acknowledgements: This work was supported by grant from the Government of the Russian Federation # 14.B25.31.0033 and a grant from the Russian Science Foundation # 16-15-10131.

RNA-SEQ DATA ANALYSIS OF RATS WITH AGGRESSIVE BEHAVIOR IN THREE BRAIN AREAS

A.O. Bragin*, A.L. Markel, V.N. Babenko, I.V. Chadaeva, E.S. Tiys, Y.L. Orlov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: ibragim@bionet.nsc.ru*

Key words: *RNA-seq, aggressive behavior, differential gene expression, brain, laboratory animals, rat*

Motivation and Aim: Manifestations of aggression and aggressive behavior in the human society are among the most urgent social problems. It is concluded from the studies conducted in recent decades that aggression should be considered as a special feature of behavior based also on the genetic and sociobiological history of the development of humans as a biologic species, of an individual, and of social group and environment [1]. With respect to the notion that aggression is a general biological phenomenon stemming from deep evolutionary roots, the investigation of aggressive behavior demands that not only the aggressive behavior of a human be analyzed but also the corresponding features of animal behavior.
Methods and Algorithms: For analyzing of genetics factors which determine aggression behavior two rat lines were selected over more than 30 years at the Institute of Cytology and Genetics. First rat line was selected for tame behavior towards human. Other rat line was selected for aggressiveness. In our work we have analyze three rat brain area: Hypothalamus; Ventral tegmental area and Midbrain raphe nuclei (aqueduct) [2]. Tissue samples were processed for RNA extraction. This was followed by RNA-sequencing and filtration of reads. After that we use TopHat for mapping reads on the reference rat genome. For assessment of gene expression level and finding differently expressed genes we used Cufflinks.
Results: We found 95 differently expressed gene in hypothalamus, 189 in Ventral tegmental area and 136 in an aqueduct between aggressive and tame rats. We use DAVID gene ontology tools to find biological processes with overrepresented rat differentially expressed genes. Many of these processes are associated with neuronal activity and behavior. Using ANDvisio system the associative genes networks of the differentially expressed genes with molecular interaction were reconstructed.
Conclusion: The RNA-seq analysis of the three brain areas of rat confirmed association of some genes with rat aggression.
Acknowledgements: The work was supported by RSF grant 14-14-00269.
References:
1. N.N. Kudryavtseva, A.L. Markel, Y.L. Orlov (2014) Aggressive behavior: genetic and physiological mechanisms, *Vavilov journal of genetics and breeding*, 18(4/3):1133–1155. (In Russian).
2. V.N. Babenko et al. (2016) Analysis of transcriptome data on gene expression in brain areas of rats, selected by aggressive behavior, *In: Proceedings of «Neuroinformatics-2016»*, *MEPhI, Moscow*, P.**2**:82-92. (In Russian).

ИСПОЛЬЗОВАНИЕ МИКРОФЛЮИДИКИ DOLOMITE ДЛЯ СЕКВЕНИРОВАНИЯ ТРАНСКРИПТОМОВ ОТДЕЛЬНЫХ КЛЕТОК

D. Brittal
ООО «Диаэм», Москва, Россия
** Corresponding author: db@dia-m.ru*

Key words: *sequencing technologies, transcriptome*

Микрофлюидика Dolomite – технология, позволяющая работать с очень малыми объемами жидкостей, газов, с кристаллическими и полимерными частицами, клетками животного, растительного и бактериального происхождения, пузырьками и каплями с возможностью наблюдать за ними, манипулировать ими и контролировать процессы, протекающие с ними.

Это дает возможность проводить «традиционные» исследования в миниатюрном формате, а также проводить исследования, которые ранее были невозможны.

Особенности и возможности микрофлюидики Dolomite: работа с микрообъектами (капли, клетки, частицы, пузырьки); работа с микро- и нанообъемами (диаметр канала от 10 нм); высокая воспроизводимость: точность дозирования – порядка пиколитра; точный контроль параметров процесса: температуры, скорости потоков, давления, смешивания; большая библиотека «стандартных» чипов; чипы произвольной конфигурации и геометрии: многослойные и составные чипы с разными свойствами поверхности каналов, интеграция на одном чипе различных стадий процессов для ускорения и автоматизации методик исследований; интеграция с приборами, детекторами, системами пробоподготовки и сенсорами (хроматографами, масс-спектрометрами, лазерами, спектрофотометрами, микроскопами и т.д.); автоматизация процессов: удобство, высокий выход, воспроизводимость, точность; объединение разных стадий методик в одном приборе; уменьшение размеров приборов; появление новых методов и приборов. Технология микрофлюидики Dolomite находит применения в таких областях как: химический синтез, аналитическая химия, физико-химические исследования; разработка лекарственных препаратов, определение эффективности и цитотоксичности; биология, диагностика и медицина; экология, производство, приборостроение. Биология, диагностика и медицина: качественный и количественный анализ фрагментов НК на чипе капиллярного электрофореза; чипы для секвенирования НК; цифровая капельная ПЦР для количественной ПЦР-диагностики с высокой точностью; анализы крови (биохимические, ИФА, на глюкозу и т.д.); изоляция ДНК из цельной крови; наблюдение за иммобилизованными эмбрионами и клетками.

Система для инкапсуляции клеток или нуклеиновых кислот в капли μEncapsulator 1: Отличная пробоподготовка для изучения экспрессии генов, ПЦР, сортировки и др.; инертные, биосовместимые материалы; возможность поддержания жизнеспособности клеток; пропускная способность: 300 000 клеток в 3 млн капель за 15 минут; инкапсуляция 100 мкл образца и реагента, или двух последовательных образцов по 100 мкл.

Система для создания библиотек единичных клеток для последующего секвенирования: Транскриптомика одиночных клеток; высокая точность, воспроизводимость, надежность, инертность материала чипа; скорость инкапсуляции, капель/с – 4 000 – выше, чем у чипов из PDMS[1].

[1] Macosko et al. (Cell 161:1202)

DELINEATING SINGLE CELL LIFE/DEATH DECISIONS IN THE CD95/FAS NETWORK

J.H. Buchbinder[1,+], D. Pischel[2,+], K. Sundmacher[2], R.J. Flassig[2], I.N. Lavrik[1]*
[1] *Department of Translational Inflammation Research, Otto-von-Guericke University Magdeburg, Germany*
[2] *Max-Planck-Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany*
[+] *equal contributors*
* *Corresponding author: inna.lavrik@med.ovgu.de*

Key words: *CD95/Fas/APO-1, cell death, NF-κB, single cell model, life/death decision, imaging flow cytometry*

Motivation and Aim: Death receptor (DR) stimulation induces apoptosis and strong anti-apoptotic responses at the same time. To solve this contradiction and understand how the decision between life and death is taken at the single cell level, we analyzed the dynamics of NF-κB activation and apoptosis responses in the same cells, which was exemplified for the CD95/Fas/APO-1 signaling pathway.
Methods and Algorithms: We have developed an imaging flow cytometry based method that enables quantitative detection of different CD95 signaling outcomes. We are able to monitor NF-κB translocation into the nucleus and activation of the apoptotic caspase-3 in single cells (Schmidt et al., 2015). With the quantitative data from a large number of cells we have developed a mathematical model describing the CD95 network at the single cell level.
Results: Our data indicate that CD95 stimulation leads to NF-κB activation and apoptosis in the same cells. Analyzing of the computational model shows that the strength of CD95-mediated NF-κB activation is invariant with respect to CD95 stimulation strength within various cells and stimulation strength. However, the ability to undergo apoptotic cell death is strongly dependent on CD95 stimulation strength. We identified the ratio between the time of survival signaling (TOS) and the time of cell death decision (TOD) as an indicator for the signaling outcome.
Conclusion: Taken together, we uncovered the regulation of these two opposing pathways in the single cell and, furthermore, our findings indicate a relatively simple strategy how a cell might avoid apoptosis and become resistant towards cell death.
Acknowledgements: The work is supported by Russian Science Foundation 14-44-00011.

References:

1. J.H. Schmidt et al. (2015) Quantification of CD95-induced apoptosis and NF-κB activation at the single cell level, *J Immunol Methods,* 423: 12-7.

MODELING OF TWO PHASES IN DROSOPHILA SENSORY ORGAN PRECURSOR CELL DETERMINATION

T.A. Bukharina[1], D.P. Furman[1, 2]*, V.P. Golubyatnikov[2, 3], M.V. Kazantsev[4]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[4] *Polzunov Altai State Technical University, Barnaul, Russia*
** Corresponding author: furman@bionet.nsc.ru*

Key words: *D. melanogaster, bristle pattern, sensory organ precursor cell formation (SOPC), Central Regulatory Circuit (CRC), mathematical model*

Motivation and Aim: The bristle pattern of *D. melanogaster* is one of the attractive model objects for studying development of ordered spatial structures in multicellular organisms. The bristle positions are strictly determined by the positions of SOPC. The goal of this work is construction of an extended mathematical model of SOPC formation under the control of CRC.
Methods and Algorithms: The description of CRC and approach to modeling of its functioning are presented in [1].
Results: SOPC specialization in *D. melanogaster* is determined by CRC functioning and takes 14-16 hours [2]. The dynamics of proteins content during this time interval has two appreciable periods. The first one (I, 0-10 hours) is characterized by predominance of interactions which amplify expression of the *AS-C* genes. During the second one (II, 10-16 hours) CRC functioning is directed to eliminate the AS-C proteins from the cell through activation of the adaptor protein Phyl followed by inhibition of the *AS-C* genes.
Conclusion: Numerical experiments generate a hypothesis that there is a mechanism which blocks appearance of Phyl during the first period.
Acknowledgements: Budget project 0324-2015-0003 and RFBR, grant 15-01-00745.
References:
1. V.P. Golubyatnikov et al. (2015) A model study of the morphogenesis of *D. melanogaster* mechanoreceptors: the central regulatory circuit, *JBCB*, 13(1): 1540006.
2. P.J. Chang et al. (2008) Negative-feedback regulation of proneural proteins controls the timing of neural precursor division, *Development*, 135(18): 3021-3030.

NUMERICAL MODEL OF DROSOPHILA SENSORY ORGAN PRECURSOR CELL DETERMINATION

T.A. Bukharina[1], D.P. Furman[1, 2], V.P. Golubyatnikov[2, 3]*, M.V. Kazantsev[4]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
[4] *Polzunov Altai State Technical University, Barnaul, Russia*
* *Corresponding author: glbtn@math.nsc.ru*

Key words: *Sensory Organ Precursor Cell (SOPC), Central Regulatory Circuit (CRC), dynamical system, stability*

Motivation and Aim: SOPC determination is the main event in development of *D. melanogaster* bristles. We describe one mathematical model of SOPC formation under the control of CRC in order to perform numerical simulations of this process.
Methods and Algorithms: Modeling of CRC functioning and analysis of the numerical results follows mathematical constructions presented in [1].
Results: We study phase portrait of 6-dimensional nonlinear dynamical system which simulates two stages formation of SOPC. This process takes up to 16 hours [2]. In our mathematical model this interval was split into two periods with different dynamics. The first period takes 10 hours, and it is characterized by absence of the Phyllopod protein in the CRC dynamics. Here, the AS-C proteins concentration grows with decreasing speed due to feedbacks structure in the system. By the end of the first period, the system approaches to its equilibrium state with maximal AS-C concentration. Appearance of Phyllopod in CRC during the second period (10-16 hours) induces decreasing of AS-C amount to almost zero values by the end of this period.
Conclusion: The two-phases model is the fullest description of SOPC determination. Results of our numerical experiments correspond to available biological data, see [2].
Acknowledgements: Budget project 0324-2015-0003 and RFBR, grant 15-01-00745.
References:
1. V.P. Golubyatnikov et al. (2015) A model study of the morphogenesis of *D. melano-gaster* mechanoreceptors: the central regulatory circuit, *JBCB*, 13(1): 1540006.
2. P.J. Chang et al. (2008) Negative-feedback regulation of proneural proteins controls the timing of neural precursor division, *Development,* 135(18): 3021-3030.

EVOLUTION FEATURES OF THE THREE CODON POSITIONS IN GENE OF ENVELOP PROTEIN E FOR DIFFERENT GENOTYPES OF THE TICK-BORNE ENCEPHALITIS VIRUS

Yu.S. Bukin[1, 2]*, Yu.P. Dzhioev[3, 4], I.V. Kozlova[4], S.E. Tkachev[5], D.O. Kiselev[3], A.I. Paramonov[3], O.N. Reva[3], V.I. Zlobin[3]

[1] *Limnological Institute SB RAS, Irkutsk, Russia*

[2] *Irkutsk State Technical University, Irkutsk, Russia*

[3] *Irkutsk State Medical University, Irkutsk, Russia*

[4] *Scientific Center of Family Health Problems and Human Reproduction SB RAS, Irkutsk, Russia*

[5] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*

Key words: *tick-borne encephalitis virus, gene of envelope protein E, molecular phylogenetic analysis, saturation substitutions, codon position, population genetic*

Motivation and Aim: Tick-borne encephalitis virus (TBEV) is one of the most dangerous infections p distributed over a wide area of the Eurasian continent [1]. Provided four TBE virus genotype - Far East, Siberian, Western (European) and dedicated enough recently in eastern Siberia 886-th genotype. The most popular marker for phylogenetic studies TBEV is gen E of envelope protein. In phylogenetic studies based on envelope protein E doesn't estimated contribution to the effect of saturation substitutions. Saturation effect is possible distortion of phylogenetic analysis. The aim of the study was to determine the effect on the results of molecular phylogenetic analysis of the saturation substitutions in different codon position in gen E of envelope protein.

Methods and Algorithms: Selection of the most appropriate model for the evolution of the nucleotide sequences of the gene E of TBE and calculation parameters was carried out with the help of «jModelTest v. 2.1.7». Test saturation substitutions of various codon position was performed in the program «DAMBE». In the population-genetic level, the sample of sequences for gene E was investigated by using of the program «DnaSP v 5». For reconstruction of the evolutionary history of genotypes of TBEV was used of a pact «ape» for the R programming language and the program «MrBayes v 3.2.0».

Results: When phylogenetic analysis TBEV virus genotypes at the third codon position of gen E saturation effect is observed, which leads to distortion at a phylogenetic evaluation of events in the sequence and dating of branch nodes. Using only the first and second codon positions gene E of TBEV not reliably split virus closely related genotypes and genotypes of the strains within the insufficient number of polymorphic sites. Genetic diversity within a virus genotypes formed by different population processes, genetic diversity of the Far Eastern and Siberian genotypes formed through the neutral genetic drift, western genotype exposed to dynamic selection, reduce the variety of coat protein gene nucleotide sequences of E.

References:

1 J. L.Goodman, D. T.Dennis, D. E. Sonenshine. (2005). Tick-borne diseases of humans. *American Society of Microbiology*.

STOCHASTIC MODEL OF SPECIATION, WHICH DESCRIBES THE EVOLUTIONARY BRANCHING PROCESS WITHIN THE SPECIES FLOCK IN A CLOSED ECOSYSTEM

Yu.S. Bukin[1, 2], D.Yu. Sherbakov[1, 3]
[1] *Limnological Institute SB RAS, Irkutsk, Russia*
[2] *Irkutsk State Technical University, Irkutsk, Russia*
[3] *Irkutsk State University, Irkutsk, Russia*

Key words: *speciation, the probability of divergence, ecological niche, stochastic processes, stochastic process with continuous time, Markov process, Kolmogorov equation for the probability of the state*

Motivation and Aim: As part of the biosphere of the earth there is such a thing as an isolated ecosystem. Isolated ecosystems actively exchanged with the environment of matter and energy, but in a much more less exchange with other organisms of different types of ecosystems. If this ecosystem is there for a long period of time, the process of biological evolution, there accumulated a large number of endemic species. Often formed of the endemic species flock, containing several species, or even several genus and family of organisms (species flock, genus flock, and families flock*)*. Examples of such ecosystems are ancient lake (Lake Baikal, Lake Tanganyika, Lake Caspian Sea); deep seeps and mud volcanoes; island ecosystems. In the study of ecosystems detailed question about how much time is required on the formation of endemic species flock. In the event of a closed ecosystem, with a large number of potential ecological niches and a small amount of initial species, how quickly things are not to be occupied ecological niches occupied by new species? This problem appears two independent parameters: probability of speciation (species divergence of existing species into two species with the occupation of unoccupied ecological niches) per unit of time; the number of potential ecological niches available for settlement.

Methods and Algorithms: To solve the problem, we used the methods of the theory of stochastic processes. The system in the process of evolution can be in a number of discrete states. Each state corresponds to a certain number of occupied and vacant ecological niches. For an infinitely small amount of time possible settlement of only one ecological niche, species, arising as a result of evolutionary divergence. Random process running in such a system can be described by Kolmogorov`s equations for the probabilities of state [1].

Results: As a result, it was found that the dependence of the mean time of settlement of ecological niches in the system is linearly dependent on the probability of speciation. From the original number of free ecological niches mean time settlement system depends on a logarithmic function. Since the logarithm function is a slow-growing function. Then for a sufficiently large number of ecological niches for a system with a large number of niches need not much longer period of time for settlement than for systems with fewer niches.

References:

1. Karlin S.A. 1968. first course in stochastic processes. Academic Press New York and London.

ONLINE SCRIPTING TOOL FOR RETRIEVING 3D HUMAN GENOME DATA

A. Butyaev*, J. Waldispüh[1]
McGill University, Montreal, Québec, Canada
** Corresponding author: alexander.butyaev@mail.mcgill.ca*

Key words: *Human Genome, 3D Genome Browser, Scripting, JavaScript*

Motivation and Aim: Our understanding of genomes has been significantly improved by the release of their three-dimensional (*3D*) structure. Our previous work [1] introduced a novel storage technology that provides a fast and convenient access to genomic data. On top of this technology, we also built an interactive genome visualization tool: the 3D Genome Browser (*3DGB*). Nonetheless, to the best of our knowledge, a fully accessible online resource that can provide an instant access to genomic data does not exist. In this work, we have implemented the Online Scripting Tool (*OST*), as an extension for *3DGB*, to easily retrieve, process and manipulate the genomic data from the web page.
Methods and Algorithms: OST is implemented in JavaScript. It improves the user experience with our system, through a wide range of tools, by maintaining a simple environment for the user (web-browser). Original API for *3DGB*, various user-written packages and scripts, gene expression dataset and libraries such as jQuery (https://jquery.com/) are already imported into the system and they could be used at any time. Moreover, user can also include 3rd party libraries to simplify the processing of data.
Results: To prove the efficiency of the OST, we designed an experiment to explore the *3D* neighborhood of the Retinoblastoma 1 gene (RB1) - a tumor suppressor gene that has been associated with many types of cancer. Starting with a 3D neighborhood, which is centered on the promoter of the RB1 gene in the 3D structure of chromosome 13 in the normal B-cells (GM06990), we retrieve the list of SNPs that can be found in the promoter region of RB1, and in other DNA strands that are not in the immediate sequence neighborhood of the RB1 promoter. In addition, we found 3 other strands in the spatial vicinity of RB1, which are located in a radius R=0.2. In total, 1199 SNPs were identified in this 3D neighborhood, for which we extracted their phenotype. As expected, we identified SNPs related to various types of cancer. Furthermore, we found one SNP (rs10492604) related to sleep disorders in one of the strand. While there are only two SNPs of this type in the chromosome, it is interesting that another SNP (rs10492507) was also found in the vicinity of the RB1 gene (R=0.53). This proof of concept experiment is implemented using the OST and researchers can use it to explore the 3D structure of the human genome.
Conclusion: The Online Scripting Tool is an interactive tool to perform a simple and efficient manipulation of the 3D Genome Browser database. We designed an experiment to explore the 3D neighborhood of RB1 gene, created JavaScript package and documented it to show usability and power of the tool.
Availability: Our Online Scripting Tool is freely available at: http://3dgb.cs.mcgill.ca/scripting.html.
References:

1. A. Butyaev et al. (2015) A low-latency, big database system and browser for storage, querying and visualization of 3D genomic data, *Nucleic Acids Research*, 43(16): e103.

UGENE: A TOOLKIT FOR TEACHING STUDENTS

I.V. Bykova[1]*, O.I. Golosova[1], A.Y. Bakulina[2, 3], D.A. Afonnikov[2, 4], D.Y. Kandrov[1], A.Y. Palyanov[2, 5], G.A. Grekhov[1], Y.E. Danilova[1]

[1] *Unipro Center of Information Technologies, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *State Research Center of Virology and Biotechnology VECTOR, Koltsovo, Novosibirsk region, Russia*

[4] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[5] *Institute of Informatics Systems SB RAS, Novosibirsk, Russia*

** Corresponding author: ibykova@unipro.ru*

Key words: *Education, bioinformatics software, data analysis*

Motivation and Aim: Modern biological experiments in many cases require bioinformatics methods for planning and subsequent data analysis with application of complex computing algorithms.

Though there are serious commercial bioinformatics packages for these purposes, they sometimes are not available for the students. On the other hand free useful program tools and algorithms are uncoordinated in many cases and young biologists have to be experienced in programming to work with them successfully. Thus, young scientists encounter big obstacles in mastering bioinformatics hands-on.

Methods and Algorithms: UGENE is developed as an open-source free software aimed to assist a molecular biologist. It comprises a lot of analysis tools, including both the experiments design and data processing without any programming skills for a user.

UGENE provides an easy way to work with DNA, RNA and protein sequences. The functionality list is very wide: sequence annotation with access to remote databases, multiple alignment and phylogenetic trees, 3D protein structures, processing of Sanger and NGS sequencing data (genome assembling and variations, processing of RNA-Seq and ChIP-Seq data), etc. UGENE can be run on MS Windows, Linux and Mac OS X platforms.

Due to its accessibility and wide functionality UGENE can be used as an excellent tool to teach different biological methods.

Results and Conclusion: Currently UGENE is used in tutorials of several universities, for example, Roskilde University (Denmark). “Practical Bioinformatics”, interdisciplinary course utilizing UGENE, starts in September 2016 at NSU. Growth of educational programs using UGENE promotes its popularization within the international community of molecular biologists.

Availability: http://ugene.net/download.html

GENE ONTOLOGY ANALYSIS AND NETWORK RECONSTRUCTION FOR GENES RELATED TO AGING DISEASES AND BEHAVIOR

I.V. Chadaeva*, O.V. Saik, V.N. Babenko
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: ichadaeva@bionet.nsc.ru*

Key words: *gene network, aggressive behavior, aging genes*

Motivation and Aim: Analysis of aging in animal models has shown the connection with neurodegenerative diseases. At the same time, it has a basis neuronal aggressive behavior that accompanies many neurodegenerative diseases. Transciptomics studies on laboratory animals allowed detection of sets of differentially expressed genes related to aggressive behavior [1]. The aim of our study was to find common genes whose expression is associated with the aging process, and the manifestations of aggressiveness.
Methods and Algorithms: Using gene network research programs, Manual Aging, Longevity, GenAge Human, Gene Ontology, Meshops, Coremine and ANDSystem [2], we analyzed a database of aggressive behavior genes and genes associated with aging. List of genes related to behavior diseases was obtained from OMIM database (http://omim.org/). We also used experimental data obtained by transcriptome profiling of aggressive and tame rats at ICG SB RAS [1].
Results: Six lists of genes related to aging as annotated in science literature were used for cross-comparison. It was found that 22 genes, such as AR, BDNF, ESR1, MAOA, POMC and others, were present in both aggressive and aging gene sets. Gene networks of aging and aggression were reconstructed using ANDSystem tool [2] and internet-available software, such as GeneMANIA (http://genemania.org/) and STRING.
Conclusion: The genes found were involved in various biological processes such as neurotransmitter signaling pathways, hormonal regulation, cell transformation and adhesion, etc. Therefore, we conclude that the ambiguous and multidimensional involvement of these genes, which in turn suggests a complicated complex nature of the phenomena of aggressive behavior under consideration.
Acknowledgements: The research is supported by ICG SB RAS project #0324-2015-0003.
References:
1. V.N. Babenko et al. (2016) Analysis of transcriptome data on gene expression in brain areas of rats, selected by aggressive behavior, *In: «Neuroinformatics-2016»*, *MEPhI, Moscow*, P.**2**:82-92. (In Russian).
2. V.A. Ivanisenko et al. (2015) ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Syst Biol*., **9**(Suppl 2):S2.

ANTIOXIDANT RESPONSE ELEMENT CONTROLS LYSOSOMAL BIOGENESIS MASTER-REGULATOR GENES

A.V. Chechushkov*, N.K. Zenkov, E.B. Menshchikova
Research Institute of Experimental and Clinical Medicine, Novosibirsk, Russia
** Corresponding author: avchechushkov@centercem.ru*

Key words: *Antioxidant Response Element, TFE3, TFEB, lysosomal biogenesis*

Motivation and aim: Antioxidant response element (ARE) is a highly evolutionary conserved transcription factor binding site that regulates expression of antioxidant defense genes and many other genes participating in cellular redox homeostasis [1]. Also, ARE was found in autophagy receptors genes and, vice versa, activity of transcription factor Nrf2, that binds ARE, has been shown to be regulated by autophagy [2]. TFEB and TFE3 are transcription factors that regulate lysosomal biogenesis and autophagy genes via CLEAR genomic motif, participating in stimulation of autophagosome formation and autophagy flux[3]. As TFEB and TFE3 participate in redox defense mechanisms, we were interested whether these transcription factors may be interrelated with ARE-binding transcription factors.
Methods and Algorithms: Using RSAT program package we created position frequency matrix for CLEAR and ARE motifs. Then genes of CLEAR-binding transcription factors (TFEB and TFE3) and ARE-binding transcription factors (Nrf1-3, BACH1-2) were searched for ARE or CLEAR motifs. Found motifs were screened for evolution conservation and ChIP-seq data for histone modifications using UCSC genome browser instruments. Real-time PCR was performed to investigate TFEB and TFE3 expression.
Results: No CLEAR motifs were found in ARE-binding transcription factors. Conversely, TFEB and TFE3 both have conservative ARE motifs with high score-similarity between mammals. ARE motif of TFEB gene is located in promoter region as indicated by H3K3me3 markers flanking the region. ARE of TFE3 is located in intron 2 of TFE3 gene and is flanked by H3K27Ac and H3K4m1 indicating intron enhancer region. Real-time PCR shows that ARE-inducing pharmacological agents tert-butylhydroquinone and its synthetic water-soluble analogue TS-13 increase TFEB and to a lesser extent TFE3 gene expression indicating conceivable functional role of found regulatory elements.
Conclusion: Our results may show possible evolutionary conservative control of master-regulator of lysosomal biogenesis genes by antioxidant response element. Such interactions may rely new pharmacological basis for control of TFEB and TFE3 activity, by ARE-inducing pharmacological agents.

References:

1. J.D. Hayes, A.T. Dinkova-Kostova, The Nrf2 regulatory network provides an interface between redox and intermediary metabolism, Trends Biochem. Sci. 39 (2014) 199–218. doi:10.1016/j.tibs.2014.02.002.
2. A. Jain et al., p62/SQSTM1 is a target gene for transcription factor NRF2 and creates a positive feedback loop by inducing antioxidant response element-driven gene transcription., J. Biol. Chem. 285 (2010) 22576–91. doi:10.1074/jbc.M110.118976.
3. M. Sardiello et al., A Gene Network Regulating Lysosomal Biogenesis and Function, Science. 325 (2009) 473–477. doi:10.1126/science.1174447.

SOFTWARE MODULE FOR INTEGRATION OF SBML-WRITTEN MATHEMATICAL MODELS OF MOLECULAR GENETIC SYSTEMS FOR THE HAPLOID EVOLUTIONARY CONSTRUCTOR 3D SOFTWARE PACKAGE

A.D. Chekantsev[1, 2]*, S.A. Lashin[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *Corresponding author: chekantsev@bionet.nsc.ru*

Key words: *Evolution modeling, SBML, software development*

Motivation and Aim: Microbes forming large communities in nature regularly exchange genes via horizontal transfer. It gives microbial cells the ability to acquire novel metabolic functions [1] and consequently may lead to ecological changes in community as a whole. Nowadays, there are a lot of resources for warehousing mathematical models of metabolic [2-4] and gene regulatory [4] systems. It is a great challenge to integrate these data in a complex hierarchical model of evolving microbial community. The Haploid Evolutionary Constructor 3D (HEC 3D) framework allows constructing and simulating such communities consisting of cells of various strains/species/populations living in spatially heterogeneous habitats. Cells consume, utilize, synthesize and secrete metabolites according to genetic programs written modeled as gene networks [5]. Some of cells consume one metabolites and synthesize another, which respectively may be consumed by third cells, i.e. they form trophic cycles of exchanging metabolites.
The aim of this study is development and implementation of a software module for HEC 3D framework in order to import mathematical models of molecular-genetic systems from the existing databases to the HEC.
Methods and Algorithms: We used models written in SBML format. The integration is provided via libSBML and SOSlib libraries. To resolve issues with different synonyms of the same metabolites, we used the REST API for ChEBI database.
Results and discussion: The module designed allows us to extract the parameters and formulas reactions from the model loaded from the repositories such as BioModels or SABIO-RK and to replace existing HEC 3D generalized synthesis strategies with real SBML models. ChEBI database of chemical names was integrated which solved the problem of metabolites wrong usage. Thus, the novel module allows users to use the “real-world” models in the HEC 3D and to investigate the behavior and the evolution of complex microbial communities.
Acknowledgement: The study has been partially supported by the RFBR grant №150703879 and Budget Project 0324-2015-0003.
References:
1. T.Y. Pang, S.S. Maslov (2011) A toolbox model of evolution of metabolic pathways on networks of arbitrary topology, PLOS Comp. Biol., 7: e1001137.
2. U. Witting et al. (2012) SABIO-RK – database for biochemical reaction kinetics, Nucleic Acid Res., 40: D790–D796.
3. S.A. Coskun et al. (2013) An online model composition tool for system biology models, BMC Syst. Biol., 7: 88.
4. F.V. Kazantsev et al. (2016) MAMMOTh: a new database for curated MAthematical Models of bioMOlecular sysTems, bioRxiv, 053652.
5. S.A. Lashin et al. (2014) HEC 2.0: improved simulation of the evolution of prokaryotic communities, In Silico. Biol. 11: 125-135.

A SPATIAL MODEL OF PLANT INTERACTOME AND LONG NON-CODING RNA

M. Chen*
Zhejiang University, Hangzhou, China
** Corresponding author: mchen@zju.edu.cn*

Key words: *plant genomics, long non-coding RNA, gene expression*

Motivation and Aim: The computer analysis of structure of transcripts and the genome organization of crop plants on the basis of integration of high-performance sequencing data is important for the solution of fundamental molecular and biological problems and applied agricultural tasks. The analysis of structure of the genes connected with architecture of an ear in crop plants, morphological parameters of draught- and cold resistance for wheat can be added with the bioinformatics methods and data for model organisms.
Methods and Algorithms: We present here joint research project devoted to development of an integrative computer platform for the analysis of high-performance sequencing data and publicly available computer genomics data. In genetic sense, the mechanisms of drought resistance can be grouped into three categories, drought escape, drought avoidance and drought tolerance. However, crop plants use more than one mechanism at a time to resist drought. Drought escape is defined as the ability of a plant to complete its life cycle before serious soil and plant water deficits develop. This mechanism involves rapid phenological development (early flowering and early maturity), developmental plasticity. Such studies need molecular-genetics approaches and integration of available data. For example, a drought resistance gene, Drt1 in rice, which is linked with genes for plant height, pigmentation, hull color and awn, and has pleiotropic effect on the root system was described.
Results: The task of the work is the analysis of structure of transcripts in plants genomes, search and the description transcripts structures and its location to the protein coding genes. The expected results include development and adaptation of the computer database of potential genes and antis-sense transcripts in model organisms of plants – Arabidopsis, rice, bread wheat.
Conclusion: Integration of data in the project will include data of sequencing technologies RNA-seq and Hi-C (for Arabidopsis). With use of the developed programs and algorithms the general platform of the bioinformatics analysis in plant genomes using high-performance sequencing data will be developed [1-3].
Acknowledgments: The work was supported by RFBR (grant 16-54-53064).

References:

1. X. Meng et al. (2016) Circular RNA: an emerging key player in RNA world, *Brief Bioinform*, pii: bbw045. [Epub ahead of print]
2. J. Wang et al. (2016) Exploring the mechanisms of genome-wide long-range interactions: interpreting chromosome organization, *Brief Funct Genomics*, pii: elv062.
3. L. Yu et al. (2016) miRNA Digger: a comprehensive pipeline for genome-wide novel miRNA mining, *Sci Rep*., **6**:18901.

LONG-TERM SPACEFLIGHT MEDIATED CHANGES IN PROMOTER LANDSCAPE IN ZEBRAFISH TISSUES

A.V. Cherkasov[1]*, K.V. Arshavsky[1], V.N. Sychev[2], M.A. Levinskikh[2], O.A. Gusev[1,3,4]
[1] *Institute of Fundamental Biology and Medicine, Kazan Federal University, Kazan, Russia*
[2] *Institute for Biomedical Problems, Russian Academy of Sciences, Moscow, Russia*
[3] *Division of Genomic Technologies, CLST, RIKEN, Yokohama, Japan*
[4] *Preventive Medicine & Diagnosis Innovation Program, CLST, RIKEN, Yokohama, Japan*
* *Corresponding author: cherkasovav@yandex.ru*

Key words: *CAGE, transcriptional activity, promoter landscape, spaceflight, zebrafish*

Motivation and Aim: Animal models are important to understanding influence of different factors of long-term spaceflights on living organisms and can be helpful for forecasting and prevention negative effects of spaceflights on humans. Some previous experiments, which performed on fish models using simulated microgravitation conditions and exposure aboard International Space Station (ISS), shows significant changes in whole genome gene expression. To define impact of spaceflight to tanscriptional activity on promoter level in Zebrafish tissues, we perform experiments using cap-analysis gene expression (CAGE) methodology.
Methods and Algorithms: Two groups of Zebrafish individuals was used as experimental and control group. The individuals from experimental group were maintained in Aquatic Habitat (AQH) in ISS and fixated in RNA stabilization reagent immediately after arriving and after 36 days of staying aboard. Part of experimental group animals were returned alive from ISS for RNA fixation in ground conditions in two time-points: 2 and 36 days after return. The animals from ground control group were fixated at the same time-points.
Results and Conclusion: First results of CAGE shows significant impact of spaceflight on transcriptional initiation landscape. More than 600 genes are changing their expression in eye samples, after arriving aboard ISS. Notably, the number of differentially expressed genes decreased to 154 after 36 days in space, it can be supposed successful adaptation to spaceflight conditions. From genes, that significantly activated in space, we found several important transcription factors: Fos, FosB and Jdp2, that regulate expression of many genes involved in growth and tissue development processes. Gene Ontology analysis of space-responsive genes shows significant over-representation of functional categories associated with circadian clock system, that confirm influence of microgravitation conditions to regulation of rhythmic processes in animals.
Acknowledgements: The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

SIMULATION OF ENHANCER EVOLUTION IN A COMPUTATIONAL MODEL OF THE DROSOPHILA GAP GENE NETWORK

A.A. Chertkova[1], J. Schiffman[2], K.N. Kozlov[1], M.G. Samsonova[1], S.V. Nuzhdin[1, 2], V.V. Gursky[1, 3]*

[1] *Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*
[2] *University of Southern California, Los Angeles, CA 90089, USA*
[3] *Ioffe Institute, St. Petersburg, Russia*
* *Corresponding author: gursky@math.ioffe.ru*

Key words: *evolution, enhancer, thermodynamic model, gap genes, drosophila*

Motivation and Aim: The evolutionary conservation of enhancers is connected with their regulatory function. We aim to study how the DNA regulatory sequence evolves under the action of selection constraints. We simulate in silico evolution of the gap gene regulatory regions in *Drosophila* and use a model connecting the genotype (regulatory sequence) and a phenotype (gene expression pattern) for this purpose [1].
Methods and Algorithms: We consider a finite-size population of individuals consisting of the putative regulatory regions (enhancers) of four gap genes, and the first generation of this population is set to the wild-type sequence. The individuals are randomly mutated and recombined in the simulation over many generations. We put this population under the negative selection with the survival rate for the progenies defined by a fitness function that does not allow large deviations of gene expression patterns for each individual from the wild-type expression patterns. We estimate the sequence variability at the level of binding affinities of transcription factor binding sites and at the level of gene expression.
Results: We show that the total number of transcription factor binding sites essentially decreases in the course of the evolution. We break the regulatory sequence into non-overlapping segments and calculate a cumulative binding energy of binding sites within each segment. The distribution of this energy across all segments evolves to larger mean values and smaller variance. We show that the binding energies of distinct regulatory segments exhibit extensive mutual correlations, which are specific for each gap gene. Less than 10 percent of binding sites stay in the population during all the evolution time. These core binding sites have higher influence on gene expression. We also demonstrate that even weak binding sites from the local vicinity of the core sites are more conserved, thus elucidating the functional role of this vicinity. Overall, the results clarify the mechanisms of how the binding affinity landscape for enhancers evolves and how this evolution is connected with the functional importance of various parts of the regulatory sequence.
Availability: The software is available from authors upon request.
Acknowledgments: The study was supported by the RSF grant 14-14-00302.
References:
1. K. Kozlov, V. Gursky, I. Kulakovskiy, A. Dymova and M. Samsonova, Analysis of functional importance of binding sites in the Drosophila gap gene network model. BMC Genomics 2015, 16(Suppl 13):S7

FUNCTIONAL ANALYSES ON THE MECHANISM OF INDUCTION OF ANHYDROBIOSIS IN THE MIDGE *POLYPEDILUM VANDERPLANKI*

R. Cornette*, K-I. Iwata, S. Kikuta, Y. Sogame, T. Okuda, T. Kikawada
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: cornette@affrc.go.jp*

Key words: *Anhydrobiosis, Oxidative Stress, Gene expression, Microarray, RNAi*

Motivation and Aim: The sleeping chironomid, *Polypedilum vanderplanki*, is the only insect able to survive almost complete desiccation in an ametabolic state known as anhydrobiosis. Previous studies identified genes involved in anhydrobiosis (1) and comparative genomics showed that those genes were concentrated into specific genomic regions (2). However, little is known about the mechanisms inducing the expression of those genes for successful anhydrobiosis. Here, we focused on the role of oxidative stress and some key genes on the induction of anhydrobiosis.
Methods and Algorithms: P. vanderplanki larvae were subjected to desiccation stress, salt stress and oxidative stress to induce the expression of anhydrobiosis-related genes. Gene expression was investigated by microarray analysis, Real-time PCR and Western blot. The implication of the oxidative stress response and heat shock response transcription factors *Cnc-C* and *Hsf* was investigated by knock-down analysis with RNAi. Plasmid transfection and ectopic expression were also used to characterize a desiccation responsive response element.
Results: Physiological experiments showed that oxidative stress was produced during the induction of anhydrobiosis and general oxidation was damaging proteins, lipids and DNA. However, oxidative stress was also directly responsible for the induction of anhydrobiosis-related protective genes and generated an expression profile similar to that observed during anhydrobiosis. Knock-down by RNAi of the main transcription factor controlling the response to oxidative stress, *Cnc-C*, showed that this gene affects the expression of some anhydrobiosis-related genes and also the survival rate after rehydration of dry larvae. RNAi was also effective on *Hsf*, showing cross-talk between heat shock response and anhydrobiosis. Finally, we identified a desiccation responsive promoter, PDDRE, which induced ectopic expression of GFP in desiccating larvae.
Conclusion: Oxidative stress is probably the main trigger of anhydrobiosis in *P. vanderplanki*. However, its integration operates through a complex cascade including *Cnc-C*, *Hsf*, PDDRE, and certainly numerous still unidentified factors.

References:

1. R. Cornette et al. (2010) Identification of anhydrobiosis-related genes from an expressed sequence tag databasein the cryptobiotic midge Polypedilum vanderplanki (Diptera; Chironomidae), JBC, 285: 35889-35899.
2. O. Gusev et al. (2014) Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge, Nature Communications, 5: 4784.

MODELING OF THE BLOOD FLOW IN THE NARROWED VESSELS

S.G. Davydova[1]*, E.A. Biberdorf[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
** Corresponding author: svetamira_davydova@mail.ru*

Key words: *one-dimensional model of hemodynamics, mathematical modeling, pulse wave velocity, aorta*

Motivation and Aim: A complex mathematical model of the cardiovascular system realized on the BioUML platform was created as a result of the cooperation of Sobolev Institute of Mathematics and Design Technological Institute of Digital Techniques SBRAS. It proved to be very effective in many experiments [1]. This model assumes that all vessels have the cylindrical form. But at the same time a number of large vessels such as the aorta have a conical shape. A transition from the cylindrical vessels to the conic ones allows to significantly improve this model, especially in questions of modeling of a pulse wave. Thus the purpose of this work is to construct an one-dimensional hemodynamics model for the vessel's conical shape and research the distinctions between the vessels of conical and cylindrical forms.
Methods and Algorithms: Theoretical methods of mathematical physics, computational methods - a method of straight lines and orthogonal pro-race.
Results: The system of hemodynamics equations for the conical shaped vessels has been obtained and its program realization in the MATLAB system has been carried out. The velocity of blood flow and the distribution of the pulse wave, the influence of a filtration coefficient on the formation of the reflected wave and the reflection of the pulse wave from the place of the joint of two cylindrical vessels have been investigated. The calculations results lead to the conclusion that using of conic vessels and also the special filtration coefficient on the ends of the terminal vessels allows avoiding of the emergence of «excess» reflections of the pulse wave.
Conclusion: The received model for the vessels of conical shape can be recommended for the improvement of the existing model on the BioUML platform, in particular, for more accurate modeling of the pulse wave profile.
Availability: http://lib.nsu.ru:8081/xmlui/handle/nsu/10420
References:
1. I. N. Kiselev, E.A. Biberdorf, V.I. Baranov, T.G. Komlyagina, V.N. Melnikov, I.Yu. Suvorova, S.G. Krivoshchekov, F.A. Kolpakov, «Personalization of Parameters and Validation of Model of the Human Cardiovascular System», Mat. Biolog. Bioinform., Vol. 10, № 2, 2015, P. 526-547.

STRUCTURAL PATTERNS AMONG THE DIVERSITY OF FLAVIN-DEPENDENT OXIDOREDUCTASES FROM LUMINOUS BACTERIA AND *E. COLI*

A.A. Deeva*, E.A. Temlyakova, A.A. Sorokin, E.V. Nemtseva, V.A. Kratasyuk
Siberian Federal University, Krasnoyarsk, Russia
Institute of Cell Biophysics RAS, Pushchino, Russia
** Corresponding author: adeeva@sfu-kras.ru*

Key words: *flavin-dependent oxidoreductase; bacterial luciferase; luminous bacteria; structural bioinformatics*

Motivation and Aim: In luminous bacteria flavin-dependent oxidoreductases play an important role in supplying unstable reduced flavin for a bioluminescencent reaction catalyzed by bacterial luciferase. Currently the problem of flavin transfer between these two enzymes is under consideration due to a diversity of oxidoreductases capable either to form a complex with luciferase [1], or to transfer flavin via free diffusion [2]. The aim of our study was to reveal probable structural similarities of various oxidoreductases with known amino acid sequences.
Methods and Algorithms: The protein-protein BLAST algorithm was used to search the NCBI database for homologs of starting sequences. Phylogenetic analysis was performed using MAFFT and PhyML software. SWISS-MODEL server was used to reconstruct unknown structures. FMN and NAD(P) binding sites were mapped by using the NCBI Conserved Domains web server. Surface electrostatic potential was described with Apbs software.
Results: Structural homologs belong to relative oxidoreductase families and the majority of conserved amino acid positions are located in proximity within FMN and NAD(P) binding sites. Electrostatic potential distribution on molecular recognition surface of the reductases show similar patterns in vicinity of the hypothetical interface involved into interaction with luciferase. A few oxidoreductases from *E. coli* have a high amino-acid similarity with corresponding oxidoreductases from luminous bacteria.
Conclusion: Structural and physical peculiarities inherent for different oxidoreductase families determine possible interaction mode with bacterial luciferase.
Acknowledgements: The reported study was funded by RFBR according to the research project No. 16-34-00746 mol_a and by the grant No.1762 from The Ministry of Education and Science of the Russian Federation.
References:
1. S. Tu (2008) Activity coupling and complex formation between bacterial luciferase and flavin reductases, *Photochem. Photobiol. Sci.* **7**: 183–88.
2. R. Tinikul et al. (2013) The transfer of reduced flavin mononucleotide from LuxG oxidoreductase to luciferase occurs via free diffusion, *Biochemistry* 52: 6834-43.

ANHYDROBIOSIS RELATED PROMOTERS IN PV11 CELL LINE

R.M. Deviatiiarov[1]*, T. Kikawada[2], R. Cornette[2], O.A. Gusev[1, 2, 3]
[1] *Institute of Fundamental Medicine and Biology KFU, Kazan, Russia*
[2] *National Institute of Agrobiological Sciences NIAS, Tsukuba, Japan*
[3] *Center of Life Science Technologies RIKEN, Yokohama, Japan*
** Corresponding author: ruselusalbus@gmail.com*

Key words: *Anhydrobiosis, promoters, CAGE, regulation of gene expression*

Motivation and Aim: Dry preservation of tissues, cells and genetic material supposed to be a more preferable approach for long term storage in biobanks than lyophilization and cryopreservation. Pv11 cell line, which is able to survive dehydration [1], is a prospective model for this purpose. The major components are already known (trehalose and LEA proteins), but the regulatory systems still stay unclear. Therefore, we have two sides of interest – fundamental, to reveal the effect of anhydrobiosis on promoters' organization, together with practical significance for the further vectors construction, as example. In the beginning, discovery of promoter regions using CAGE could give an insight about possible transcription factors and general features, like GC content, shape, and active genes, which are described in the current research.
Methods and Algorithms: CAGE libraries were performed in three replicates for six stages of desiccation/rehydration of Pv11 cell line and sequenced on HiSeq 2500 (high output mode, 50-bp single end). R packages "edgeR", "CAGEr" were applied for differential expression analysis, peaks clustering, plotting and other objects. The MEME Suite and AgriGO online tools for motifs and gene ontology tags enrichment, respectively.
Results: CAGE revealed 9218 active promoters and 8139 active coding genes in the cell culture. If we consider 200 bp area around TSS, the average GC% is 26.9, which is lower than in genome in general (28%), and for 50 bp regions – about 30.3%. Among gene-associated promoters ~30% classified as peaked and the rest have wide type distribution of CAGE signal. These peaked promoters have significantly ($p<2.2e-16$) higher GC content. DE analysis with expression profiles clustering uncovered the desiccation specific group of genes (n=1202), including 40 of 68 active cryptogenes. GO enrichment showed terms about homeostasis, like GO:0019725, GO:0045454 and other, confirming importance of selected genes. While GC distribution in promoters for this group has no essential bias, the shape is much wider, and motifs enrichment analysis showed particular traits, as specific positions around +1 and +30 bp from TSS, including stress factors binding sites, like ABRE.
Conclusion: We defined active and desiccation specific genes in Pv11, described the basic properties of their promoters. Here we found that wide shape not associated with high GC content, suggesting a lack of conventional CpG islands.
Acknowledgements: This work was supported by Russian Science Foundation, grant №14-44-00022

References:

1. Y. Nakahara et al. (2010) Cells from an anhydrobiotic chironomid survive almost complete desiccation, *Cryobiol*ogy, **60**, 138-146.

GENE EXPRESSION PROFILING OF FLAX (*LINUM USITATISSIMUM* L.) UNDER EDAPHIC STRESS

A.A. Dmitriev[1], A.V. Kudryavtseva[1], N.V. Koroban[1], G.S. Krasnov[1], A.S. Speranskaya[1, 2], A.A. Krinitsina[2], M.S. Belenikin[1, 2], A.V. Snezhkina[1], A.F. Sadritdinova[1], O.Yu. Yurkevich[1], N.V. Kishlyan[3], T.A. Rozhmina[3], O.V. Muravenko[1], N.L. Bolsheva[1], N.V. Melnikova[1]*

[1] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*
[2] *Lomonosov Moscow State University, Faculty of Biology, Moscow, Russia*
[3] *All-Russian Research Institute for Flax, Torzhok, Russia*
** Corresponding author: mnv-4529264@yandex.ru*

Key words: *Linum usitatissimum, flax, edaphic stress, expression alterations, nutrients*

Motivation and Aim: Flax (*Linum usitatissimum* L.) is an important crop, which is widely used in textile, food, pharmaceutical and chemical industries. Edaphic stresses result in crop losses and decrease of flax oil and fiber production. In the present work, the impact of nutrient stress on flax plants has been studied.

Methods and Algorithms: Flax plants of line Stormont cirrus and cultivar Bethune were grown under normal, deficient and excess nutrition. RNA was extracted from upper leaves and cDNA libraries were constructed. Next generation sequencing (NGS) of cDNA libraries was performed on Illumina platform. SOAPdenovo assembler was used for the assembly of flax transcriptome. For quantitative PCR (qPCR) data analysis, $\Delta\Delta C_t$ method and 2 reference genes (*ETIF3H* and *ETIF3E*) were used [1]. Statistical analysis was performed using Kruskal-Wallis and Mann-Whitney tests.

Results: Consequently, 69M, 27M, and 24M of 100 bp reads were obtained for flax plants grown under phosphate deficient (P), normal (N), and nutrient excess (NPK) conditions respectively. In total, 43471 transcripts were identified and their expression levels were assessed. 14 differentially expressed genes were chosen on the basis of obtained NGS data and their expression was evaluated on extended sampling (10 flax plants grown under N, 11 – under P, and 10 – under NPK conditions) using qPCR. The expression alterations under imbalanced nutrients were revealed for genes encoding WRKY DNA-binding protein family and JAZ (jasmonate-zim-domain) protein family.

Conclusion: WRKY protein family regulates numerous processes in plants, including abiotic stress response [2]. Proteins of JAZ family are jasmonate signalling repressors, which control plant growth and development [3]. The role of WRKY and JAZ encoding genes in flax response to imbalanced nutrition was shown for the first time. These data provide new insights into edaphic stress response of flax.

Acknowledgements: This work was financially supported by the Russian Science Foundation (grant 16-16-00114).

References:

1. N.V.Melnikova, A.A.Dmitriev, et al. (2016) Identification, expression analysis, and target prediction of flax genotroph microRNAs under normal and nutrient stress conditions, *Frontiers in Plant Science*, **7:** 399.
2. A.Banerjee, A.Roychoudhury (2015) WRKY proteins: signaling and regulation of expression during abiotic stress responses, *The Scientific World Journal,* **2015:** 807560.
3. L.Pauwels, A.Goossens (2011) The JAZ proteins: a crucial interface in the jasmonate signaling cascade, *The Plant Cell,* **23(9):** 3089-3100.

GENETICS AND PHYSIOLOGY OF WHEAT INFLORESCENCE DEVELOPMENT

O.B. Dobrovolskaya[1,5]*, P. Martinek[2], Yu.L. Orlov[1], A.A. Krasnikov[3], E.D. Badaeva[4], K.I. Popova[5], J. Salse[6], N. Watanabe[7]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Agrotest Fyto, Ltd, Kroměříž, Czech Republic*
[3] *Central Siberian Botanical Garden SB RAS, Novosibirsk, Russia*
[4] *Vavilov Institute of General Genetics RAS, Moscow, Russia*
[5] *Novosibirsk State Agrarian University, Novosibirsk, Russia*
[6] *INRA-UBP UMR-1095, Clermont – Ferrand, France*
[7] *College of Agriculture, Ibaraki University, Ibaraki, Japan*
** Corresponding author:oxanad@bionet.nsc.ru*

Key words: *plant architecture, inflorescence development, mutants, SEM, molecular genetic mapping, synteny, map based cloning*

Motivation and Aim Wheat is one of the most important food crops in the world. The solution of practical problems of modern plant breeding and the establishment of new technological groundwork is impossible without the new fundamental knowledge and comprehensive information on the structural and functional organization of the plant genome and without a deep understanding of the molecular-genetic mechanisms underlying agronomic traits. Such traits include qualitative and quantitative characteristics of wheat inflorescence (spike) directly related to yield. The aim of this project is genetic dissection of wheat inflorescence morphology and development, structural and functional characterization of genes involving in the control of inflorescence development using a collection of wheat mutants with altered spike morphology.
Methods and Algorithms Our strategy integrates analysis of inflorescence development of wheat mutants using light microscopy and SEM, molecular genetic mapping, a candidate gene approach, and map-based cloning.
Results Here we report the structural and functional characterization of the *WFZP* gene as one of the key regulators of wheat inflorescence development, driving supernumerary spikelets that provide new resources for the genetic manipulation of grain number; and molecular-genetic mapping of QTLs for the multifloret trait that influences number of grains, one of the major yield components.
Acknowledgements This work is supported by grants from Russian Foundation for Basic Research 15-04-05371 and 16-54-53064.

THE MANIFESTATION AND PHYTOHORMONE RESPONCE OF LEAF PUBESCENCE GENES IN BREAD WHEAT

A.V. Doroshkov*, A.V. Simonov, D.A. Afonnikov, T.A. Pshenichnikova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author:ad@bionet.nsc.ru*

Key words: *wheat leaf, trichomes, leaf hairiess genes, image analysis*

Motivation and Aim: The leaves of many angiosperm species develop trichomes. In dicots they have been exploited for the study of the determination of cell fate, plant cell differentiation mechanisms and cell morphogenesis. This trait is known to make a significant contribution to the protection from pests and adaptation to environmental factors in bread wheat. However the genetic basis of wheat trichome formation is poorly understood although a wide variation was found among *Triticeae* species with different ploidy level. Currently Catalogue of Gene Symbols for wheat contains only two loci associated with this trait: the gene *Hl1* in 4B chromosome and the gene $Hl2^{aesp}$ in 7B chromosome. Molecular function and regulation of these genes are currently not known.The present research sought to establish the individual and joint effect on trichome patterning and growth of each of three known wheat leaf pubescence genes (*Hl1, Hl3* and $Hl2^{aesp}$) and effect of phytohormone treatment on their phenotypic expression.
Methods and Algorithms: Various lines carrying *Hl1, Hl3* and $Hl2^{aesp}$ and specially created nearly isogenic lines were used to quantitatively compare leaf pubescence using a high throughput phenotyping method (wheatdb.org/lhdetect2). This method allows to obtain rapidly quantitative characteristics of leaf pubescence (length of individual trichomes and their number) among many plants.
Results and conclusion: These genes differed in their effect on trichome formation. *Hl1* and *Hl3* more affected trichome initiation and growth, while $Hl2^{aesp}$ modified trichome length. Their action was independent to a large extent. A model of the action and interaction of *Hl1, Hl3* and $Hl2^{aesp}$ has been proposed to explain the genetic basis of trichome length and number. The effects of phytohormones on trichome cell growth and initiation while *Hl1, Hl3* and $Hl2^{aesp}$ genes manifestation were explored. The effects of auxin (IAA), gibberellic acid (GA3), cytokinins (6-BAP, Kinetin), metyl jasmonate (MeJa), ethylene (ACC) have been investigated and described. Our data revealed a key role of cytokinin signaling pathway in *Hl1* and *Hl2* gene manifestation.
Acknowledgements: This work was supported by Russian Science Foundation (RSCF) grant № 14-14-00734.

A SOFTWARE SYSTEM FOR SIMULATING SOCIAL AND GENETIC ASPECTS OF DEAFNESS IN HUMAN POPULATIONS

I.S. Dyachenko[1, 2]*, O.L. Posukh[1, 2], M.S. Bady-Khoo[1], M.V. Zytsar[1, 2], V.Yu. Mikhalskaia[1, 2], G.P. Romanov[3], N.A. Barashkov[3, 4], Yu.G. Matushkin[1, 2], S.A. Lashin[1, 2]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Institute of Natural Sciences, North-Eastern Federal University, Yakutsk, Russia*

[4] *Yakut Scientific Center of Complex Medical Problems, Yakutsk, Russia*

** Corresponding author: i.s.djachenko@gmail.com*

Key words: *Human genetics, modeling, populations, hearing loss*

Motivation and Aim: Hearing loss (HL) caused by environmental or genetic factors, affects a large part of human population worldwide. Genetic factors account for at least half of all cases of profound congenital deafness. Hereditary HL could be caused by several dozen genes and characterized by various inheritance modes: autosomal dominant (~ 20%) or recessive (~ 75-80%), and X-linked or mitochondrial (~ 2-3%). The most common form of recessive HL is due by mutations in gene *GJB2*. Prevalence of genetic HL is often determined by specific population genetic structure. Nance et al. [1] suggested that obvious accumulation of *GJB2*-caused deafness could be also influenced by intense assortative marriages among deaf people combined with their improved biological adaptation (genetic fitness). Study of socio-demographic and genetic characteristics of deaf people' communities is important for predicting the prevalence of hereditary deafness and understanding a role of social factors in evolutionary processes in human populations.

Methods and Algorithms: We combined multilayer population model [2] with models of assortative mating [3] taking into account genetic and social structure of population. Our model represents the spreading of deafness-causing mutations across the population considering deaf people's tendency to assortative ("deaf x deaf") marriages.

Results: We have developed prototype of program model based on the agent-based simulation framework "Diploid evolutionary constructor" [2] for deafness prevalence. Basic objects of this model are 'an individual' (agent) and 'population structure'. Phenotype strategy computes agent's phenotype by his genotype. It allows changing phenotype expression mechanism without modifying whole model structure. Partner choice strategy defines the choice of deaf individual for marriage partner, taking into account both common for all humans and specific for current problem factors.

Conclusion: Developed software package will allow investigating the trends in prevalence of genetic deafness considering socio-demographic, molecular genetic and population genetic parameters.

Acknowledgements: This study was partly supported by the RFBR grant #15-04-04860_a.

References:

1. W.E. Nance et al. (2000) Relation between choice of partner and high frequency of connexin-26 deafness, *Lancet*, 356 (9228): 500-501.
2. S.A. Lashin, Yu.G. Matushkin. (2013) Multi-layer computer models of gene network evolution in diploid populations, *FEBS Journal*, 280 (Suppl. 1): 563.
3. S. Karlin. (1972) Some mathematical models of population genetics, *Amer. Math. Monthly*, 79: 699-739.

SOME ASPECTS OF MOLECULAR EVOLUTION AND RECOMBINATIONAL VARIABILITY OF THE ZIKA VIRUS

YP. Dzhioev[1, 2]*, A.I. Paramonov[2], Y.S. Bukin[3, 4], I.V. Kozlova[2], V.I. Zlobin[1]
[1] *Irkutsk State Medical University, Institute of Biomedical Technology, Irkutsk, Russia*
[2] *Scientific Center of family health and human reproduction problems, Irkutsk, Russia*
[3] *Limnological Institute SB RAS, Irkutsk, Russia*
[4] *Irkutsk National Research Technical University, Irkutsk, Russia*
* *Corresponding author:alanir07@mail.ru*

Key words: *Zika virus, molecular evolution, recombination*

Motivation and goal. Zika virus is a flavivirus of the Spondevi serotype. Its hosts are mosquitoes of the genus Aedes and monkeys, but people are infected occasionally. ZIKV genome consists of a single-stranded positive sense RNA molecule with 10794 kb of length with two flanking noncoding regions. The aim was to reconstruct the evolution of molecular variability and recombination of the virus Zika using methods of bioinformatics and phylogenetics.
Methods and algorithms. Genomes 30 strains 1 and Zika virus genome Spondveni from GenBank data base were used in this research. Sequences were aligned with MAFFT program on the Galaxy web-based platform. Phylogenetic trees were constructed by Neighbor-joining and Maximum Composite Likehood methods of MEGA6 software package. Spondveni virus genome was used as the outgroup. Statistics were obtained by bootstrap with 10,000 repetitions. Evaluation of selection pressure was carried out by SLAC. Recombination breakpoints were calculated by the means of RDP v4.61 software package. Violations of topology were visualized by using NeighborNet SplitsTree 4.14.1 program and further confirmed statistically by PhiTest.
Results. Analysis of selection pressure through the number of synonymous and nonsynonymous substitutions has shown that many sites in the genome of the virus Zika are under strong negative selection pressure, but there were no sites under positive selection. This pattern is typical for highly adapted genotypes with low polymorphism in functionally important genes. Accumulation of synonymous mutations may be associated with the change of arthropods and vertebrates as hosts. Phylogenetically Zika virus strains are divided into three clusters: African, Asian and American. Strains of the first two clusters are closely related and were identified during the recent virus outbreaks. African cluster includes relatively old samples and is further divided into several subclusters. However low statistical support for some nodes of the phylogenetic tree within the cluster suggested non-linear nature of evolution at least in some sequences of the Zika virus. Phylogenetic networks contain several large loops, indicating the inability to resolve the phylogenetic relationships between strains in the form of a phylogenetic tree. PhiTest with high confidence ($p < 0.0001$) had shown presence of recombination sites in the genomes of African cluster strains. RDP programs with high reliability ($p < 0.01$) revealed 18 unique sites of recombination. Furthermore, some genomes contained more than one recombination point. Thus, despite the high stabilizing selection pressure, Zika virus can have a high potential for adaptation.

THE DISTANCE MATRIX BOOTSTRAPPING IN THE CASE OF QUANTITATIVE TRAITS

V.M. Efimov[1–4]*, K.V. Efimov[5], V.Y. Kovaleva[2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Systematics and Ecology of Animals SB RAS, Novosibirsk, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
[4] *Tomsk State University, Tomsk, Russia*
[5] *Moscow Institute of Physics and Technology (State University), Moscow, Russia*
** Corresponding author:vmefimov@ngs.ru*

Key words: *nucleotide sequences, object–character matrix, principal coordinate analysis*

Motivation and Aim: Felsenstein suggested bootstrapping for "object-character" matrix referring to both the nucleotide sequences and quantitative traits (namely, measurements of the skull). He didn't take into account the problem of correlation between quantitative traits deliberately, unlike he saw the problem: "A more serious difficulty is lack of independence of the evolutionary processes in different characters. If the characters are correlated (as measurement characters often are), then, in effect, we have fewer characters in the study than we believe. ... For the purposes of this paper, we will ignore these correlations and assume that they cause no problems; in practice, they pose the most serious challenge to the use of bootstrap methods." [1]. This opinion was supported by Efron et al. [2]. Nowadays bootstrap resampling is strongly recommended for any trait matrices and without any justification: "This data factory applies to data matrices only. After bootstrapping the data, a new distance matrix is computed." [3]. But see [4]. We propose new approach for Euclidean distance matrix (EDM) bootstrapping based on one distances only.
Methods: The basic idea of our approach is initially to transform the EDM into the principal component (coordinate) matrix (PCM). Then we add a column Y of unique object identificators to obtained matrix. Looping through components, we form a pair Z=(X, Y), where X – a succeeding component, and sort Z on X. On i-th bootstrap iteration of Z by rows we receive Zi=(Xi, Yi), sort it by Xi, replace Yi by Y, sort Zi by Y, and combine all components Xi for the given i in i-th bootstrap copy of PCM – PCMi. At last we compute Euclidean distances between rows of PCMi and obtain EDMi – the i-th bootstrap copy of initial EDM. If initial distance matrix isn't EDM the nonmetric multidimension scaling (NMDS) can be used.
Results: The EDMi is Euclidean matrix because it is computed by coordinate matrix.
Conclusion: The new approach for Euclidean distance matrix bootstrapping based on one distances only is proposed.
Acknowledgements: This work has been supported by Budget Project (grant #0324-2015-0003).
References:
1. Felsenstein J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 39(4). 783-791.
2. Efron B., Halloran E., Holmes S. (1996). Bootstrap confidence levels for phylogenetic trees. PNAS. 93(23), 13429-13429.
3. Pouseele H., Vauterin P., Vauterin L. (2011). A resampling strategy for reliable network construction. Molecular phylogenetics and evolution. 60(3). 273-286.
4. Lin Y., Rajan V., Moret B.M. (2012). Bootstrapping phylogenies inferred from rearrangement data. Algorithms for Molecular Biology. 7(1). 1

METABOLITE PROFILING OF THE MOSS *PHYSCOMITRELLA PATENS* INOCULATED WITH *PSEUDOMONAS*

E.D. Egorova, N.A. Baraeva, S.V. Vinogradova*
Research Center of Biotechnology RAS, Moscow, Russia
** Corresponding author: svetlana.vinogradova@biengi.ac.ru*

Key words: *Physcomitrella patens, Pseudomonas, moss, plant-pathogen interaction, metabolome*

Motivation and Aim: Mosses are ancient, nonvascular, seedless plants, belonging to the bryophytes. *Physcomitrella patens* is a new model organism for various studies in system biology, plant physiology, plant-microbe interactions. The interactions between necrotrophic pathogens such as *Botrytis*, *Pythium*, *Pectobacterium* and moss *P. patens* were previously reported. However, reports about the ability of specialized bacterial pathogens to colonize *P. patens* are absent. The aim of this work is study of *P. patens* and *Pseudomonas* interactions via metabolome profiling.
Methods and Algorithms: *P. patens* was cultivated on Knop's medium and inoculated with suspensions of *P. fluorescens*, *P. syringae* and *P. viridiflava* (OD 0.4). Two and five post inoculation with *Pseudomonas* gametophores were washed off and grinded to determine number of bacteria colonies in moss. These suspensions were serially diluted and spread evenly on Petri dishes. Metabolites were isolated and identified by gas chromatography–mass spectrometry (GC-MS).
Results: We found 75 metabolites with concentration higher or lower compare with the control. Were detected functional groups of carboxylic and fatty acids, sugars, sterols and polyhydric alcohols. The involvement of components of the carbohydrate metabolism such as hexadecanoic acid, phthalic acid, propanedioic acid, butanoic acid and octadecanoic acid was previously shown. In addition, the sugar alcohols or polyols, such as hexitol, pentitol, threitol and myo-inositol were found in many gametophore samples. Their specific role in osmoprotection has been reported for many seed plants.
Conclusion: We found that *Pseudomonas* bacteria induced the accumulation of compounds required for bacterial growth and involved in plant's systemic resistance. Thus, the obtained data confirmed that *P. fluorescens*, *P. syringae* and *P. viridiflava* are the pathogens of *P. patens.*
Acknowledgements: The reported study was funded by grant of President of Russian Federation MK -7138.2015.4 and by RFBR according to the research project №13-04-40104.

DYNAMIC RECOGNITION OF 8-OXOGUANINE BY DIFFERENT PROTEIN FOLDS

A.V. Endutkin[1, 2]*, C. Simmerling[3], D.O. Zharkov[1, 2]
[1] *SB RAS Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Stony Brook University, Stony Brook, NY, USA*
* *Corresponding author:Aend@niboch.nsc.ru*

Key words: *8-Oxoguanine, DNA repair, protein–DNA interactions, enzyme kinetics, molecular dynamics*

Motivation and Aim: 8-Oxoguanine, a mutagenic DNA lesion generated under oxidative stress, differs from its precursor guanine by only two substitutions. 8-Oxoguanine DNA glycosylases of different structural superfamilies can locate and remove 8-oxoguanine through extrusion and excision. To date, how these enzymes efficiently distinguish 8-oxoguanine from a large excess of undamaged DNA bases remains unclear.
Methods and Algorithms: Here, we use molecular dynamics simulations to explore the mechanism by which 8-oxoguanine DNA glycosylases from *E. coli* (Fpg) and human (OGG1) discriminate between 8-oxoguanine and guanine. Using biochemical experiments some key points of the simulations have been confirmed.
Results: We have found that discrimination between the oxidative DNA lesion and its normal counterpart, by DNA glycosylases likely involves multiple gates. Interestingly, we found that OGG1 and Fpg share remarkable similarities in their dynamic damage recognition mechanisms despite lack of structural homology. Our simulations, supported by biochemical analysis, suggest that OGG1 distinguishes between 8-oxoguanine and guanine using their chemical dissimilarities not only at the active site, but also at earlier stages during base eversion, and this mechanism is at least partially conserved in Fpg.
Conclusion: The similarity suggests that lesion recognition through multiple gating steps may be a common theme in DNA repair. Our results provide new insight into how enzymes can exploit kinetics to probe the chemical modifications present in DNA lesions and efficiently recognize damage.
This work was supported by Russian Science Foundation (14-24-00093). A.E. was supported by a Russian Presidential Fellowship.

AGEING OF MULTICELLULAR ORGANISMS AS A STAGE OF ONTOGENESIS

I.L. Erokhin
National Biotechnological Company LLC, Moscow, Russia
** Corresponding author: i.erokhin@inbox.ru*

Key words: *ageing, ontogenesis, junk DNA*

As a rough approximation, the multicellular organism genome is structured (in terms of the graph theory) in the form of an oriented binary tree, with tops corresponding to homogeneous logical elements "iteration step of the genome tree", and arcs corresponding to transfer of management between the "steps". Each cell of a multicellular organism corresponds to its "step". It receives management from the "step" of the mother cell, initiates realization of the "cell program" that determines its development and division, returns management upon termination of the cell program realization and transfers management to the "steps" of daughter cells. Many steps of the genome tree (GT) cycle create the largest by volume structure of the multicellular organism genome, its junk DNA. The "cell program" is a network where the regulatory, as well as structural genes, are engaged in certain order.

This genome model is based on the hypothesis in accordance with which the management between the "steps" is transferred with the help of short RNAs. Being the product of expression of non-coding interspersed repeats, these short RNAs which make up a part of nucleoproteins formed with general factors of polymerase transcription Pol III, can initiate transcription of other repeats. The cell program iteration step may be represented as a pair of regulatory genes with identical complex promotors, and GT step – with three regulatory genes. The process of organism ageing is determined by the structure of extreme GT branches.

1. Termination of GT branches may be represented as a chain. One arc of each step transfers management to the next step, while the second arc does so to the step that initiates apoptosis. One daughter cell replaces the mother cell, while the second one is dismantled. If the final step of the chain refers to the apoptosis program, the death of the organism occurs without signs of ageing. If the final step of the chain refers to postmitotic program, radicals and double-stranded breaks in DNA will be accumulated. The life span of such an organism shall be determined by the length of chains at the terminal GT branches.
2. All the second last GT steps may be partially looped to themselves. One arc transfers management to the step from which it originates, the second arc – to the step that initiates apoptosis. One daughter cell replaces the mother cell, while the second one is dismantled. All the cells of such an organism shall be timely renewed, and the organism shall not age. If the last step transfers management to the postmitotic program, the cells under its management may be separated and specialized as blood cells or form a gradually removed level like skin cells.
3. One of GT branches forms an organ that emits the cell division factor into the blood. The terminal steps of this branch are structured in the form of a chain. Final steps of the remaining GT branches partially looped. If the organ produces sufficient amount of the factor, all the cells accepts it, are divided and constantly renewed. At the end of the chain, the organ cells are no longer renewed, and the amount of the factor is reduced. Not all the cells of the organism receive the factor, and they stop renewing as well. This process aggravates with the time, and the organism is subjected to ageing. The ageing model and the accompanying hypotheses belong to the author.

OPISTHORCHIIDAE TRIAD: COMPARATIVE GENOMICS OF THE CARCINOGENIC LIVER FLUKES USING A DRAFT GENOME OF *OPISTHORCHIS FELINEUS*

N. Ershov[1]*, G. Fan[2, 3], E. Prokhortchouk[4], V. Solovyev[5], D. Afonnikov[1, 6], H. Yang[2], V. Mordvinov[1], X. Liu[2], K. Skryabin[4, 7] and The Opisthorchis Genome Consortium
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *BGI-Shenzhen, Shenzhen, China*
[3] *State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences, University of Macau, Macao, China*
[4] *Russian Federal Research Center for Biotechnology, Moscow, Russia*
[5] *Softberry Inc., Mount Kisco, NY, US*
[6] *Novosibirsk State University, Novosibirsk, Russia*
[7] *National Research Centre, Kurchatov Institute, Moscow, Russia*
* *Corresponding author: nikotinmail@mail.ru*

Key words: *liver flukes, O. felineus, genome assembly, gene expression, molecular evolution*

Motivation and Aim: The liver flukes *Opisthorchis felineus*, *O. viverrini*, and *Clonorchis sinensis* are the three epidemiologically significant species of the family Opisthorchiidae (class Trematoda, phylum Platyhelminthes), affecting the human and animal liver and bile ducts. The International Agency for Research on Cancer classified the liver flukes *O. viverrini* and *C. sinensis* as group 1 agents, i.e., the agents carcinogenic to humans, and as the major factors inducing cholangiocarcinoma in endemic regions. Despite clinical significance of these species they are poorly investigated and little is known about their functional genomics and molecular mechanisms underlying their host-parasite interactions. Here we provide a draft annotated genome and updated transcriptome of *O. felineus* – the third major Opisthorchiidae species – which is a main cause of opisthorchiasis in nothern Eurasia.
*Methods and Algorithms:*Genomic DNA libraries of *O. felineus* (adult worm; paired-end libraries) and mRNA-seq libraries (cDNA libraries) were obtained according to the manufacturer's instructions (Illumina, USA). Preprocessed raw sequencing data were used for genome assembly and transcriptome analysis. We used gene prediction for *O. felineus* genome based on combination of ab initio gene finding and additional data on trematode transcripts/protein sequences.
Results: The genomic sequence of *O. felineus* was obtained. Gene models of *O. felineus* genome were reconstructed. We also assessed evolutionary conservativeness of various aspects of the genome and transcriptome structure between the Opisthorchiidae species.
Conclusion: Taken together, these data allowed us to get a more precise picture of the parasite's genome organization which can shed light onto the evolution of Opisthorchiidae species and assist their further functional genomics research.
Acknowledgements: The study supported in part by the RSF 14-24-00123 grant.

ELUCIDATION OF MOLECULAR SIGNAL OF TRANSCRIPTION RESPONSE TO DESICCATION STRESS IN CHIRONOMID *P. VANDERPLANKI*

E.I. Shagimardanova[1]*, R.M. Deviatiyarov[1], T. Kikawada[2], O.A. Gusev[1, 3]
[1] *Kazan Federal University, Kazan, Russia*
[2] *National Institute of Agrobiological Sciences, Tsukuba, Japan*
[3] *RIKEN, Yokohama, Japan*
* *Corresponding author: rjuka@mail.com*

Key words: *anhydrobiosis, stress response, cell culture, trehalose, transcriptome*

Motivation and Aim: Pv11 is the cell culture derived from embryonic stem cells of anhydrobiotic insect *Polypedilum vanderplanki*. This cell line as well as insect itself is able to survive complete desiccation. Protocol development of cell culture maintenance and survival after water loss have shown that for successful anhydrobiosis induction the preculture with 300 mM trehalose or sucrose is necessary. To reveal mechanisms to stress adaptation and identify correct molecular signal of response we performed whole genome gene expression analysis in cell culture after different stress effect.
Methods and Algorithms: We performed mRNA sequencing of Pv11 cell culture subjected to desiccation stress, oxidative stress and salt stress using HiSeq 2500 Illumina platform. Reads were mapped on *Polypedilum vanderplanki* genome available at http://bertone.nises-f.affrc.go.jp/midgebase.
Results: We compared expression level of "cryptogenes" in response to different kinds of stress. It was shown that in comparison to other stresses, the most drastic changes were observed during desiccation. Moreover, genes typically induced by water loss were highly upregulated by adding trehalose or sucrose to culture media. In case of several genes, such as Lea or Tret-transporter, expression level was higher during sugar preincubation that desiccation itself.
Conclusion: Whole transcriptome analysis showed that preincubation with trehalose and incubation with sucrose induce expression level of the majority of desiccation stress response genes. It suggests that trehalose molecule triggers transcription response to water loss in *P. vanderplanki* midge.
Acknowledgements: The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University and supported by federal program (ID RFMEFI58414X0002).

STRUCTURAL BASIS FOR THE RECOGNITION AND PROCESSING OF DNA CONTAINING BULKY LESIONS BY THE MAMMALIAN NUCLEOTIDE EXCISION REPAIR SYSTEM

A. Evdokimov[1]*, A. Popov[1], I. Petruseva[1], O. Lavrik[1, 2]
Institute of chemical biology and fundamental medicine, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: an_evdokimov@mail.ru*

Key words: *DNA repair, NER, artificial DNA substrates*

Motivation and Aim: Mammalian NER eliminates the broadest diversity of bulky lesions from DNA with wide specificity. At that the double incision efficiency for structurally different adducts can vary over several orders of magnitude. Therefore, great attention is drawn to the question of the relationship among structural properties of bulky DNA lesions and the rate of damage elimination. The synthetic DNA structures (model DNA) which imitate NER intermediates and substrates, e.g. double-stranded DNA bearing an appropriate modification are widely used instruments of NER investigations *in vitro*. Our present work concerns the properties of several structurally diverse model DNAs containing bulky modifications. According existing concepts, it was expected, that the values of melting temperature decrease, bending angles and K_D values clearly define the model DNAs substrate properties, but the experimentally estimated levels of the substrate properties were far away from these expectations.
Results and Conclusion: Present work studies the properties of several structurally diverse model DNAs containing bulky modifications. Model DNAs have been designed using modified nucleosides (exo-*N*-{2-*N*-[*N*-(4-azido-2,5-difluoro-3-chloropyridin-6-yl)-3-aminopropionyl]aminoethyl}-2'-deoxycytidine (Fap-dC) and 5-{1-[6-(5(6)-fluoresceinylcarbomoyl)hexanoyl]-3-aminoallyl}-2'-deoxyuridine (Flu-dU)) and the nonnucleosidic reagents *N*-[6-(9-antracenylcarbomoyl)hexanoyl]-3-amino-1,2-propandiol (nAnt) and *N*-[6-(5(6)-fluoresceinylcarbomoyl)hexanoyl]-3-amino-1,2-propandiol (nFlu). The substrate properties of model DNAs under investigation were estimated. The impact of artificial lesions on spatial organization and stability of the model DNA was evaluated. Their affinity for the damage sensor XPC was also studied. It was expected, that the values of melting temperature decrease, bending angles and K_D values clearly define the row of model DNAs substrate properties as: nFlu-DNA≈Flu-dU-DNA>>nAnt≈Fap-dC-DNA. The experimentally estimated levels of the substrate properties were actually in the row: nAnt-DNA≥ nFlu-DNA>> Flu-dU-DNA >> Fap-dC-DNA. Molecular dynamics simulations have revealed structural and energetic basement of the discrepancies observed. A several lesion-specific regions of DNA secondary structure stabilization and destabilization were found, and their possible impacts on efficiencies of DNA damage recognition and subsequent excision was suggested.
This work was supported by the Russian Science Foundation [142400038 to O.L.].

PRINCIPAL ORGANIZATION OF PHYSIOLOGICAL REGULATOR

V.I. Fedorov
Institute of Laser Physics SB RAS, Novosibirsk, Russia
** Corresponding author: vif41@mail.ru*

Key words: *systems biology, physiological system, hierarchical control, cybernetical biology*

Hierarchical division of physiological system. Each complex physiological process can be decomposed into elementary processes. Each elementary physiological process can be decomposed into individual elementary acts. Each elementary act is regulated by a special control system. I call such control system for regulation of an elementary act as the elementary control unit. Thus, individual elementary physiological process is the set of elementary acts. Therefore, the set of elementary control units is the control system of an individual physiological process. I call such set as the functional ensemble. The set of functional ensembles controls physiological process as a whole. I call such set as the functional system. Hence for the management of a functional process each time the subset of elementary control units from the entire set of units and the subset of functional ensembles is generated dynamically. Various elementary processes can be combined from the same elementary acts in various combinations. Combination of individual functional processes is complex process. There are several types of relationships between functional ensembles for the control of complex processes.

Regulation of elementary control unit status. The status of elementary control unit is controlled by activating and inhibiting signals. Each of them affects the specific receptor. Signals for each input have various intensity at each time point and act simultaneously. Input elements can be active or inactive. Only those receptors react that are in an active state in a given time. Their activity has a different degree. The reaction depends on the extent of receptor activity in a given time. The receptor activity is defined by its conformation in a given time. The degree of acceptance of signal for each input depends on the steric affinity of receptor to input signal in a given time. The quantitative expression of receptor affinity will called the weight of input. Weight of input varies over time depending on the internal and external tuning. The weight of input causes a partial contribution to change of elementary control unit state. Tthe combination of the number of active inputs (activating and inhibiting) expressed as the intensity of input signals multiplied by the weight of inputs uniquely determines the state of elementary control unit at a given time.

Types of interaction between functional ensembles. 1. Antagonism. It is used for maintain a predetermined level of the process. 2. Synergism. It is used for transfer the process to a new level of functioning. 3. Crossing. It is used for regulation of cyclic processes and multicomponent process transition to a new level of functioning (process of development, attenuation, or extinction of process. 4. Balancing. It is used for correction of level of ensemble activity for coordination of parallel occurring processes. 5. Cascade. It is used when one functional ensemble starts the development of process and activates a new ensemble. The functioning of the organism is done through a combination of individual functional processes. Each process is controlled by a separate functional ensemble. The combination of functional assembles, down to control the integrative function of the organism, called the functional system. The consideration of principal organization of physiological regulator will illustrate by organization of nervous control.

PHYLOGENY DEVELOPED OVER THE TRIPLET COMPOSITION OF MITOCHONDRIAL GENOMES: HIGH SYNCHRONY IN THE EVOLUTION OF TWO GENETIC SYSTEMS

V.S. Fedotova[1]*, M.G. Sadovsky[2], Yu.A. Putintseva[1, 3]
[1] *Siberian Federal University, Krasnoyarsk, Russia*
[2] *Institute of Computational Modeling, Krasnoyarsk, Russia*
[3] *Forest Institute, Krasnoyarsk, Russia*
* *Corresponding author:vasilinyushechka@gmail.com*

Key words: *population genomics, cluster, k-means, coevolution, elastic map*

Motivation and Aim: Genomics is an up-to-date field of molecular biology studying the genome of the organism in its entirety. Nowadays, a new direction in Genomics that is Population Genomics is rapidly developing. One of the main problem in Population Genomics is to identify the features of genomes that manifest at the population level.
In this study, we present some preliminary results obtained in this new direction. Rather nontrivial genetic object, namely mitochondrial genomes of different species, have been chosen for a study. Previously, it was reported that there is fundamental connection between the genome structure of organells and taxonomy of the bearers. Here we explore the problem and expand the results obtained earlier.
Methods and Algorithms: In our study, a population diversity has been studied by genomic methods: we examine the mitochondrial genome clusterization of different organism groups, in the space of triplet frequencies. All mitochondrial genomes were taken from EMBL-bank and were used to build up the database (http://www.ebi.ac.uk/ena).
Source database of mitochondrial genomes contained variety of species in different genuses. It included 3726 entries. Variety numbers of species in genuses caused bias in results. Unbiased dataset was created with uniform distribution of species in genuses and it includes 2990 genomes of different organisms.
We used *K*-means to cluster the dataset. The clusterisation was conducted in ViDaExpert program. As usual, *K*-means classifies a part genomes stably (in a series of the runs of k-means), while others show unstable classification [1]. The construction of layered graph is based on *K*-means: the classes developed for specific *K* yield a layer.
Results: The high synchrony in the evolution of two genetic systems manifesting in nonrandom distribution of taxonomic categories in over the vertices of layered graph has been found. As *K* grows up, the species comprising the classes redistribute among them in highly nonrandom pattern. The stability of clusterisation has been found.
Conclusion: High synchrony of classic phylogeny and clusterisation developed over the triplet composition of mitochondrial genomes has been founded. Synchrony has also been founded for the group of species in a low taxonomic genera – genuse.
References:
1 Sadovsky M. G., Putintseva Yu. A., Fedotova V. S., Chernyshova A. I. (2015) Genome structure of organells strongly relates to taxonomy of bearers. *LNCS* vol. 9044, Part 2: 481-490

INTEGRATION OF TRANSCRIPTOMIC AND PROTEOMIC DATA TO ELUCIDATE THE MECHANISM OF ACTION OF NOVEL COMPOUNDS: THE CASE OF THE ANTITUMOR PEPTIDE CIGB552

J.R.F. Massó[1]*, T.N. de Villavicencio-Díaz[1], Y.R. Gómez[1], B.O. Argüelles[2], A.R. Ulloa[1], Y.C. García[3], O.G. Cruz[1], Y.P. Riverol[1], L.J. González[1], I. TiscorniaI[4], S. Victoria[4], M.B. Fogolín[4], V.B. Pérez[1], L.D. Roche[5], D.P. Gardon[1], I.G. Perez[1], D.V. Blomquist[1], L.N. Perez[1], Y.G. Rodriguez[2], M.G. Vallespi[2]

[1] *Division of System Biology, CIGB, Havana, Cuba*

[2] *Division of Pharmaceuticals, CIGB, Havana, Cuba*

[3] *Department of Preclinical Studies, National Institute of Oncology and Radiobiology,Cuba*

[4] *Cell Biology Unit, Institut Pasteur of Montevideo, Uruguay*

[5] *Center of Studies for Research and Biological Evaluations, Pharmacy and Food Sciences College, University of Havana, Havana, Cuba*

[*] *Corresponding author e-mail: julio.fernandez@cigb.edu.cu*

Key words: *antitumoral peptide, transcriptomics, proteomics, oxidative stress, network analysis*

Motivation and Aim: CIGB-552 is a novel molecule with antineoplastic and cell-penetrating capacity in several tumor cell lines. Systemic injection of immunocompetent and nude mice with established solid tumor resulted in regression of tumor mass and apoptosis. The molecular target and the mechanism of action were unknown [1].

Methods: Microarray experiments and proteomic approaches were used to identify the molecular target and the set of pathways regulated in tumor cells following treatment with the CIGB-522. Data integration was conducted using Bisogenet plugin [2].

Results: We identified a set of 349 genes differentially expressed when compared treated versus untreated cells using oligonucleotide microarray. In addition the nuclear subproteome of HT-29 colon adenocarcinoma cells treated with CIGB-552 peptide was identified and analyzed. Pathway analysis enrichment using bioinformatic tools reveals NF-kB signaling and oxidative stress as relevant pathways in the antitumor activity of CIGB-552. In addition, using proteomic and network analysis tools COMMD1 protein was identified as a target of CIGB-552 peptide. The relevance of COMMD1 as the molecular target of the CIGB-552 was validated using shRNAi.

Conclusion: COMMD1 protein was identified as the molecular target of the antitumoral peptide CIGB-552. NF-kB signaling and oxidative stress modulation by CIGB-552 contributes with its cytotoxic effect.

References:

1. Vallespi, M.G. *et al.* Identification of a novel antitumor peptide based on the screening of an Ala-library derived from the LALF(32-51) region. *Journal of peptide science : an official publication of the European Peptide Society* **16**, 40-47 (2010).
2. Martin, A. *et al.* BisoGenet: a new tool for gene network building, visualization and analysis. *BMC bioinformatics* **11**, 91 (2010).

THE SPATIAL MAP OF AVIAN GENOME

V. Fishman[1, 2]*, N. Battulin[1, 2], A. Maslova[3], O. Serov[1, 2], A. Krasikova[3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Saint-Petersburg State University, St. Petersburg, Russia*
** Corresponding author: minja-f@ya.ru*

Key words: *Gallus gallus, 3-dimentional genome architecture, Hi-C*

Motivation and Aim: Eukaryotic gene expression is subjected to precisely coordinated multi-layer controls, across the levels of epigenetic, transcriptional and post-transcriptional regulations. Recently developed high-throughput genomic methods of mapping higher-order chromatin structure and chromatin–nuclear matrix showed that 3-dimentional DNA architecture makes an important contribution to transcription regulation by triggering physical interactions of enhancer with their target genes [1,2]. Moreover, recently appeared evidences of significant role of 3-dimential DNA architecture during evolution of vertebrate. However, our knowledge of animals genome architecture is mainly restricted to a relatively small sample of mammals and a model insect *D. Melanogaster*. This inspired us to investigate a spatial map of first avian genome, *Gallus gallus,* using a high-resolution *Hi-C.*
Methods and Algorithms: Hi-C libraries were generated using *in situ* Hi-C protocol [1]. Data was aligned to an existing genome assembly (*Gallus gallus v5.0*) and iteratively corrected [3]. LACHESIS software and homemade algorithms were used to validate and improve existing genome assembly.
Results: We examined spatial contacts of DNA in chicken embryonic fibroblasts and adult red blood cells using Hi-C. A current chicken genome assembly was updated using obtained data. Genome-wide contact maps show a typical plaid-pattern described earlier in mammals and insects [1,2,3]. We identified topologically associated domains (TADs) and A/B compartments both in fibroblasts and in red blood cells genomes and show a similar localization of domain borders in these cell types. As in mammals, distribution of TAD borders reflects chromatin organization and genes expression. We also found several specific features of chicken genome that were not observed in mammals, such as increased number of interchromosomal contacts of microchromosomes comparing to interchromosomal contacts of macrochromosomes.
Conclusion: The first spatial map of avian genome shows that principles of genome architecture are similar in different vertebrates.
References:

1 Rao et al (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, 159(7):1665–1680
2 Dixon et al (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539): 331–336
3 Imakaev et al (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods*, 9(10):999-1003.

A NEW ALGORITHM TO THE RECONSTRUCTION OF A SET OF POINTS FROM THE MULTISET OF N^2 PAIRWISE DISTANCES IN N^2 STEPS FOR THE DE NOVO SEQUENCING PROBLEM

E.S. Fomin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author:fomin@bionet.nsc.ru*

Key words: *Algorithm optimization, amino acid sequences, cyclic peptides*

The problem of amino acid sequence reconstruction in cyclic peptides is particularly important, because such peptides are of special significance in a cell. They include antibiotics, antitumor agents, toxins, immunomodulatory agents, and a variety of peptides with still unknown properties. The sequences of many cyclic peptides cannot be reconstructed from the DNA sequence, because they are not encoded by genes and their synthesis occurs in a nonribosomal way [1].

De novo sequencing, i.e., the direct reconstruction of the primary sequences of peptide chains from mass spectra, has been in use since the early 2000s [2]. Its major advantage is that it can be applied when no genomic information is available. The problem of sequencing linear and cyclic peptides is reduced to the known turnpike and beltway problems [3], the latter of which having no polynomial-time algorithm in the general case.

Many algorithms were proposed for de novo sequencing, see the review of [4]. Despite the enormous efforts made in recent years, the turnpike and beltway problems are still considered open. A new simple approach to reducing overhead costs for the problems, based on the sequential removal of redundant information from inputs, is proposed. For error-free inputs that simulate de novo sequencing spectra with high accuracy, up to 10^{-3} Da, the size of inputs decreases drastically, from n^2 to O(n), which permits one to eliminate exhaustive search from the algorithm almost completely and reconstruct sequences in numbers of steps in direct ratio to the input size, n^2. Computational experiments show high efficiency of the algorithm for both the turnpike and beltway cases, with the reconstruction time for sequences of lengths up to several thousand elements being within one second on a modern PC.

References:

1. M.A. Marahiel et al. (1993) Regulation of peptide antibiotic production in Bacillus, Mol Microbiol, 7(5): 631-636.
2. R.S. Johnson, J.A.Taylor. (2002) Searching sequence databases via de novo peptide sequencing by tandem mass spectrometry, Mol Biotechnol, 22: 301-315.
3. P. Lemke et al. (2003) Reconstructing Sets From Interpoint Distances, Discrete and Computational Geometry Algorithms and Combinatorics, 25: 597-631.
4. J. Allmer (2011) Algorithms for the de novo sequencing of peptides from tandem mass spectra, Expert Review of Proteomics, 8(5): 645-657.

MOLECULAR MODELING OF THE INTERACTION BETWEEN INDOLE LUPANE DERIVATIVES AND C-MYC/MAX HETERODIMER

T.S. Frolova[1, 2, 3]*, D.S. Baev[2], A.V. Petrova[4], E.K. Khusnutdinova[4], O.I. Sinitsyna[1, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk Institute of Organic Chemistry SB RAS, Novosibirsk, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
[4] *Bashkir State University, Ufa, Russia*
** Corresponding author: frolova@bionet.nsc.ru*

Key words: *betulonic acid derivatives, selective cytotoxicity, molecular docking, c-Myc/Max interaction*

Motivation and aim: Pentacyclic triterpene acids are one of the most interesting compounds because of their various biological properties, for example, antioxidative, antiviral, antiallergenic, antiangiogenic and antispasmodic activity [1]. One of the most promising triterpenes in plant raw material are lupane acids, which are very popular for chemical modifications to enhance their biological activity, especially selective cytotoxicity against cancer cells. According to the literature, one the possible mechanisms of this effect is the inhibition of c-Myc/Max interaction [2].
Methods and Algorithms: Previously we synthesized four new lupane derivatives with indole functional group, all tested compounds have selective cytotoxic properties. Presently we assessed binding energy between c-Myc/Max heterodimer and indole lupane derivatives using molecular docking.
Results: It was shown that all tested compounds probably could interact with binding site of c-Myc and prevent the association of c-Myc/Max heterodimer which results in cell death. *Conclusion:* Considering that *c-Myc* gene is overexpressed in many cancers, its directed inhibition may be used in antineoplastic therapy.
References:

1. A. Bishayee et al. (2011) Triterpenoids as potential agents for the chemoprevention and therapy of breast cancer, *Front Bios*ci, 16: 980–96.
2. H. Wang et al. (2015) Direct inhibition of c-Myc-Max heterodimers by celastrol and celastrol-inspired triterpenoids, Oncotarget, 32:32380-95.

ALTERED CATECHOLAMINERGIC, SEROTONERGIC, GABAERGICS, AND GLUTAMATERGIC GENES EXPRESSION IN THE VENTRAL TEGMENTAL AREA OF MALE MICE UNDER CHRONIC SOCIAL DEFEAT STRESS: RNA-SEQ DATA

A.G. Galyamina, I.L. Kovalenko, D.A. Smagin, N.N. Kudryavtseva*
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: n.n.kudryavtseva@gmail.com*

Key words: *RNA-Seq, ventral tegmental area, chronic social defeat stress*

Motivations and aims: Chronic social defeat stress (CSDS) leads to the development of mixed anxiety/depression-like state in male mice similar to those in humans. It has been shown that, under CSDS, the adult brain undergoes numerous changes in the function of neurotransmitter systems, including changes in gene expression in different brain regions. In this experiment we are focused on the analysis of genes encoding proteins related with the metabolism and receptors of serotonergic, catecholaminergic, GABAergic and glutamatergic systems in the ventral tegmental area which is important in the regulation of cognition, motivation, drug addiction, and emotions relating to several psychiatric disorders.
Methods: Mixed anxiety/depression-like state was generated in male mice by exposure to CSDS in daily agonistic interactions [1, 2]. The ventral tegmental area was dissected according to the map presented in the Allen Mouse Brain Atlas. The collected samples were sequenced at JSC Genoanalytica (http://genoanalytica.ru/, Moscow, Russia). The Cufflinks program was used to estimate the gene expression levels in FPKM and then to detect the differentially expressed genes in the analyzed and control groups. Genes were considered to be differentially expressed at the level of statistical significance $p < 0.05$ and $q < 0.05$.
Results: We found that genes, related with serotonin (Maob, Htr4, Htr1a) were upregulated but expression of Htr3a gene was decreased in the ventral tegmental area of depressive mice. Besides, upregulation of dopaminergic Ddc, Slc6a3, Drd2, and Maob genes were upregulated while noradrenergic Slc6a2, Adra2c, and Adra2a genes were downregulated. Expression of GABAergic Gabra1, Gabra2, Gabrg2, Gabrg1, Gabrq, Gad1, and Gad genes and glutamatergic Gria1, Gria2, Grik2, Grm2, Grm5, and Slc17a8 genes were increased under CSDS.
Conclusion. These results are evidences of altered noradrenergic, dopaminergic, serotonergic, GABAergic, and glutamatergic neurotransmissions in the ventral tegmental area of chronically defeated mice.
Acknowledgments. Supported by Russian Science Foundation (No 14-15-00063); Program of Presidium RAS “Molecular and Cellular Biology” (No 6.25).
References:
1. Kudryavtseva N.N. et al. (1991) Social model of depression in mice of C57BL/6J strain. Pharmacol., Biochem. Behav. 38: 315-320.
2. Avgustinovich D.F. et al. (2004) Dynamic changes of brain serotonergic and dopaminergic activities during development of anxious depression: Experimental Study. Usp. Fiziol. Nauk. 35(4):19-40.

RECONSTRUCTION OF TRANSCRIPTION CONTROL NETWORK IN GENOME-REDUCED BACTERIA BY HIGH-THROUGHPUT PROMOTERS IDENTIFICATION

I.A. Garanina*, G.U. Fisunov, D.V. Evsutina, V.M. Govorun
Scientific Research Institute of Physical-Chemical Medicine SRI PCM, Moscow, Russia
** Corresponding author: irinagaranina24@gmail.com*

Key words: *mollicutes, genome reduction, transcription, promoters*

Motivation and Aim: Class Mollicutes is a specialized clade of gram-positive bacteria that lack cell wall and feature significant genome reduction. Genome reduction resulted in the loss of most of the conserved transcriptional regulators. Attempts to elucidate transcriptional regulation in Mollicutes using high-throughput technologies resulted in a limited progress. In current work we aimed to investigate transcription regulation and reveal most important transcription factors for Mollicutes.
Methods and Algorithms: We utilized comparative genomics methods and algorithms to analyze conservation of regulators throughout Mollicutes phylogeny. For regulation network reconstruction we prepared and sequenced RNA 5`-ERS libraries of 4 samples of every bacteria analyzed. To identify transcription start sites we searched for a local maximum in the read coverage that was supported by at least 5 reads. Then, we modeled the coverage at each local maximum while considering 5 nt in each direction as background using a GLM with a quasi-binomial distribution.
Results: We described all transcription factors and DNA-binding proteins of 47 Mollicutes and found that there is only one common function requiring transcriptional control in all species: protein folding maintenance via chaperones. We carried out whole-genome mapping of transcription start sites of *Acholeplasma laidlawii* and *Spiroplasma melliferum* and compared promoter structure between three bacteria. *A. laidlawii* demonstrate the most organized promoter structure, whether *M. gallisepticum* show some extent of promoter degeneration. Based on comparative genomics approach we predicted binding sites of transcription factors and probable targets of these regulators. Most of the transcription factors with identified binding sites in the three species are involved in the metabolism control. Progressive loss of the conserved TFs from *A. laidlawii* to *S. melliferum* and *M. gallisepticum* is associated with drastic loss of metabolic pathways.
Conclusion: We suggests that only one regulatory component is conserved between analyzed bacteria. This fact reflects high plasticity and adaptive potential of Mollicutes. Our study thus provides insight into evolution and organization of transcriptional regulation of genome-reduced bacteria.

GENOME-WIDE TRANSCRIPTOMICS AS A PLATFORM FOR UNDERSTANDING THE UNUSUAL RESISTANCE TO MUSCLE ATROPHY IN HIBERNATING DORMICE

G.R. Gazizova[1]*, O.V. Tyapkina[2], O.S. Kozlova[1], M.D. Logacheva[1, 3], L.F. Nurullin[2], I.M. Vikhlyantsev[4], O.A. Gusev[1, 5]

[1] *Kazan Federal University, Kazan, Russia*

[2] *Kazan Institute of Biochemistry and Biophysics KSC RAS, Kazan, Russia*

[3] *Lomonosov Moscow State University, Moscow, Russia*

[4] *Institute of Theoretical and Experimental Biophysics RAS, Puschino, Russia*

[5] *RIKEN, Yokohama, Japan*

** Corresponding author: grgazizova@gmail.com*

Key words: *hibernation, immobilization, muscle atrophy, transcriptomics, dormice*

Motivation and Aim: Muscle atrophy during prolonged periods of immobilization or microgravity remains an unsolved problem in public health, physiology and space medicine. At the same time, some animals are able to experience long periods of hypokinesia without any significant changes in muscle structure during hibernation. Thus, hibernating mammals are convenient objects for studying the mechanisms that prevent atrophy or trigger skeletal muscle recovery. The aim of present research is to identify molecular mechanisms of protective muscular adaptation via analysis of whole-transcriptome sequencing of muscle and lumbar spinal cord samples of hibernator edible dormice (*Glis glis;* Rodentia).

Methods and Algorithms: We performed mRNA and total RNA sequencing of dormice' muscles and spinal cord using HiSeq 2500 Illumina platform. Four groups of animals (n=12) were examined: 1) active animals, 2) animals immobilized in special fixing cages in laboratory, to estimate the effect of general muscle disuse 3) hibernated animals and 4) animals just aroused after 14 days of hibernation. For the reference, *de novo* transcriptome was assembled to yield 30000 + primary contigs.

Results: To test hypothesis that dormice are able to resist muscle atrophy independently of type of limb immobilization, we conducted morphometric and molecular analysis of skeletal muscle. Based on sequencing data, profiles of differential gene expression in different types of muscles and lumbar spinal cord were determined. Enriched by differentially expressed genes molecular signaling pathways were identified. Status of expression of sarcomeric and muscle-specific proteins was analyzed. Remarkably, associated with muscle atrophy process in mammalian muscles E3-ubiquitin ligases MuRF1 and MAFbx revealed no changes in mRNA expression both in response to hibernation and induced hypokinesia in dormice.

Conclusion: Obviously, edible dormice have distinct regulatory mechanisms preventing muscle atrophy during immobilization. Whole-genome analysis of gene expression has allowed determining transcriptional program in response to different types of muscle disuse. Further studies, including genome sequencing, would lead a deeper understanding of the strategies of hibernators to survive extreme environmental conditions.

Acknowledgements: The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University and supported by RFBR JSPS_a №14-04- 92116.

SYSTEMIC RESPONSE TO GENETIC AND CHEMICAL MODULATION OF DDR REGULATING WILD TYPE P53-INDUCED PHOSPHATASE IN SKIN, INTESTINE AND HEMATOPOIETIC SYSTEM

A.R. Goloudina[2], B.B. Grigorash[1], E.Y. Kochetkova[1], E. Appella[3], V.A. Pospelov[1], O.N. Demidov[1,2]*

[1] *Institute of Cytology RAS, St. Petersburg, Russia*
[2] *University of Burgundy, France*
[3] *NCI, NIH, Bethesda, USA*
* *Corresponding author: demol13@yahoo.com*

Key words: *cell death, DNA damage, aging, senescence, tumor suppressor, genetic models*

Motivation and Aim: Wild type p53-induced phosphatase is one of the key regulators of DNA damage response in cancer cells. It is also involved in regulation of various processes including cell death and senescence. The aim of our study was to analyze the expression patterns of wild type p53-induced phosphatase gene, Ppm1d, in normal epithelial tissues and hematopoietic system and consequences of modulation of its expression or activity.
Methods and Algorithms: The Ppm1d reporter and conditional deletion animal models were created to study Ppm1d expression and Ppm1d- deficient phenotypes. GSK2830371 was used to study effects of chemical modulation of Ppm1d activity. Ppm1d expression in tissues was studied with RNA-FISH techniques and staining for LacZ activity as Ppm1d reporter.
Results: Ppm1d expression was observed in skin, intestinal epithelium and hematopoietic cells at mRNA and protein levels. In skin and intestine Ppm1d expressed mainly in the compartment of epithelial progenitor cells. In hematopoietic system Ppm1d expression was an important factor in regulation of differentiation at the different stages in various hematopoietic lineages. The genetic deletion of Ppm1d gene sensitized all tissues to DNA –damage inducing agents such as UVB or chemotherapeutic agents. It is also increased spontaneous cell death and senescent phenotype with age in mentioned tissues and affected normal aging in comparison with wild type animals. Another consequence of Ppm1d deletion was more severe pro-inflammatory response to UVB and other agents. The phenotypical changes after Ppm1d deletion correlated with changes in gene expression profiles. Chemical or genetic modulation of Ppm1d activity activated anti-tumor defense and led to tumor suppression in several animal models of human cancer.
Conclusion: Ppm1d was confirmed to be an important regulator of DNA damage response, cell death, aging, inflammation and tumorigenesis in epithelial and hematopoietic tissues. The development of new therapeutic strategies based on modulation of Ppm1d activity would help to improve the treatment of several types of human pathologies.

PREDICTING OF THERMODYNAMIC DATA OF MORPHOLINO ANALOGOUS OF NA BY COMPUTER APPROACH AND COMPARING WITH EXPERIMENTS

V.M. Golyshev*, A.A. Lomzov
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
** Corresponding author: golyshevvictor@gmail.com*

Key words: *molecular dynamics (MD), novel analogous of NA, thermodynamics of binding and cooperativity*

Motivation and aim: Development of approaches for reliable calculations of physic-chemical properties of novel analogous of nucleic acids (NA) is a big challenge. Recently, the ability to calculate enthalpy of binding ΔH^0 for native oligonucleotides (ON) with DNA was shown. The applicability of proposed method for NA derivatives is under dissuasion. A new morholino derivative [1] is an attractive object for verification of the predictive capacity. In this work the hybridization enthalpy of native and morpholino derivative oligonucleotide complexes were calculated using MD simulation and determined experimentally.

Methods and Algorithms: The structures of the two isomers of morpolino adenine was minimized using QM calculation at HF/6-311+G(d,p) in Gaussian'09 and particular atoms charges were calculated using RESP method. Based on this data the library files for MD simulation were prepares. MD simulations of native and the morpholino derivative (two isomer and one isomer mixed chain) oligonucleotide complexes were performed using AMBER12 software. The 1 μs trajectories were obtained in the explicit solvent in periodic condition (TIP3P water model, 12Å cuboid box) and NPT ensemble (1 bar, 300K). MD data we analyzed using MMGBSA calculation.

Thermal stability of pentamers of adenine and morhpholino derivative of adenine with oligothymidines (15, 20, 25 and ~300 base) was studied by thermal denaturation with optical registration of signal method. The data analyses were performed with originally developed method which includes simultaneous fitting of three different complex denaturation curves.

Results: The values of enthalpy changes of binding (ΔH^0 in kcal/mol) and cooperative interaction (ΔH^0_k in kcal/mol) obtained by MD were $\Delta H^0 = 28.3\pm0.2$, $\Delta H^0_k = 11.1\pm0.4$ for native and $\Delta H^0 = 25.3\pm0.2$, $\Delta H^0_k = 14.0\pm0.2$ for morpholino complexes . Corresponding values, obtained by experimental techniques were $\Delta H^0 = 37.7\pm2.1$, $\Delta H^0_k = 27.2\pm3.2$ for deoxyriboadenosine and $\Delta H^0 = 22.6\pm1.1$, $\Delta H^0_k = 15.5\pm0.7$ morpholino derivative complexes.

Conclusion: We have shown that thermodynamics parameters of binding and oligonucleotide cooperative interaction in tandem duplexes can be determined by proposed using experimental and computer simulation techniques. Thermodynamics data obtained with MD and thermal denaturation approaches are perfectly matched. Thus the MD simulations are compatible for prediction of hybridization enthalpy of analogous of NA.

References:

1. Abramova, T. V., et al. (2012). Synthesis and properties of nucleic acid methylene carboxamide mimics derived from morpholine nucleosides. Russian Journal of Bioorganic Chemistry. 38(4) 400-411.

TUMOR-SPECIFIC CELL FREE DNA AS A BIOMARKER OF METASTASIS

T.M. Gorbacheva[1]*, S.A. Solodskikh[1], V.Yu. Bashmakov[1], V.Yu. Panevina[1], A.Y. Maslov[2], V.N. Popov[1]
[1] *Voronezh State University, Voronezh, Russia*
[2] *Albert Einstein College of Medicine of Yeshiva, USA*
* *Corresponding author: gorb.tanya@gmail.com*

Key words: *RNA-seq, aggressive behavior, differential gene expression, brain, laboratory animals, rat*

Motivation and Aim: The development of novel techniques for the evaluation of blood biomarkers (liquid biopsies) in cancer such as circulating tumor cells (CTCs) and circulating tumor DNA (ctDNA) or tumor-specific cell free DNA (cfDNA) have changed the translational medicine as a non-invasive biomarker. Circulating biomarkers might represent both primary tumor and metastatic deposits and provide ways for investigating metastatic processes.
Methods and Algorithms: The present study investigated the levels of patient-specific mutations in circulating cell-free DNA (cfDNA) in plasma from patients with clear-cell renal cell carcinoma. Pairs of normal-tumor tissues were collected and sequenced, analyzed to establish the mutation profile for each patient based on which the primers were designed.
Results: Allele-specific PCR were conducted to monitor mutations in cfDNA in plasma of different points – cfDNA from blood taken in 1, 3 and 6 months after the operation. It was observed that there is a tendency of increasing the level of mutant alleles in plasma in patients with metastasis.

DNA DAMAGE INITIATING DEMETHYLATION: A REPAIR–EPIGENETIC CONNECTION

I.R. Grin[1, 2]*, A.A. Ishchenko[3]
[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Gustave Roussy Cancer Campus, Villejuif, France*
** Corresponding author: grin@niboch.nsc.ru*

Key words: *DNA damage, DNA repair, BER, MMR, epigenetic demethylation*

Motivation and Aim: Active DNA demethylation (ADDM) in mammals occurs via hydroxylation of 5-methylcytosine (5mC) by TET and/or deamination by AID/APOBEC family enzymes. The resulting 5mC derivatives are removed through the base excision repair (BER) pathway (1). At present, it is unclear how the cell manages to eliminate closely spaced 5mC residues while avoiding generation of double-strand breaks and whether alternative DNA repair pathways participate in ADDM (2). Recently, was proposed, that cytidine deaminase induced damage can lead to indirect ADDM through the BER or non-canonical DNA mismatch repair (ncMMR) (3).
Methods and Algorithms: We used a phagemid DNA containing oxidative base lesions and methylated sites are used to examine the involvement of various DNA repair pathways in ADDM in murine and human cell-free extracts.
Results: We demonstrate that, in addition to short-patch BER, 5-hydroxymethyluracil and uracil mispaired with guanine can be processed by ncMMR and long-patch BER with concomitant removal of distant 5mC residues. Furthermore, the presence of multiple mispairs in the same MMR nick/mismatch recognition region together with BER-mediated nick formation promotes proficient ncMMR resulting in the reactivation of an epigenetically silenced reporter gene in murine cells.
Conclusion: These findings suggest cooperation between BER and ncMMR in the removal of multiple mismatches that might occur in mammalian cells during ADDM.
This work was supported by Russian Science Foundation 14-24-00093; Fondation ARC PDF20101202141.

References:

1. S.Moréra, I.Grin, A.Vigouroux, S.Couvé, V.Henriot, M.Saparbaev, A.A.Ishchenko (2012) Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyuracil-containing DNA, *Nucleic Acids Res.*, **40(19):** 9917-9926.
2. D.O.Zharkov, I.R.Grin (2012) Mechanisms of somatic mutagenesis and active DNA demethylation in epigenetic regulation, In: *Epigenetics*, S.M.Zakian, V.V.Vlasov, E.V.Dementieva, 31-52 (Novosibirsk: Publishing House of SB RAS).
3. R.D.Kasymov, I.R.Grin, A.V.Endutkin, S.L.Smirnov, A.A.Ishchenko, M.K.Saparbaev, D.O.Zharkov (2013) Excision of 8-oxoguanine from methylated CpG dinucleotides by human 8-oxoguanine DNA glycosylase, *FEBS Lett.*, **587(18):** 3129-3134.

CONSERVATION LEVEL OF THE KEY MEIOTIC PROTEINS REFLECTS THEIR FUNCTION AND INDEPENDENT EVOLUTION IN DIFFERENT LINEAGES OF EUKARYOTES

T.M. Grishaeva*, Yu.F. Bogdanov
Vavilov Institute of General Genetics RAS, Moscow, Russia
** Corresponding author: grishaeva@vigg.ru*

Key words: *proteins, amino acid sequence, functional domains, recombination, cohesins, synaptonemal complex, conservation, evolution*

Motivation and Aim: Meiosis is conserved in terms of its cytological appearance in the evolutionarily row of Eukaryota. Meiosis requires proteins participating in (1) recombination and (2) chromosome remodeling and synapsis. Among them are meiosis-specific proteins and proteins shared with mitosis. According to a widespread opinion, group 1 proteins are more conserved than group 2 proteins, mainly due to similarity of their functional domains. Published estimates are rarely presented in a quantitative form. We analyzed the amino acid sequences of some key meiotic proteins from both of the groups in model eukaryotic species and quantitatively estimated the level of their evolutionary conservation.
Methods and Algorithms: Protein Blast was used to perform pair-wise amino acid sequence comparisons for orthologs of seven proteins in eight eukaryotic species representing fungi, plants, invertebrates and vertebrates. Because the similarity index Score depends on the protein size, percent similarity was estimated for each protein, taking its similarity for itself as 100%.
Results: The recombination mediators RAD51 and DMC1 and the mismatch repair protein MLH1 were shown to be moderately conserved (from 13% to 76% similarity between proteins from different phyla), while meiosis-specific endonuclease SPO11 had only a low similarity in the tested proteomes (5-32%). Among structural chromosomal proteins, cohesins RAD21 and REC8 demonstrated low conservation (0-25% and 0-5%, respectively). Synaptonemal complex components possessing the HORMA domain (Hop1 and its orthologs) displayed only a minor similarity between the taxa (2-14%). In a separate study, the level of conservation was estimated for functional domains of some proteins because they are thought to be more conserved than the full-size protein molecules. Unexpected results were obtained. The longer (TOPRIM_topoIIB_SPO) and short (TPGA_N) domains of SPO11 showed the same or even lower conservation as compared with the full-size SPO11 (14-38% and 1-36%, respectively). The HORMA domain was only two to three times more conserved than the full-size protein.
Conclusion: Proteins responsible for the fidelity of meiotic recombination (RAD51, DMC1, and MLH1) are the most conserved among key meiotic proteins, while cohesins, synaptonemal complex proteins, and endonuclease SPO11 are less conserved. Meiotic proteins evolved independently in different phylogenetic lineages of Eukaryota.
Acknowledgements: This work was supported by the Russian Foundation for Basic Research (Grant no. 16-04-01447-a).

A COMPUTATION SYSTEM FOR RANDOMIZATION-BASED ENRICHMENT ANALYSIS USING GPU: PERFORMANCE INVESTIGATION

M. Grishenko[1], A. Yakimenko[1, 2], M. Khairetdinov[1, 2], K. Gunbin[3]
[1] *Novosibirsk State Technical University, Novosibirsk, Russia*
[2] *Institute Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*
[3] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: mikhail.grishch@gmail.com*

Key words: *Randomization, GPU*

Motivation and Aim: Processing of genetic data for the analysis genetic determination of traits is very important problem for modern biology. Resampling methods are widely used to solve such problems. However, they require much computational resources. Common trend of increasing the simultaneously tested set of genetic characteristics, reflecting the simultaneous contribution of many factors in the formation of a single biological trait is a challenge for computer science. The aim of this paper is improving of existing algorithm as well as it usability and performance.
Methods and Algorithms: We used graphic processor units (GPUs) for increasing the performance of permutation test. The whole algorithm workflow can be divided into three main stages: 1) reading the input file and forming the input data array according to convenient format for further calculations; 2) the sum calculation of the measured values of the gene functioning in the context of the experiment under analysis for various gene characteristics, the cycle composed of mixing of the array elements and after that collecting the quantities necessary for the statistics; 3) the exact calculation of p-values and the formation of output file. The most consuming and parallelizable is the second stage. It performed by multiplying two matrices: matrix containing numerical values obtained by the experiments and a matrix of gene characteristics. The cuBLAS library of matrix-matrix multiplication, which allowed for a transition of this algorithm to the architecture of GPUs, was used.
Results: A software for the permutation test aimed at finding statistically significant overrepresented gene characteristics under different external and/or internal conditions for computing devices: PC with NVIDIA GPU. We improved multiple testing of genetic characteristics overrepresentation by replacing matrix-vector multiplication with matrix-matrix multiplication. The resulting matrix contains a set of arrays with the sums reflecting various gene characteristics. We investigated performance of this software depending on the: (1) number of genes; (2) number of genetic characteristics. The execution time of program using the matrix-matrix multiplication for the simultaneous testing of characteristics overrepresentation increases slightly with the number of testable characteristics and experiments.
Conclusion: This software allows user to find statistically significant overrepresented characteristics of genes simultaneously.
Availability: Upon the requests to the authors.
Acknowledgements: This work was supported by the Russian foundation for basic research (Grant No. 16-37-00240)

ACTIVE MAINTENANCE OF PHYLOTRANSCRIPTOMIC HOURGLASS PATTERNS IN PLANT AND ANIMAL EMBRYOGENESIS

H.-G. Drost[1], A. Gabel[1], I. Grosse[1, 2], M. Quint[3, 4]
[1] *Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany*
[2] *German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany*
[3] *Department of Molecular Signal Processing, Leibniz Institute of Plant Biochemistry, Halle (Saale), Germany*
[4] *Institute of Agricultural and Nutritional Sciences, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany*
** Corresponding author: grosse@informatik.uni-halle.de*

Key words: *evolution, gene expression, hourglass model*

The developmental hourglass model has been used to describe the morphological transitions of related species throughout embryogenesis.
Recently, quantifiable approaches combining transcriptomic and evolutionary information provided novel evidence for the presence of a phylotranscriptomic hourglass pattern across kingdoms. As its biological function is unknown it remains speculative whether this pattern is functional or merely represents a nonfunctional evolutionary relic. The latter would seriously hamper future experimental approaches designed to test hypotheses regarding its function. Here, we address this question by generating transcriptome divergence index profiles across embryogenesis of Danio rerio, Drosophila melanogaster, and Arabidopsis thaliana. To enable meaningful evaluation of the resulting patterns, we develop a statistical test that specifically assesses potential hourglass patterns. Based on this objective measure we find that two of these profiles follow a statistically significant hourglass pattern with the most conserved transcriptomes in the phylotypic periods. As the transcriptome divergence index considers only recent evolutionary signals, this indicates that the phylotranscriptomic hourglass pattern is not a rudiment but possibly actively maintained, implicating the existence of some linked biological function associated with embryogenesis in extant species.

THE EVOLUTION OF LANGUAGE-READINESS IN THE HOMININ LINEAGE: AN ANALYSIS OF OPEN CHROMATIN REGIONS IMPLICATED IN GENE REGULATION

K. Gunbin[1,*], A. Benítez-Burraco[2], F. Gusev[1], E. Rogaev[1,3]
[1]*Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2]*Department of Philology, University of Huelva, Huelva, Spain*
[3] *University of Massachusetts Medical School, Worcester, USA*
e-mail: genkvg@gmail.com [*]*Corresponding author*

Key words: DNAseI Hypersensitive Sites (DHSs), language-readiness, extinct vs. extant hominins, self-domestication hypothesis.

Motivation and Aim: Language evolution studies rely on indirect evidence, mostly related to speech organs and human behaviour. The successful retrieval of ancient genomes is expected to contribute significantly to our knowledge of how cognition and communication evolved. Comparative genomic studies have demonstrated >99.9% of identity between modern human (*M*) and Neanderthal (*N*) and Denisovan (*D*) protein-coding gene sequences, which is in sharp contrast with the significant (~1 MYA) evolutionary distance between these three hominin species and with their behavioral and physical distinctive features [1,2]. Therefore, it is reasonable to anticipate an elevated rate of changes in the regulatory regions of the genes.

Methods and Algorithms: First, we aligned SNPs and indels from *D* [1], *N* [2], and *UI* [3] that are not found in *M*. Second, we mapped common changes in the *D-N* and *D-N-UI* clades onto DHSs from *M* with an annotated effect (activation or repression) on gene expression [4]. Third, we checked if the positions evolved in *D-N* and *D-N-UI* overlapped with transcription factor binding sites showing signatures of negative or positive selection in *M* lineage (using ENCODE RegTfbs data track and INSIGHT approach [5]). This allowed us to put forth a list of genes affected by changes in DHSs. In our past research we have put forth a list of candidate genes that could be associated with language-readiness [6-10]. Some of these genes show differences with the *N* and *D* homologs in regulatory (including methylation patterns) or coding regions [6-8]. Finally, we analyzed the nearest interactions between these two lists using GeneMania data (http://genemania.org/).

Results: Our results are suggestive of potential differences between *M* and *N*, *D* and *UI* regarding the regulation pattern of genes involved in osteogenesis, brain function, and immunity. This is interesting in view of the link between the emergence of language-readiness and skull-brain cross-talk [6], but also in the brain-immune system cross-talk [11]. We have also found that the regulation of *FOXP1* (in *D*), and *CNTNAP2* (in *N*), two well-known genes related to language function could have been affected by severe changes in DHSs of these hominin species. Finally, we found that several of our candidates for language-readiness and self-domestication are predicted to interact with genes affected by changes in DHSs. In the whole, our findings give support to the view that differences in expression and/or interactions of known candidates for cognitive development and language evolution may account for the presumed differences in cognitive and linguistic abilities between extinct and extant hominins.

Availability: Upon the requests to the authors. *Acknowledgements:* The study is supported by the grants 14.B25.31.0033 [Resolution No.220], FFI2014-61888-EXP, FFI2013-43823-P).

References:

1. M. Meyer et al. (2012) Science 338:222-226.; 2. G. Renaud et al. (2014) Nature 505:43-49.; 3. Q. Fu et al. (2014) Nature 514:445-449.; 4. N.C. Sheffield et al. (2013) Genome Research 23:777-788.; 5. L. Arbiza et al. (2013) Nature Genetics 45:723-729.; 6. C. Boeckx, and A. Benítez-Burraco (2014). Front. Psychol. 5:282.; 7. C. Boeckx and A. Benítez-Burraco, (2014). Front. Psychol. 5:1324.; 8. A. Benítez-Burraco and C. Boeckx (2015). Front. Psychol. 6:794.; 9. A. S. Wilkins et al. (2014). Genetics 197:795-808.; 10. B. Hare, and M. Tomasello (2005). Trends Cogn. Sci. 9:439-444.; 11. A. Benítez-Burraco and J. Uriagereka, (2016) Front. Mol. Neurosci. 8:84.

SEARCHING FOR CYTOLYTIC GENETIC MARKERS OF NEWCASTLE DISEASE VIRUS USING COMPUTER ASSISTED ANALYSIS

K.V. Gunbin[1, 4], M.R. Kabilov[2], K.S. Yurchenko[3], A.V. Glushchenko[3], A.M. Shestopalov[3], N.V. Gubanova[1, 4]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
[3] *Research Center of Clinical and Experimental Medicine SB RAMS, Novosibirsk, Russia*
[4] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: genkvg@gmail.com*

Key words: *Newcastle Disease Virus, whole genome sequencing, 3D protein structure*

Motivation and Aim: Newcastle disease virus (NDV) is one of the potential candidate viruses for therapy of human tumors. This is enveloped virus containing a non-segmented, single-stranded, negative-sense RNA genome. The replication of many viruses results in lysis of the tumor cells and inherent cytotoxicity. However, profound understanding of the mechanisms of virus replication is still illusive.
Methods and Algorithms: In order to distinguish the cytolytic genetic markers, 15 NDV class II [1] strains (s.) were tested for their ability to lyse the human tumor cell cultures. Whole genome sequencing of 8 new NDV s. was carried out. NDV genomes were aligned by SaTe 2.2.7; NDV proteins were aligned using Promals3D. In searching for the genetic basis underlying cytotoxicity, various RNA and protein characteristics were analyzed: (1) common SNPs and indels; (2) the localization and number of long inverted, tandem, and palindrome RNA repeats; (3) local RNA similarity and (4) the values of various dinucleotide RNA properties obtained by various window size; (5) codon adaptation indexes as described in [1]; (6) the 3D protein structure assisted comparison of conservative and variable positions. For this analysis M, HN, and F protein structures were taken directly from PDB, the 3D protein structures of L, NP, and V proteins were modelled using I-TASSER 4.4 and target structures with DELTA-BLAST E-value<1E-40. The protein structural flexibility was solved by CABS-flex.
Results: In their ability to lyse tumor cells, the strains were divided into two groups: cytolytic (8 s.) and non-cytolytic (7 s.). Only tiny differences between these groups were identified in two protein coding regions (<200nt length each) using various dinucleotide RNA properties. The projection of variable sites on 3D protein structures shown that non-cytolytic strains accumulate mutations in NP, L, and M proteins in sterically interacted regions with high level of structural flexibility and high total number of intramolecular contacts significantly more frequently than cytolytic strains. The same is observed for V proteins in cytolytic strains. Taking into consideration that these three proteins are critically important to NDV RNA replication and virion synthesis, it is reasonable to speculate that the cause of NDV cytotoxicity is the fast kinetic of proteins folding. The structure of NDV V protein resembles Simian virus 5V protein, which is known to modulate the activity of DDB1-CUL4-ROC1 E3 apparatus in favor of viral infection [3]. Therefore, the development of the same molecular mechanism promoting the viral infection is probable in cytopathic NDV strains.
Availability: Data available upon the requests to the authors. *Acknowledgements:* This work was supported by the grants 14-04-01196 and 2013–2020 (VI.53.1.4, VI.62.1.3, 0309-2014-0007).
References:

1. D.G. Diel et al. (2012) Infect Genet Evol. 12:1770-1779.
2. Y.Y. Waldman et al. (2010) Nucleic Acids Res. 38:2964-2974;
3. S. Angers et al. (2006) Nature. 443:590-593.

TRANSCRIPTOMIC ANALYSIS OF WHEAT ROOT IN RESPONSE TO ESSENTIAL NUTRIENT DEFICIENCY: A GEMOME-WIDE COMPARATIVE STUDY

S. Gupta[1], B.S. Yadav[2], S. Freilich[3], P.K. Varadwaj[1]*
[1] *Department of Bioinformatics, Indian Institute of Information Technology-Allahabad, India*
[2] *Deparment of Molecular Biology and Ecology of Plants, Tel Aviv University, Israel*
[3] *Systems Biology and Ecology ARO-Volcani Center-Bet-Dagan, Israel*
* *Corresponding author: Pritish@iiita.ac.in*

Key words: *genomics, wheat, nutrient deficiency*

Plant roots are essential for providing anchorage and uptake of nutrient and water, required for growth and development. The deficiency of necessary macro-nutrients, i.e. Potassium (K), Magnesium (Mg), Nitrogen (N), Phosphorus (P), and Sulfate (S) in soil directs to a reduction in crop yield and growth as a resulted of change in its gene expression profile. Aim of this study is to identify and annotate such deficiency related responsible genes and proteins as well as highlights the cellular and metabolic pathways response which is essential for survival of the plant. To differentiate the gene expression levels, various nutrient deficiencies (ND) treated, microarray data of wheat root tissue sample were analyzed. To annotate the Gene Ontology (GO) term for each nutrient deficiency response genes (NDRGs), Singular Enrichment Analysis (SEA) was performed.Further, the functional enrichment analysis was performed using MapMan annotations scheme. Macro-nutrients regulated genes were assigned into a functional category within the hierarchy of MapMan pathway scheme. Pathway enrichment in each experiment was determined by calculating the cumulative hypergeometric p value for the probability that a gene group is over-represented within a functional bin at a rate higher than chance expectation. Multiple tests were performed for all gene group functional categories at all hierarchical levels and Heatmap for each gene group was generated using "gpplaot" package in R. To revalidate the identified results a meta-analysis was performed using root specific earlier identified sample data of wheat through Genevestigator tool. A set of 435 statistically significant NDRGs was identified which got differentially expressed in response to at least one among all ND under study. Total, 58 NGRs were reported to be expressed in response to minimum two ND. P deficiency was seen to be affecting large numbers of NRGs in- comparison to other ND. The SEA predicts individual GO classification of NRGs in different biological process, cellular component and molecular function. MapMan analysis showed the carbohydrate and lipid pathways are up-regulated during potassium deficiency, hormonal (jasomonate and ethylene) as well amino acids synthesis (Asp) in Mg deficiency. In P deficiency, lipid degradation, MYB transcription factors, transportation and poly amines pathway were highly expressed but the cell-wall, fatty acid synthesis, protein translation modification, DNA synthesis and repairs and abiotic stress pathways were down-regulated. Amino acid and carbohydrate degraded and photosynthesis process down during nitrogen deficiency. Polyamine metabolism and protein synthesis pathways enriched in S deficiency in plant. Moreover, meta-analysis performed the significance analysis of NRGs, which is based on a large-scale systematic combination of normalized and quality-controlled expression data of root specific 109 samples of wheat. Study reveals several uncharacterized proteins in response to ND and other stress with provide a set of known genes, mRNA, proteins. Through functional annotation of these uncharacterized genes, components involved in different metabolic pathways and reaction mechanism of proteins/enzymes encoded by 155 transcripts were expressed during ND condition. The study demonstrated the deeper understanding of nutrient utilization in wheat root.

TRANSCRIPTION BY ALTERNATIVE SIGMA FACTORS: REVISING THE RIGIDNESS PARADIGM

J. Guzina[1*], M. Djordjevic[1]
[1] *Faculty of Biology, University of Belgrade, Belgrade, Serbia*
** Corresponding author: jelenag@bio.bg.ac.rs*

Key words: *ECF sigma, bacterial promoters, transcription initiation, σ^{70} family*

Motivation and Aim: In distinction to the housekeeping σ factors, which transcribe a majority of bacterial genes, alternative σ factors have a more specialized regulon, necessary for coping with stress or development. Among alternative σ factors, ECF (ExtraCytoplasmic Function) σs are the most numerous and diverse group, which is, on the other hand, insufficiently studied [1]. In particular, the current ECF paradigm, which assumes recognition of rigid promoters with well conserved elements, is supported by a very limited data originating from a few (canonical) representatives. This is contrary to the mix-and-matching paradigm, which is well established for the housekeeping σ^{70} factors, implying flexibility in the promoter recognition. With the goal of gaining a better insight in the transcription initiation by ECF σ factors, we did a comprehensive study of ECF σ protein and DNA recognition motifs, and analyzed flexibility (mix-and-matching) in ECF σ promoter recognition.
Methods: By combining protein multiple global and local sequence alignment, domain search and DNA regulatory elements detection, we extensively computationally analyze all the available bacterial ECF σ subgroups [3]. We also use the ECF group outliers obtained by recently sequenced bacteriophages, as a source of independent (self-contained) data, which we analyze through a novel procedure for detecting phage-promoters [2]. Additionally, we systematically (quantitatively) analyze the canonical bacterial ECFs, through a biophysics based procedure, to investigate contribution of their promoter elements to transcription activity.
Results: We found an extreme qualitative example of mix-and-matching for phage ECFs, where a long -10 element extension - interacting with an extension of σ_2 domain - complements the absence of the major (-35) promoter element. We also report examples of the putative novel interactions between ECFs and their promoters, exhibited by the conserved promoter spacer elements and the σ-motifs outside of the main σ_2 and σ_4 DNA-binding domains [3]. Finally, we provide quantitative evidence of substantial promoter element complementation (mix-and-matching) in ECFs [4].
Conclusion: We reveal a much larger flexibility in ECF σ functioning than previously recognized, which suggests that mix-and-matching may provide a common kinetic framework for promoter recognition in the entire σ^{70} family Novel protein and DNA recognition motifs that we discovered, will guide future experiments on alternative σs.
Acknowledgement: This work is funded by the Swiss National Science Foundation (IZ73Z0_152297), European Commission (GA-2010-276996) and Serbian Ministry of Science and Technology (ON173052).
References:
1. Staroń A. et al., 2009. Molecular microbiology 74:557-581.
2. Guzina J., Djordjevic M., BMC, Evolutionary Biology 15: S(1) (2015).
3. Guzina J., Djordjevic M., J Bacteriol, in press (2016), doi: 10.1128/JB.00244-16
4. Guzina J., Djordjevic M., to be submitted.

SCORING OF PROTEIN DOCKING BY GENE ONTOLOGY

A. Hadarovich[1, 2]*, I. Anishchenko[1], A.V. Tuzikov[2], P.J. Kundrotas[1], I.A. Vakser[1]
[1] *Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas, USA*
[2] *United Institute of Informatics Problems, National Academy of Sciences, Minsk, Belarus*
** Corresponding author: ahadarovich@gmail.com*

Key words: *template-based predictions, modeling of protein complexes, molecular recognition*

Motivation and Aim: Structural characterization of protein-protein interactions (PPI) is important for understanding life processes at the molecular level. Experimental techniques, due to their inherent limitations, can determine structures only for a fraction of known PPI. Thus, most structures of protein complexes have to be modeled by docking techniques. In the template-based (comparative) docking, suitable templates are detected by sequence and/or structure similarity between the target and the template. When, structures of individual proteins are not available, comparative docking has to deal with modeled structures of the interactors, which are typically less accurate than the ones determined by experimental techniques [1], thus reducing target/template structure similarity score. This requires additional scoring of the target/template match, which would compensate for such drop in structure similarity.
Results: We present a functional score, based on target/template similarity of the Gene Ontology (GO) annotations [2] (GO-score), which is complementary to the structure-based scoring (TM-score) [3]. A scoring function that combines TM- and GO-scores was tested on a non-redundant set of 165 protein-protein complexes. The set includes six models for each structure, generated with predefined C-alpha root mean square deviations from the native structure [4]. For templates, we used the set of 4,950 template complexes [5] from the DOCKGROUND resource [6].
Conclusion: The results show that the new combined score improves the template detection and can be successfully applied to template-based docking of modeled proteins.
References:

1. I. Anishchenko et al. (2014) Protein models: The Grand Challenge of protein docking, Proteins, 82: 278-87.
2. Gene Ontology Consortium (2012) Gene Ontology annotations and resources, Nucleic Acids Research, 41: D530-D535.
3. Y. Zhang et al. (2005) TM-align: A protein structure alignment algorithm based on the TM-score, Nucleic Acids Research, 33: 2302-2309.
4. I. Anishchenko et al. (2015) Protein models docking benchmark 2. Proteins, 83: 891-897.
5. I. Anishchenko et al. (2015) Structural templates for comparative protein docking, Proteins, 83: 1563-1570.
6. D. Douguet et al. (2006) DOCKGROUND resource for studying protein-protein interfaces, Bioinformatics, 22: 2612-2618.

KATIS: INTEGRATIVE INFORMATION SYSTEM FOR COMPLEMENTARY MEDICINE

R. Hofestädt*, V. Ogultarhan, A. Shoshi
University Bielefeld, AG Bioinformatics and Medical Informatics, Bielefeld, Germany
** Corresponding author: Ralf.Hofestaedt@uni-bielefeld.de*

Key words: *Information system, complementary medicine*

Motivation and aim: The Internet represents a lot of databases, information systems and portals specialized in complementary medicine. Most of these systems are specialized and represent only information about one category of alternative medicine. In addition, these systems are not able to support the computerized detection of suitable therapies based on patient specific indications. Furthermore, based on these systems various apps are available today. To develop and implement an information system which represents the main knowledge of complementary medicine is the goal of our work.
Methods: Using information fusion methods it was our idea to integrate the most relevant knowledge of alternative medicine. Therefore, we developed the warehouse based web-based information system KATIS. Furthermore, based on KATIS we developed the a new app called ALMEKO, which was implemented for mobile usage of the KATIS.
Results: We developed and implemented a web based information system which represents the knowledge of complementary medicine. This system is called KATIS. Based on KATIS we developed the app
Conclusion: We implemented an integrative information system for complementary medicine which allows the user to search for individual therapies based on patient-specific indications.
Availability: KATIS: http://www.komplementäre-medizin.de/
ALMEKO: https://play.google.com/store/apps/details?id=de.icancode.almeko&hl=de
References:
1. V. Ogultarhan, A. Shoshi, R. Magnucki, B. Kormeier, R. Hofestädt (2016) KATIS: An eHealth System for Complementary Medicine. In: Health Informatics Meets eHealth, Schreier, G. et al. (eds.), Proceedings of the 10th eHealth2016 Conference, Wien 2016, IOS Press, 167-173.

LONG NON-CODING RNAS IN FANTOM5

C.-C. Hon[1], J.A. Ramilowski[1], the FANTOM consortium, Y. Hayashizaki[2], P. Carninci[1], A. Forrest[1, 3]*

[1] RIKEN Center for Life Science Technologies (Division of Genomic Technologies), Suehiro-cho, Tsurumi-ku, Yokohama, Japan

[2] RIKEN Preventive Medicine and Diagnosis Innovation Program, 2-1 Hirosawa, Wako, Saitama, Japan

[3] Harry Perkins Institute of Medical Research, QEII Medical Centre and Centre for Medical Research, University of Western Australia, Nedlands, Western Australia, Australia

Key words: *transcriptomics, gene expression, non-coding RNA, GAGE*

The FANTOM5 project used Cap Analysis of Gene Expression (CAGE) to systematically map promoter and enhancer elements in the human and mouse genomes. The first phase of the project profiled a broad collection of primary cell types, tissues and cancer cell lines, to generate steady state 'snapshots' of these elements. In the second phase we studied multiple time courses of stimulation and differentiation, to identify elements that are dynamically regulated. Here I will describe our efforts using the FANTOM5 CAGE data, RNA-seq data and public transcript models to generate a comprehensive catalog of human 5' complete long non-coding RNAs with matched CAGE expression profiles across more than 1,800 samples. Integration of these lncRNA models with their expression profiles, overlap with conserved elements, trait associated polymorphisms, and epigenome data has provided convincing evidence that a significant fraction of these loci are functional elements of the genome.

IDENTIFICATION OF NEW CANDIDATE GENES FOR ELEVATED BODY MASS INDEX NEAR GWAS SNPS USING TRANSCRIPT ANNOTATIONS FROM ENSEMBL AND HAVANA PROJECTS

E.V. Ignatieva*, V.G. Levitsky
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: eignat@bionet.nsc.ru*

Key words: *GWAS, candidate genes, feeding behavior, body mass index*

Motivation and Aim: Feeding behavior (FB) and body mass index (BMI) are controlled by a complex network, which involves genes expressed in many organs and tissues. Genome-wide association study (GWAS) technology is broadly used to reveal SNPs and candidate genes significantly associated with BMI. However, for a substantial part of GWAS SNPs the risk-genes were not identified yet. In this study we used automated/manual annotation of human transcripts from Ensembl and HAVANA gene tracks to search new candidate genes relevant to BMI regulation. Our goals were to find new genes near GWAS SNPs and to explain their functional significance in the context of BMI regulation.
Methods and Algorithms: 164 SNPs (GWAS SNPs) and 184 candidate genes (known GWAS BMI genes) were collected from nine GWAS meta-analysis publications. All GWAS SNPs were associated with elevated BMI at the genome-wide significance level ($P<5.0\times10^{-8}$). The annotations of transcripts for the GRCh38 and Hg19 assemblies of the human genome were extracted from the Ensembl archives by the Biomart data-mining tool. Biological functions and relevance of novel genes to BMI regulation were analyzed using STRING system, RefSeq, KEGG Pathway, Reactome and a compilation of human genes controlling FB or BMI (FB/BMI genes), described previously [1].
Results: We found 138 genes, which contain one of 164 GWAS SNPs in their exons, introns or 10 kb upstream regions. Among these genes we revealed 56 novel genes which were not present in the list of 184 known GWAS BMI genes. Using STRING we found that five out of 56 novel genes have gene-gene or protein-protein interactions with genes involved in regulation of FB or BMI (FB/BMI genes). Additional manual verification using RefSeq, Reactome and KEGG Pathway databases confirmed that some genes from the list of novel genes might be involved in the network controlling BMI.
Conclusion: Our results demonstrate that modern annotation of human transcriptome from Ensembl and HAVANA teams provides advantages in scanning genome for new candidate genes. We suggested new candidate genes relevant to FB and BMI regulation. This result may give a deeper view of molecular-genetic basis of FB and BMI abnormalities and may be useful for designing new pharmacological approaches for the treatment of human diseases.
Acknowledgements This work was supported by the Government of Russian Federation grant no. 14.B25.31.0033 (to E. I. Rogaev) and Federal Agency of Scientific Organizations (project 0324-2015-0003).
References:
1. E.V. Ignatieva et al. (2014) Human genes controlling feeding behavior or body mass and their functional and genomic characteristics: a review, *Vavilov Journal of Genetics and Breeding*, 18:867-875.

THE COMPENDIUM OF HUMAN GENES CONTROLLING FEEDING BEHAVIOR OR BODY WEIGHT, RECONSTRUCTION OF NETWORKS AND ANALYSIS OF THEIR PROPERTIES

E.V. Ignatieva*, O.V. Saik, D.A. Afonnikov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: eignat@bionet.nsc.ru*

Key words: *Feeding behavior, Body mass index, Network reconstruction and analysis*

Motivation and Aim: Elevated body mass index (BMI) is a substantial risk factor for human disease emergence. The goals of this study were to create the list of genes controlling feeding behavior (FB) or BMI and to analyze a network formed by interactions between genes/proteins.
Methods and Algorithms: Data were collected from scientific publications, OMIM, Human Protein Atlas, Genotype-Tissue Expression project and were analyzed using DAVID, STRING, Cytoscape and GeneMANIA.
Results: The compendium of genes controlling FB or BMI included 571 human genes. Among them 103 genes were collected from scientific publications as involved in regulation of FB in humans (or in mice or rats). 73 genes have OMIM-annotated allelic variants associated with FB abnormalities (hyperphagia, anorexia) or obesity. 263 genes have OMIM evidences for involvement in FB regulation, but have no data on allelic variants associated with FB abnormalities or obesity. 22 genes were included into compendium because they are associated with Bardet-Biedl, Prader-Willi or Alstrom syndromes. 184 genes have evidences from GWAS meta-analysis: these genes are located near SNPs associated with elevated body mass index at the genome-wide significance level ($P<5.0\times10^{-8}$). Genes were ranged according to their significance in regulation of BMI and classified according to expression patterns or functional characteristics. Considerable portions of genes from the compendium encoded cell surface receptors, signaling molecules (hormones, neuropeptides, etc.), and signal transduction proteins. We identified 27 pathways from KEGG, REACTOME and BIOCARTA whose genes were overrepresented in the compendium. Networks formed by physical interactions between proteins, homology or involvement into pathways were reconstructed and their main characteristics (number of neighborhoods, clusters) were analyzed. Nodes, which correspond to GWAS genes, which were not interpreted yet, were selected and ranged according to the number and biological significance of their first neighborhoods.
Conclusion: A compendium of human genes controlling FB or BMI was deigned. Analysis of a networks formed by genes/proteins interactions provided new biological interpretations for some GWAS genes. Our results may give a deeper view of molecular-genetic basis of FB and BMI abnormalities and may be useful for designing new pharmacological approaches for the treatment of human diseases
Acknowledgements: This work was supported by the Government of Russian Federation grant no. 14.B25.31.0033 (to E. I. Rogaev) and Federal Agency of Scientific Organizations (project 0324-2015-0003).

SYNTHESIS AND ACCUMULATION OF A NOVEL FUNCTIONAL FOOD COMPONENT IN TOMATO

A. Ito, S. Hano, N. Imoto, T. Shibuya, Y. Kanayama*
School of Agricultural Science, Tohoku University, Sendai, Japan
** Corresponding author: yoshinori.kanayama.a7@tohoku.ac.jp*

Key words: *functional food component, metabolome, Solanum lycopersicum, serotonin*

Motivation and Aim: Serotonin is an aromatic amine neurotransmitter that controls several physiological functions such as mood and sleep in humans. Approximately 98% of serotonin is synthesized and stored in the peripheral system, although its function in the peripheral system is not fully understood. Here, we focused on serotonin as a new functional food component for promoting health, because serotonin in the peripheral system enhances lipid metabolism and exerts an anti-obesity effect in mice according to some recent reports.
Results and Conclusion: Because serotonin consumed through food cannot cross the blood–brain barrier and functions within the peripheral system, it is important to focus on vegetables and fruits as its sources having anti-obesity effects in the fields of horticultural and food science. During the determination of serotonin content in various vegetables and fruits, tomato fruit was found to be rich in its content; high content was observed in the mesocarp tissue of ripe tomato fruit, whereas low content was observed in processed tomato products such as juice and ketchup. These results indicate that fresh tomato fruit is a promising source. Although the serotonin biosynthesis pathway in plants is under dispute, it was identified in tomato fruit using metabolome analysis. Serotonin is likely synthesized using tryptamine by tryptophan decarboxylase (TDC) and tryptamine-5-hydroxylase. Overexpression of the TDC gene increased serotonin content in tomato fruit and some phenotypes of the gene were found in transgenic plants, suggesting that TDC plays a key role in serotonin synthesis and that serotonin metabolism has some physiological functions.

NOVEL APPROACH FOR COMPUTATIONAL DESIGN OF SMALL MOLECULE INHIBITORS OF PROTEIN/ PROTEIN INTERACTIONS IN CD95/FAS PATHWAY

N.V. Ivanisenko[1, 2]*, A.S. Ishchenko[1, 2], I.N. Lavrik[1, 3], V.A. Ivanisenko[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Otto-von Guericke-University, Magdeburg, Germany*
* *Corresponding author: ivanisenko@bionet.nsc.ru*

Key words: *Virtual Screening, Molecular Docking, Molecular Dynamics, Solvation, Apoptosis, CD95*

Motivation and Aim: Apoptosis is an important form of the programmed cell death of all eukaryotic organisms. Defects in regulation of programmed cell death can lead to number of oncological and neurodegenerative diseases. Protein-protein interactions play one of the main roles in regulation of apoptosis. Compounds targeting key protein/protein interactions involved in apoptosis inhibition have shown great potential as a new therapeutics for cancer treatment. Despite recent progress in rational structure based drug design of protein/protein inhibitors, design of small molecules able to inhibit protein/ protein interaction is still of great challenge. One of the reasons is a lack of well-defined binding site on target of interest. In addition, solvent effects have to be considered that can enhance precision of binding site prediction and molecular docking accuracy. In this work we developed and applied new *in silico* approach for improvement of virtual screening accuracy of solvent exposed small molecule inhibitors.
Methods and Algorithms: Molecular dynamics simulations and molecular docking methods were used in this work. Analysis of trajectories from molecular dynamics simulations was done using Python-based scripts.
Results: An algorithm for prediction of conservative water clusters in the vicinity of small molecule binding site based on analysis of molecular dynamics trajectories have been developed. Further hydrogen-bond network graph analysis with implemented empirical scoring function allow to estimate binding site «druggability» and predict possible energetically favorable displacement of water molecules by small molecule inhibitors. Validation on set of X-RAY structures of complexes with solvent exposed small molecule inhibitors showed significant increase of enrichment score in comparison with conventional virtual screening. We have applied this approach for design of small molecules targeting proteins involved in CD95/FAS regulation pathway to predict potential enhancers and inhibitors of apoptosis induction.
Conclusion: Developed approach for prediction and analysis of conservative waters clusters allow to significantly increase molecular docking accuracy and can be applied for design of new small molecule inhibitors of protein/protein interactions.
Availability: Software is freely available as a Python-script.
Acknowledgements: The work was supported by Russian Scientific Foundation grant No. 14-44-00011.

POLYMORPHISM OF *LOXL1* GENE IN WEST SIBERIA PATIENTS WITH OPEN ANGLE GLAUCOMA AND PSEUDOEXFOLIATION GLAUCOMA.

D.E. Ivanoshchuk[1,2]*, E.O. Datskih[3], N.A. Konovalova[3], I.Y. Bychkov[4], A.Z. Fursova[5], O.S. Konovalova[3], S. V. Mikhailova[1] , M.N. Ponomareva[3], M.I. Voevoda[1,2], A.G. Romaschenko[1]

**dinara@bionet.nsc.ru*

[1]Federal Research Center Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

[2]Institute of Internal and Preventive Medicine, Novosibirsk, Russia

[3]Tyumen State Medical Academy, The Ministry of Health of the Russian Federation, Tyumen, Russia

[4]The Novosibirsk Branch of FSBI "The Acad. S.N. Fyodorov Eye Microsurgery Complex" of the Minpublhealth of Russia, Novosibirsk, Russia

[5]Novosibirsk State Regional Clinical Hospital, Novosibirsk, Russia

Key words: open angle glaucoma, single nucleotide polymorphism, *LOXL1* gene

Motivation and Aim: Open-angle glaucoma (OAG) is a chronic neurodegenerative eye disease caused by axons degeneration of the retinal ganglion cells (RGC) and subsequent neuron death. Pseudoexfoliative glaucoma (PEXG), which is characterized by severe course of the disease and poor prognosis in vision, is one of OAG types. The association of *LOXL1* gene with PEXG in a genome-wide association study was shown [1]. The aim of this study was estimation of a possible association the *LOXL1* gene rs1048661 (R141L) and rs2165241 (IVS1) with OAG and PEXG in West Siberia patients.

Methods and Algorithms: In this case/control study, 197 unrelated patients with OAG, 197 patients with PEXG, and unrelated healthy individuals were genotyped using RLFP analysis. For intergroup comparison by the genotype frequencies, the exact Fisher test was calculated using the SPSS 11.0 computer program. The intergroup differences were considered statistically significant if the p- value was less than 0.05 ($p < 0.05$).

Results: The analysis of the OAG and PEXG samples revealed a significant increase in the rs2165241 (IVS1) TT genotype frequencies (51.3% for OAG ($p<0.0001$) and 53.8% for PEXG ($p<0.0001$) respectively) compared to the control group (28.5%). The rs1048661(R141L) GG genotype frequencies were significantly higher in OAG (69.0%, $p= 0.023$) and PEXG (67.5%, $p= 0.005$) patients compared to the control (54,8%). A significant decrease in the frequencies of the carriers of TT genotype was detected in OAG (3.0%, $p= 0.023$) and PEXG (1,5%, $p= 0.001$) patients compared to the control group (8.0%). The frequency of GT genotype at the rs1048661 (R141L) was significantly decreased in OAG (27.9%, $p= 0.034$) patients vs control (37.2%).

Conclusion: The *LOXL1* gene rs1048661 (R141L) and rs2165241(IVS1) polymorphisms correlate with an increased risk of POAG and PEXG in West Siberia patients. The observed effect should be considered for further studies in order to estimate the contribution of the gene environment interaction to POAG and PEXG phenotypes.

References:

1. G. Thorleifsson et al. (2007) Common sequence variants in the LOXL1 gene confer susceptibility to exfoliation glaucoma, *Science,* **317:** 1397–1400

IDENTIFICATION OF PROTEINS ASSOCIATED WITH DRUG-INDUCED LIVER INJURY USING IN SILICO PREDICTION OF DRUG-TARGET INTERACTIONS

S.M. Ivanov[1, 2]*, M.I. Semin[1, 2], A.A. Lagunin[1, 2], D.A. Filimonov[1], V.V. Poroikov[1, 2]
[1] *Institute of Biomedical Chemistry, Moscow, Russia*
[2] *Pirogov Russian National Research Medical University, Medico-Biological Faculty, Moscow, Russia*
** Corresponding author: sergey.ivanov@ibmc.msk.ru*

Key words: *drug-induced liver injury, idiosyncrasy, acute liver failure, drug-target interactions, off-targets, structure-activity relationships, PASS Targets, Gene Ontology*

Motivation and Aim: Drug-induced liver injury (DILI) is the leading cause of acute liver failure as well as one of the major cases of drug withdrawn from clinical trials and the market. Understanding of DILI-related mechanisms may help to improve the existing and develop new methods of DILI detection on the earliest stages of drug development. Most of the investigations focused on the formation of toxic or reactive metabolites, whereas specific interactions with protein targets are accepted to be the primary cause of many other adverse drug effects [1, 2].
Methods and Algorithms: Specific DILI-related protein targets were identified through the analysis of drug-target interactions which were predicted by PASS Targets software [3]. It predicts interactions with 1534 human targets. The study was carried out using a dataset containing 178 severe DILI-causing drugs, 310 moderate DILI-causing drugs and 211 non-DILI-causing drugs, which was created based on mainly SIDER (http://sideeffects.embl.de/) and LiverTox (http://livertox.nih.gov/) databases.
Results: Statistical analysis of predicted drug-target interactions of dataset's compounds coupled with analysis of Gene Ontology allows revealing 145 protein targets putatively associated with DILI as well as cellular pathophysiological processes leading to DILI. Most of the revealed processes were associated with hepatocytes, the main from which was apoptosis. Interactions with proteins which were involved in immune system regulation were also identified. About half of DILI-causing drugs from various chemical-therapeutic classes interact with the revealed targets. We clustered drugs based on their interactions with 145 targets and confirmed correlations with DILI within clusters for 61 from those targets. These 61 protein targets are possibly the most essential for DILI development.
Conclusion: We found that interaction with the identified specific protein targets has a major role in the development of severe DILI.
Acknowledgements: This work was supported by the Russian Foundation for Basic Research grant 16-34-01077.
References:
1. M. Chen et al. (2015) Drug-induced liver injury: Interactions between drug properties and host factors, *J. Hepatol.*, 63:503-514.
2. S.M. Ivanov et al. (2016) In silico assessment of adverse drug reactions and associated mechanisms, Drug Discov. Today, 21:58-71.
3. P.V. Pogodin et al. (2015) PASS Targets: Ligand-based multi-target computational system based on a public data and naïve Bayes approach, *SAR QSAR Environ. Res.,* 26:783-793.

USING THE BIOINFORMATIC SOFTWARE TECHNICUES TO SEARCH CRISPR / CAS SYSTEMS IN THE GENOME OF *ESCHERICHIA COLI* STRAIN O157:H7

E.I. Ivanova[1]*, Yu.P. Dzhioev[1, 2], A.Yu. Borisenko[2], A.I. Paramonov[1], V.I. Zlobin[2], N.L. Belkova[1, 3]

[1] *Scientific Center of the Family Health and Human Reproduction Problems, Irkutsk, Russia*

[2] *Irkutsk State Medical University, Irkutsk, Russia*

[3] *Limnological Institute SB RAS, Irkutsk, Russia*

* *Corresponding author: ivanova.iem@gmail.com*

Key words: *E. coli O157:H7, CRISPR / Cas-system, bacteriophages, bioinformatics*

Motivation and Aim: CRISPR / Cas-system (Clustered Regularly Interspaced Short Palindromic Repeats / CRISPR-associated proteins, or short palindromic repeats regularly arranged in groups with CRISPR - associated protein) is a prokaryotic adaptive protection system from foreign genetic material. CRISPR-cassette consists of palindromic repeats, between which there are spacers - sections of DNA that are complementary to protospacers of phages and plasmids. Bacteria containing spacers exhibit resistance to the corresponding phage or plasmid. CAS-proteins possess helicase and nuclease activity and are usually in close proximity to each other. The aim of this work is to find and analyze CRISPR / Cas-systems sites in the genome sequence of *Escherichia coli* serotype O157:H7 strain using bioinformatics methods.

Methods and Algorithms: Genomic sequence of *E. coli* O157:H7 strain was taken as an object of research. It was downloaded from the GenBank database (accession number NC_002695). Methods from MacSyFinder ver 1.0.2 (Macromolecular System Finder) software package were used to find CRISPR / Cas-system sites. Identification of structural and functional characteristics of discovered cas genes were carried by auxiliary programs of the makeblastdb ver.2.2.28 and HMMER ver. 3.0 packages. Visualization of the results was carried out through MacSyView web interface. «CRISPI: a CRISP RInteractive database» online application on Gen Ouest BioInformatics Platform were used for explanation of produced CRISPR-cassettes.

Results: CRISPR / Cas-system locus was identified in the studied strain of *E. coli* in positions 2920680-2921322, ie its length was 642 nd. 7 cas 2 sse genes belonging to CAS-TypeIE were detected and visualized. Their structural and functional characteristics were determined. Consensus size of repeats was 29 nd. CRISPR-cassette structure was identified containing repeats and 11 spacer sequences. Their size ranged from 29 to 35 nd.

Conclusion: Analysis of the decoded CRISPR-cassette in the *E. coli* O157:H7 strain allows evaluating its ability to defend against phages for which complementary spacers are identified. Presence of mandatory sse and cas proteins shows high anti phage activity of its CRISPR / Cas-system. The number of revealed spacers evidences its long evolutionary history. Information about the CRISPR / Cas-system of this strain makes it possible to select phages for strain-specific phage therapy.

MIR-619-5P BINDING SITES IN PROTEIN CODING REGION OF ORTHOLOG GENES MRNA

A.T. Ivashchenko*, S.A. Atambayeva, R.E. Niyazova, A.Y. Pyrkova
Al-Farabi Kazakh National University, Almaty, Kazakhstan
** Corresponding author: a_ivashchenko@mail.ru*

Key words: *miRNA, miR-619-5p, mRNA, binding sites, ortholog genes*

Motivation and Aim: miRNA binds to the 5′UTR, CDS, and 3′UTR of mRNA. miR-619-5p has binding sites in coding regions of mRNA for the C8orf44, ISY1, NANOGNB, ZNF429, and ZNF714 genes. It is important to know how conservative the prediction is for the binding sites in mRNA of orthologous genes.
Methods and Algorithms: The target genes for miR-619-5p were revealed using the MirTarget program [1]. This program defines: a) the origin of the initiation of miRNA binding to mRNAs; b) the localisation of miRNA binding sites in the 5′-untranslated region (5′UTR), the coding domain sequence (CDS) and the 3′-untranslated region (3′UTR) of the mRNA; c) the free energy of hybridisation (ΔG, kJ/mole); and d) the schemes of nucleotide interactions between the miRNAs and the mRNAs. The ratio ДG/ДGm (%) was determined for each site (ДGm equals the free energy of a miRNA binding with its perfect complementary nucleotide sequence).
Results: miR-619-5p (5′-GCUGGGAUUACAGGCAUGAGCC-3′) has 526 binding sites in 3′UTRs, 10 sites in 5′UTRs, and 5 sites in CDSs in mRNAs of 508 genes, with ΔG/ΔGm equal to 98-100%. miR-619-5p binding sites in CDS may encode a protein in the three reading frames, two of which encode the WLMPVIP and AHACNPS oligopeptides, and the third has a stop codon. The miR-619-5p binding site in the mRNA of the C8orf44 gene was perfectly complementary (ΔG/ΔGm=100%) in the orthologs Homo sapiens, Gorilla gorilla, and Macaca nemestrina, and it encoded the GRARWLMPVIPALWE polypeptide, which contains the WLMPVIP oligopeptide. In the C8orf44 ortholog gene of Macaca fascicularis, Macaca mulatta, Cercocebus atys, and Mandrillus leucophaeus, the binding site is encoded in the WLMPAIP oligopeptide (ΔG/ΔGm=98%). miR-619-5p binding sites (ΔG/ΔGm = 100%) in the mRNA of the ISY1 gene of H. sapiens and Pan troglodytes are encoded by the RQVRWLMPVIPALWE polypeptide. In mRNA of the NANOGNB gene of H. sapiens, Nomascus leucogenys, Pan paniscus, P. troglodytes, Pongo abelii, miR-619-5p binding sites are encoded by the HRARWLTPVIPALWE polypeptide (ΔG/ΔGm=98%). miR-619-5p binding sites in mRNA of the ZNF429 and ZNF714 genes are oligopeptides encoded in another reading frame and were synthesised as the oligopeptides AHACNPS (ΔG/ΔGm=100%) and AHACNPN (ΔG/ΔGm=98%). mRNA of the ZNF429 gene of H. sapiens, M. nemestrina, M. fascicularis, and M. mulata encoded the MGVVAHACNPSTLG polypeptide. mRNA of the ZNF714 gene of H. sapiens, G. gorilla, P. paniscus, and P. troglodytes encoded the QGMVAHACNPNTLR polypeptide.
Conclusion: miR-619-5p binding sites in CDS mRNA of C8orf44, ISY1, ANOGNB, ZNF429, and ZNF714 orthologous genes are highly conserved.
Reference:

1. A.T. Ivashchenko et al. (2014) MiR-3960 binding sites with mRNA of human genes, Bioinformation, 10: 423-426.

FEATURES OF MIRNA INTERACTION WITH MRNA GENES IN CORONARY HEART DISEASE

A.T. Ivashchenko*, R.E. Niyazova, S.A. Atambayeva, A.Y. Pyrkova
Al-Farabi Kazakh National University, Almaty, Kazakhstan
** Corresponding author: a_ivashchenko@mail.ru*

Key words: *miRNA, miR-619-5p, mRNA, binding sites, ortholog genes*

Motivation and Aim: In total, 268 binding sites for 2578 miRNAs were found in mRNA of 187 genes involved in the development of coronary heart disease. Of these, 52 were located in the coding domain sequence (CDSs), 23 were located in the 5′-untranslated region (5′UTRs), and 193 were located in the 3′-untranslated region (3′UTRs). From the database of miRNAs involved in the development of coronary heart disease, none of the miRNA had binding sites on the mRNA of 187 genes that play a role in coronary heart disease. It is possible that these miRNAs act on mRNA genes that are not included in the database of genes involved in the development of coronary heart disease. *Methods and Algorithms:* The target genes for miR-619-5p were revealed using the MirTarget program. This program defines: a) the origin of the initiation of miRNA binding to mRNAs; b) the localisation of miRNA binding sites in the 5′UTR; the CDS and the 3′UTR of the mRNA; c) the free energy of hybridisation (ΔG, kJ/mole); and d) the schemes for nucleotide interactions between the miRNAs and the mRNAs. The ratio ΔG/ΔGm (%) was determined for each site (ΔGm equals the free energy of an miRNA binding with its perfect complementary nucleotide sequence).
Results: Features of the interaction of miRNA with the mRNA of 85 target genes are described below. mRNA of some genes can bind five or more miRNAs. Five miRNAs bind to the mRNA of the *MTHFR, PLA2G7* genes. Six miRNAs bind to the mRNA of the *AS3MT, F2RL3, MLXIPL, PPP1R3B*, and *TGFB1* genes. Seven miRNAs bind to the mRNA of the *IL6R, LDLR, MLXIPL,* and *NPC1L1* genes. These data indicate a strong dependence of the expression of these genes on miRNA. The mRNA for the *CD36* and *PLA2G7* genes has multiple binding sites for miR-466, which belongs to a class of unique miRNAs. The mRNA for the *IGF1, NOS1*, and *PPARA* genes has multiple binding sites for miR-574-5r, which also belongs to a class of unique miRNAs. We have previously shown that unique miRNAs are encoded in the human genome, and they have more than 300 binding sites. These miRNAs include miR-619, miR-5095, miR-5096, miR-3960, miR-1322, and some of the miRNAs of the miR-1273 family. Expression of a significant part of the genes involved in the development of coronary heart disease may depend upon these unique miRNAs. For example, miR-619-5p has 14 target genes, whereas miR-5095 and miR-5096 have 10 target genes each. miR-1273a,c,d,e,f,g,h family miRNAs have 38 binding sites, including 19 binding sites for miR-1273g-3p in mRNA for 17 genes. miR-6089-5p, consisting of 24 nt, binds to the mRNA of the *TGFB1* gene at two sites with free energies of binding of -132 kJ/mole and -136 kJ/mole. The same miRNA binds to the mRNA of the *IL6R* gene with a binding free energy of -138 kJ/mole (93% of the maximum binding free energy).
Conclusion: These data show that the interactions between the examined miRNA and mRNA can serve as a basis for selecting associations between miRNA and mRNA for diagnostic evaluation of coronary heart disease. Association means the connection of one miRNA with mRNA of one or more genes, or one or more miRNAs with the mRNA of a single gene.

SEX CHROMOSOME EVOLUTION IN PAMPHAGIDAE GRASSHOPPERS

I.E. Jetybayev[1, 2]*, A.G. Bugrov[2, 3], O.G. Buleu[2, 3], A.G. Bogomolov[1], N.B. Rubtsov[1, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Systematics and Ecology of Animals SB RAS, Novosibirsk, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: jetybayev@bionet.nsc.ru*

Key words: *Pamphagidae grasshoppers, karyotype, neo sex chromosome evolution, the neo-X, the neo-Y*

Motivation and Aim: In evolution of grasshoppers, formation of neo sex chromosome is a rare event. However, numerous neo sex chromosomes were revealed in Pamphagidae family. It was shown that neo-Y chromosome had undergone heterochromatinistion and degradation [1]. The neo-Y chromosome of studied closely related species showed different stages of heterochromatinistion and degradation making this group to be perspective model of sex chromosome evolution.
Methods and Algorithms: Classical banding, molecular cytogenetic mapping and FISH of microdissected DNA probes were used for analyzing of the neo sex chromosomes. VISSIS software [2] was used to analyze images of cross hybridization results.
Results: Twelve previously unstudied species were karyotyped. Two different types of neo-Y chromosomes were revealed. The first type was similar to the XR arm of the neo-X but contained two small proximal interstitial C-bands. The second type of the neo-Y chromosome was smaller than the XR of the neo-X and more heterochromatinized. Cross hybridization of microdissected DNA probes derived from neo sex chromosomes with chromosomes of different species revealed partial homology of C-positive block in neo-Y chromosomes of Glyphotmethis and Asiotmethis genera. Partial homology was also observed in neo-Y chromoosmes of species belonged to Nocarodeini tribe.
Conclusion: Most studied Pamphagidae species from the Western Asian region showed a neo sex chromosome. There are two evolutionary lineages of Pamphagidae grasshoppers in this region, characterized with different neo sex chromosome systems. Analysis of cross hybridization results indicated that X-chromosome-autosome fusions were probably independent events in the ancestors of the group of Trinchinae species and the Nocarodeini tribe. The more advanced reorganization of the sex chromosomes in Nocarodeini in comparison with the neo sex chromosomes in Trinchinae species pointed to the “older” history of their formation. Further investigation of the Pamphagidae family neo sex chromosome systems can help to clarify the mechanisms of neo sex chromosome formation and evolution in grasshoppers.

References:

1. Bugrov AG, Jetybayev IE, Karagyan GH, Rubtsov NB. Cytogenetics Sex chromosome diversity in Armenian toad grasshoppers (Orthoptera , Acridoidea , Pamphagidae). 2016;10:45–59.
2. Bogomolov AG, Zadesenets KS, Karamysheva TV, Podkolodnyi NL, Rubtsov NB. Visualization of chromosome-specific DNA sequences by fluorescence in situ hybridization of microdissection DNA probes with metaphase chromosomes. Vavilov jaurnal Genet. Sel. 2012;16:348–57.

DNA REPAIR AND DEATH SIGNALING TARGETED BY ALKYLATING ANTICANCER DRUGS

B. Kaina
Department of Toxicology, University Medical Center, Mainz, Germany
** Corresponding author: kaina@uni-mainz.de*

Key words: *DNA repair, cancer*

Alkylating agents are first-line therapeutics for the treatment of malignant brain tumors and several other types of cancer. Although they alkylate DNA at several sites, the minor product O6-alkylguanine represents a major cytotoxic DNA lesion, activating the apoptotic pathway. Autophagy and replicative senescence are also induced by the damage, indicating a compex signaling network evoked by a single type of damage. While for methylating agents the DNA repair enzyme MGMT protects against all these effects, mismatch repair is essentially required for eliciting these responses. Downstream in the pathway are DNA double-strand breaks (DSBs), representing the key trigger of cell death and presumably also replicative senescence. Searching for drug modifyers, we identified homologous recombination as a major pathway for repairing alkylating agent-induced DSBs, with XRCC2, XRCC3, Rad51, BRCA2 and other repair proteins being involved causing cancer cell resistance. We further showed that DNA damage signaling (DDR) activating ATR-CHK1 and, to less extent, ATM-CHK2, results in resistance to methylating agents. Data will be presented demonstrating that pharmacological inhibition of homologous recombination and DDR ameliorates the killing response of alkylating anticancer drugs; PARP inhibition causes synthetic lethality and histone deacetylases may have an impact on DNA repair and survival. The role of p53 and JNK/AP-1 in upregulating DNA repair and apoptosis genes will also be discussed, highlighting their importance at the cutting edge between survival and death in cancer therapy. Supported by DFG KA724.

FUNCTIONAL ANALYSIS OF RNA-SEQ TRANSCRIPTOMES FROM OESOPHAGEAL CANCER SPECIMENS OF KAZAKHSTANI PATIENTS

U. Kairov[1]*, A. Molkenov[1], S. Rakhimova[1], A. Abilmazhinova[1], M. Zhalbinova[1], D. Yerezhepov[1], A. Akhmetova[1], Y. Zhukov[2], M. Omarov[2], M. Popova[3], A. Zinovyev[4], A. Akilzhanova[1], Zh. Zhumadilov[1]

[1] *Center for Life Sciences, NLA, Nazarbayev University, Astana, Kazakhstan*
[2] *Oncology Center, Astana, Kazakhstan*
[3] *Department of Pathology, Astana Medical University, Astana, Kazakhstan*
[4] *Institute Curie, Paris, France*
* *Corresponding author: ulykbek.kairov@nu.edu.kz*

Key words: *oesophageal cancer, transcriptome, RNA-seq, next-generation sequencing*

Motivation and Aim: Oesophageal cancer is the eighth most common cancer in the world and the highest in Eastern Asia. The incidence rate in Kazakhstan is 10.1: 100 000. Oesophageal squamous cell carcinoma (ESCC) is the dominant histological type (>90%) of esophageal cancer cases. Transcriptomic profiling of cancer specimens with next-generation sequencing technologies has provided a comprehensive opportunity for in-depth investigation of gene expression and affected molecular pathways. In our study we aimed to perform whole-transcriptome sequencing to identify affected molecular pathways and extract meaningful molecular signals from oesophageal cancer specimens of Kazakhstani patients.
Methods and Algorithms: Twenty three patients with esophageal cancer that underwent surgery at Oncology center (Astana, Kazakhstan). Fresh frozen cancer and its adjacent normal tissue specimens were obtained from each patient (in total 23 tumor center samples and 23 normal tissue samples). Whole-transcriptome sequencing was performed on Illumina HiSeq2000 platform at the Center for Life Sciences, Nazarbayev University. mRNA libraries were prepared using TruSeq RNA library prep kit according to standard protocol. Raw *.bcl files were converted and demultiplexed using bcltofastq. STAR and HTSeq have been used for alignment and mapping of sequencing reads. Differentially expressed genes have been identified using DeSeq. KEGG and Reactome databases were processed for analysis of signaling networks.
Results: Grouped analysis of cancer and normal samples has identified 1072 down-regulated and 1963 up-regulated genes. Functional analysis of up-regulated genes revealed the most significant enrichment for genes encoding products in the category of 'cell cycle' (p-value = 2.3×10^{-6}), 'DNA replication' (p-value = 1.8×10^{-4}) and 'lysosome'(p-value = 2.31×10^{-5}), whereas down-regulated genes in the category 'metabolism of lipids and lipoproteins' (p-value=8.62×10^{-4}), 'valine, leucine and isoleucine degradation' (p-value = 1×10^{-6}) and 'propanoate metabolism' (p-value = 6.4×10^{-6}).
Conclusion: Here, we report functional analysis of transcriptomic profiles from oesophageal cancer and matched adjacent normal specimens from twenty three Kazakhstani patients.

DIVERGENCE OF PARALOGOUS GROWTH HORMONE GENES IN SALMONIDS

D.N. Kamenskaya[1]*, M.V. Pankova[2], D.M. Atopkin[2, 3], V.A. Brykov[1, 2]
[1] *A.V. Zhirmunsky Institute of Marine Biology FEB RAS, Vladivostok, Russia*
[2] *Far Eastern Federal University, School of Natural Sciences, Vladivostok, Russia*
[3] *Institute of Biology and Soil Science FEB RAS, Vladivostok, Russia*
** Corresponding author: kamenskaya.daria@mail.ru*

Key words: *growth hormone genes, nucleotide diversity, gene duplication, Salvelinus, Salmonidae*

Motivation and Aim: Duplication of DNA sequences, especially genes, is one of the main evolutionary factors. The stage of whole-genome duplication is believed to take place in the history of all organisms, and many duplicated genes could originate from it. Subsequently, full or partial diploidization stages occurred in various phylogenetic branches of eukaryotes, and these duplications can be manifested in the form of multiple gene families [1]. Salmonids are a unique group, which developed in the past, after the last autotetraploidization event. Many genes in this taxonomic group were multiple, including the growth hormone (GH) gene [2]. The GH gene is represented by two unrelated paralogous genes, *gh1* and *gh2*, in salmonids' genome. Both genes exist throughout the time of divergence of species in this group. Therefore, salmonids are a suitable model system for investigating the origin, evolution, and functions of duplicated genes. The aim of the current research is to compare paralogous GH genes of three genera in the family Salmonidae–*Salvelinus*, *Salmo* and *Oncorhynchus*–in order to determine their potential differences and possible functions.
Methods and Algorithms: To obtain nucleotide sequences of the GH genes, the following conventional techniques of molecular genetics were used: PCR, electrophoresis, molecular cloning, and sequencing. Nucleotide sequences were analyzed using the MEGA-6.0 and DnaSP-5.10.01 software packages.
Results: A comparison of the complete paralogous GH genes *gh1* and *gh2* of salmonids has shown that the conserved regions are associated with exons, and the peaks of variable regions correspond to intron sequences. It should be noted that not all intron sequences are variable; conserved regions can also be found. The presumable regulatory elements, localized in some introns (Pit-1 motifs, CRE, ERE), are also conserved. We found that *gh1* gene in charrs is more conserved than *gh2* gene; the amount of variability in *gh2* gene is 2–3 times as large as that in *gh1* gene, but it is not so obvious when comparing genes of the all investigated species.
Conclusion: A high conservation of coding sequences (exons) in paralogous GH genes of salmonid fish can be determined by the fact that both genes are functional or probably subfunctional. Both exons and regulatory regions are under the influence of negative selection. But the different rate of changes can be explained by the different selection intensity in paralogous genes.

References:

1. J. Zhang (2003) Evolution by gene duplication: an update, Trends Ecol. Evol., 18: 292–298.
2. F.W. Allendorf, G.H. Thorgaard (1984) Tetraploidy and the evolution of salmonid fishes, In: Evolutionary Biology of Fishes, B.J.Turner (Eds.), 1-Klimenko A.I., Matushkin Yu.G., Kolchanov N.A., Lashin S.A. Bacteriophages affect evolution of bacterial communities in spatially distributed habitats: a simulation study // BMC Microbiology, 2016, Vol. 16 (Suppl. 1), No.1, p. 31.

DINAMIC METBOLIC REGULATION BY A CHROMOSOME SEGMENT FROM A WILD SPECIES DURING FRUIT DEVELOPMENT IN A TOMATO INTROGRESSION LINE

Y. Kanayama
School of Agricultural Science, Tohoku University, Sendai, Japan
** Corresponding author: yoshinori.kanayama.a7@tohoku.ac.jp*

Key words: *metabolome, Solanum lycopersicum, Solanum pennellii, transcriptome*

Motivation and Aim: Tomato (*Solanum lycopersicum*) is one of the most important fruit crops, and useful as a fruit crop model. The genetic variation of cultivated tomato has been narrowed by domestication, and tomato wild relatives with useful genes that have been lost from cultivated tomatoes are useful for breeding. Tomato introgression lines have been developed from a cross between *S. lycopersicum* and wild relative *S. pennellii* for the efficient evaluation and use of useful quantitative trait loci from *S. pennellii*. Each line carries a single *S. pennellii* chromosome fragment in the genetic background of cultivated tomato cv. M82, and a line, IL8-3, was used as a near-isogenic line for analyzing traits associated with fruit quality. Comparative metabolome and transcriptome analyses were performed during fruit development using M82 and IL8-3, with interesting and useful traits such as high sugar content.
Results and Conclusion: Marked differences between M82 and IL8-3 were found in ripe fruit and young fruit at 20 days after flowering (DAF) in the hierarchical clustering analysis of the metabolome, whereas patterns were similar between the two genotypes at 10 and 30 DAF. The metabolome analysis showed that 20 DAF is an important stage for fruit metabolism and that the *S. pennellii* introgressed region in IL8-3 plays a key role in metabolic changes at this stage. Carbohydrate and amino acid metabolism were promoted in IL8-3 at 20 DAF and ripening stage, respectively, whereas transcriptome pattern showed no marked differences between the two genotypes, indicating that dynamic metabolic regulation at 20 DAF and ripening stage was controlled by relatively few genes. The expression of the cell wall invertase and sucrose synthase genes in starch and sucrose metabolic pathway and that of the glutamate synthase gene in the amino acid metabolic pathway in IL8-3 fruit were higher than those in M82. Our results suggested that sugar metabolism activated by the invertase and sucrose synthase in IL8-3 fruit at 20 DAF affects amino acid metabolism and accumulation by higher sugar concentration at the late stage of fruit development.
References:
1. H. Ikeda, T. Shibuya, Y. Kanayama et al. (2016) Dynamic metabolic regulation by a chromosome segment from a wild relative during fruit development in a tomato introgression line, IL8-3. Plant and Cell Physiology, doi:10.1093/pcp/pcw075
2. H. Ikeda, Y. Kanayama et al. (2013) Analysis of a tomato introgression line, IL8-3, with increased Brix content. Scientia Horticulturae, 153:103-108.

BASED ON THE LOCAL SEQUENCE SIMILARITY METHOD FOR PREDICTION OF AMINO ACID POSITIONS RELATED TO THE PROTEIN-LIGAND SPECIFICITY

D.A. Karasev[1, 2]*, A.V. Veselovsky[1], N.Yu. Oparina[3, 4], A.V. Rudik[1], D.A. Filimonov[1], B.N. Sobolev[1]

[1] *Institute of Biomedical Chemistry, Moscow, Russia*
[2] *Russian National Research Medical University, Moscow, Russia*
[3] *Engelhardt Institute of Molecular Biology, Moscow, Russia*
[4] *Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden*
** Corresponding author: w.dmitrykarasev@gmail.com*

Key words: *local sequence similarity, ligand-specific positions, protein kinase inhibitors*

Motivation and aim: Detection of the residues responsible for ligand specificity is applied in protein engineering and discovery of new drug targets. However, many existing methods require the precise superposition of functionally specific residues in aligned sequences. This is not always possible in the case of diverged protein family. We applied original method SPrOS (Specificity Projection On Sequence) to detect amino acid positions associated with ligand specificity of proteins belonging to the same family.
Methods and algorithms: The method SPrOS allows detecting the amino acid residues specific to user-defined groups. SPrOS compares the sequence segments from the studied protein and proteins from the training set. Contrary to other segment-comparison approaches extracting the string motifs, SPrOS calculates the scores for single positions by the similarity of their surroundings. We tested our method on the sequences of protein kinases classified by interaction with small molecular compounds known as the promising leads for drug development.
Results: The prediction of ligand-specific positions was performed on kinases with known 3D structure. The significant specificity estimates were obtained for residues located in ATP-binding cleft, which is a known binding-site for kinase inhibitors. The impact of several found residues is confirmed by the published experimental studies. Filtering out the close homologues of the test protein at the sequence comparisons, we were able to locate specific residues with the more precision.
Conclusions: We showed the applicability of our method for recognition of the amino acid residues associated with ligand specificity. The method was successfully applied to complicatedly partitioned protein family when functional classification differs from phylogeny. Based on inspecting the 3D structures, we suggest that predicted positions determine specific interaction by directly contacts with the ligand molecule.
Available: http://www.way2drug.com/spros/
Acknowledgments: This work is carried out with the financial support of Russian Foundation for Basic Research (grant no. 16-04-00491).
References
1. Karasev D.A. et al. (2016) Prediction of amino acid positions specific for functional groups in a protein family based on local sequence similarity, J Mol. Recognit., 29(4),159-169.

SEQUENCING FROM ROCHE: WHAT THE FUTURE WILL BRING FOR YOU?

I.Y. Karpova*
LCC "Roche Diagnostics Rus", Moscow Russia
** Corresponding author: irina.karpova@roche.com*

Key words: *NGS, high-throughput, SMRT sequencing, targeted sequencing.*

DNA sequencing is used widely to solve of a large set of studies: sequencing de novo of different organisms; targeted gene sequencing and search of mutations; analysis of transcriptomes and methylomes and etc.

High-throughput sequencing has become a routine method and is applied both in science and in medicine. DNA tests are utilized actively at all steps of diagnostics: from search of patients and diagnosis to monitoring of drug action. This approach is used more and more and it will become the same essential part in diagnostics as Real-Time PCR.

The Company Roche based on its experience and knowledge in the field of diagnostic and pharmaceutical industry develops an integrated approach that would allow using NGS in medicine quickly and efficiently. It will include all basic steps: biological sampling, DNA extraction, preparation of libraries, high-throughput sequencing, data analysis and annotation.

Today targeted sequencing of different genes and regions of the human genome is preferred for medical research, because it is much cheaper and more effective than whole genome sequencing. Germline mutations associated with define inherited disease is determined by targeted sequencing. In addition, there are somatic mutations in genome, they can also cause diseases. Not all of targeted sequencing methods are able to identify somatic mutations with high accuracy. Roche offers a new enrichment technology based on amplification. This method overcomes previous limitations of the targeted approach to enrichment by the use of molecular inversion probes. These probes improve capture performance, the detection of alleles as a function of allele frequency in somatic reference samples.

Besides, Roche in collaboration with Pacific Bioscience introduces a new high-throughput sequencer, based on single molecular sequencing in real-time (SMRT). The new sequencer allows to obtain long reads to a few thousand base pairs, as well as to analyze the modified bases without additional manipulation with DNA during sample preparation. SMRT technology is already used widely in science to sequencing small genome de novo, to improve the assembly of large genomes, searching for new splice variants and isoforms of transcripts, the targeted sequencing, etc.

New SMRT sequencer from Roche will be the first step to the active use of NGS testing in the medicine diagnostics.

BIOMOLECULAR SYSTEMS MODELS SEMI-AUTOMATIC RECONSTRUCTION BASED ON STRUCTURAL AND QUANTITATIVE INFORMATION

F.V. Kazantsev[1, 2]*, I.R. Akberdin[1, 5], S.A. Lashin[1, 2], N. Ree[1], V. Timonov[2], A.V. Ratushny[3, 4], T.M. Khlebodarova[1], V.A. Likhoshvai

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Center for Infectious Disease Research, Seattle, USA*

[4] *Institute for Systems Biology, Seattle, USA*

[5] *San Diego State University, San Diego, USA*

Keywords: *Mathematical models; generalized Hill functions; respiration; de novo nucleotide synthesis; gene network, gene expression, database*

Motivation: Living systems have a complex hierarchical organization that can be viewed as a set of dynamically interacting subsystems. Thus, to simulate the internal nature and dynamics of the whole biological system we should use the iterative way for a model reconstruction, which is a consistent composition and combination of its elementary subsystems. In accordance with this bottom-up approach, we have developed MAMMOTh (MAthematical Models of bioMOlecular sysTems) database that allows integrating manually curated mathematical models of biomolecular systems, which are fit to the experimental data. The database entries are organized as building blocks in a way that the model parts can be used in different combinations to describe systems with higher organizational level (metabolic pathways and/or transcription regulatory networks).

Results: The database supports export of single model or their combinations in SBML or Mathematica standards. The database currently contains more than 100 mathematical models for *Escherichia coli* elementary subsystems (enzymatic reactions and gene expression regulatory processes) that can be combined in at least 5100 complex/sophisticated models concerning such biological processes as: de novo nucleotide biosynthesis, aerobic/anaerobic respiration, and nitrate/nitrite utilization in *E. coli*.

Conclusions: The database provide REST API that can be used for programmatic data access and the integration with external software tools. Currently, we have done integration with the biomedb.ru resource. And now, one can create the structural model on biomedb.ru resource and obtain the mathematical model with subsystems extracted from MAMMOTh in a semi-automatic way.

Availability: http://mammoth.biomodelsgroup.ru

Acknowledgements:

This work is partially supported by RFBR grants № 15-07-05889 and № 15-07-03879.

REGULATORY ROLE OF SINGLE CPG METHYLATION

A. Khamis[1], A.V. Artemov[2], A.V. Lioznova[2], V.B. Bajic[1], Y.A. Medvedeva[2]*
[1] *King Abdullah University of Science and Technology*
[2] *Research Center of Biotechnology RAS*

Key words: *epigenetics, methylation, CpG, CAGE*

DNA methylation is a well studied epigenetic process. Several methods have been developed to determine genome methylation at various scales, including those capturing DNA methylation with a single base resolution, such as bisulfite sequencing. Still, for a downstream analysis the most common strategy is to average methylation levels along regulatory regions, based on the assumption of the homogeneous distribution of DNA methylation within genomic regions. Despite the well known observations of unmethylated CpGs co-localized within CpG islands (CGIs) and methylated CpGs co-localized within repetitive elements, the role of single CpG methylation has also been reported. Our previous study demonstrated that a share of gene-proximal CpGs exhibited a significant negative correlation of their methylation profiles with the expression profiles of neighboring genes across various cell types. We called such CpG positions CpG traffic lights. Although they can be co-localized in short CpG-clusters, they are quite often single CpGs. Previously we have demonstrated that CpG traffic lights are unlikely to be widely involved in regulation of transcription factor binding. Now we show that CpG traffic lights are over-represented within enhancers and transcriptional start sites determined by CAGE (Cap Analysis of Gene Expresion). They are as well co-located with regions of histone modifications, supporting their regulatory potential. We also show depletion of SNP in such positions, suggesting the presence of natural selection. We conclude that thought the regulatory role of CpG traffic lights in not completely clear, they can represent regulatory regions and their methylation levels at very least can serve as markers for gene expression.

GENETIC DIVERSITY IN NATIVE SIBERIAN POPULATIONS: CORRELATION WITH CLIMATIC AND GEOGRAPHICAL PARAMETERS

V.N. Kharkov[1, 2]*, A.V. Markov[1, 2], I.Yu. Khitrinskaya[1], V.A. Stepanov[1, 2]
[1] *Research Institute for Medical Genetics, Tomsk, Russia*
[2] *Tomsk State University, Tomsk, Russia*
** Corresponding author: vladimir.kharkov@medgenetics.ru*

Key words: *human populations, genetic diversity, gene pool, adaptation, natural selection, human evolutionary genetics*

Motivation and Aim: Adaptive evolution to adverse or extreme climatic and geographical conditions mediated by natural selection, probably played the substantial role in shaping the genetic structure of modern human populations. We have investigated the distribution of genome-wide SNPs, correlated with climatic and geographic parameters, in native Siberian populations comparing to worldwide human populations in order to detect the natural selection.
Methods and Algorithms: Our data on genome-wide SNPs frequencies in 5 native Siberian populations (Buriat, Yakut, Tuva, Khants, Kets) were pooled with data on worldwide populations and analyzed by means of positional search of association of allele frequencies with climatic and geographic parameters and search for signals of natural selection.
Results: For a considerable number of SNPs allele frequency demonstrate an increase of heterozygosity from tropical to northern populations. The level of genetic diversity and genetic differentiation of these SNPs is significantly different from the average for the genome. Among the genes demonstrating significant selection signals are: EPHA8, GRB2, LINGO2, LINC00669, YES1, CSMD1, DAAM1, DLGAP1, DRD3, KAZN, FRMD4B. Haplotypic tests show traces of natural selection in genomic regions of this genes.
Conclusion: We suppose that genetic diversity in the substantial part of the human genome in native Siberian populations were shaped by adaptation to cold climate.
Acknowledgements This work was supported by Grant of President of Russian Federation (grant № MD-8886.2016.4).

THE INFLUENCE OF RARE MUTATIONS IN THE APOB GENE TO THE LEVEL OF OXIDIZED LDL

E.Yu. Khlebus*, N.V. Shcherbakova, I.S. Zhanin, A.A. Zharikova, A.I. Ershova, A.V. Kiseleva, S.A. Boytsov, A.N. Meshkov
National Research Center for Preventive Medicine? Moscow, Russia
** Corresponding author: elkhlebus@gmail.com*

Key words: *oxidized lipoproteins, rare variant association analysis, APOB*

Motivation and Aim: Atherosclerosis represents one of the major problems in the modern medicine and public health. Cardiovascular disorders, which result from atherosclerosis, are the leading cause of mortality in developed countries. Oxidized LDL (oxLDL) plays an important role in atherogenesis by promoting an inflammatory environment and lipid deposition in the arterial wall. The detection of mutations affecting the level of oxLDL is relevant.

Our work aims to investigate the functional effect of genetic markers affecting oxLDL levels, major risk factor for atherosclerosis, on incidence/onset of atherosclerosis and atherosclerotic phenotypes in a cohort of 725 patients.

Methods and Algorithms: Genotyping of 725 patients was performed using Cardio-Metabo Chip (Illumina) which allows to genotype 196 000 SNPs. Targeted sequencing of genomic region 2p24-p23 was performed with the TargetSeq™ Custom Enrichment Kit (Applied Biosystems, USA) using the SOLiD 5500W system (Applied Biosystems, USA). Alignment and search SNPs were implemented by data analysis tools with the LifeScope™ Genomic Analysis Software.

Results: We performed the genome-wide association study (GWAS) using microarrays Cardio-Metabo Chip (Illumina). We found genetic locus, 2p24-p23, capturing a total number of 14 SNPs in the APOB gene or near it, significantly associated (after adjustment for multiple testing) with levels of oxLDL.

We conduct sequencing of APOB gene and surrounding areas (locus of about 500 000 bp, Chr2: 20996301-21494945) in 96 patients (48 with high levels and 48 with low levels of oxLDL).

ApoB locus was analyzed. For analysis we used data for 725 patients from Cardio-Metabo Chip (Illumina) and for 96 patients from targeted sequencing. We identified single-nucleotide polymorphisms (SNPs) and filtered them to narrow the search for the causal variant. It seems possible that much of the genetic control of common diseases is due to rare and generally deleterious variants that have a strong impact on the risk of disease in individual patients. It is also likely that the variants with the largest effect sizes will be those that have obvious functional consequences. So we focused only on nonsynonymous (protein-altering) changes, other variants have been removed from further consideration. To analyze the impact of the cumulative effect of various rare alleles to the level of oxidized LDL we used R package and statistical tests for rare mutations - CMC (Combined Multivariate and Collapsing Methods) and SKAT (Sequence kernel association tests).

Conclusion: Our results may indicate the influence of functional mutations at the level of oxidized LDL.

It is supported by RFBR #13-04-01259

WHAT WE USUALLY STUDY WHEN WE THINK WE STUDY AGING

A.N. Khokhlov
Evolutionary Cytogerontology Sector, School of Biology, Lomonosov Moscow State University, Moscow, Russia
** Corresponding author: khokhlov@mail.bio.msu.ru*

Key words: *aging, gerontology, death probability, life span, non-aging organisms, cell cultures, systems approach*

The recent considerable increase in interest in experimental gerontological research has led to a paradoxical situation: although the number of studies in this field is ever increasing, only a small part of them actually deals with aging mechanisms. This situation is determined, among others, by the following circumstances. (1) As a rule, the classic definition of aging as a set of age-related changes leading to an increase in the probability of death is ignored. (2) The focus in these studies is on an increase or a decrease in the lifespan, although this had nothing to do with modification of aging (in particular, it is possible to successfully extend the lifespan for non-aging organisms; on the other hand, the very existence of aging does not necessarily suggest a short lifespan). (3) The animals with certain abnormalities (such as genetic diseases) are often used as a control; thus, any favorable impact on the corresponding pathological processes leads to an increase in lifespan. (4) Too much importance is attached to an increase or a decrease in the AVERAGE lifespan, which is in many respects determined by the factors that are in no way associated with aging. (5) The ever increasing number of gerontological experiments involves the model systems that give only indirect information about the aging mechanisms and whose interpretation, in many respects, depends on the basic concept shared by the corresponding researchers. In particular, this refers to the situation with the term “cell senescence.” Initially, this term was introduced to denote various adverse changes in normal cells RESULTING from depletion of their mitotic potential. On the contrary, this term now is ever more frequently used to denote the inhibition of cell proliferation (including cancer cell proliferation) accompanied by a certain cascade of intracellular reactions and caused by various DNA-damaging factors. (6) Finally, there is the issue that may be referred to as the “reductionism problem.” The overwhelming majority of gerontological theories that have appeared during the last decades reduced all the mechanisms underlying both “normal” and modified (accelerated or slowed down) aging of multicellular organisms to certain macromolecular alterations (it is not important whether they are stochastic or programmed) in the constituent cells. This has given rise to numerous cytogerontological model systems for studying the “age-related” changes in the cells freed from a “body-level noise” associated with the functioning of the neurohumoral system. However, many of the conclusions reached earlier on the basis of the results of experiments conducted on Hayflick's model (aging in vitro), were subsequently found to be wrong. In addition, our cytogerontological studies of various anti-aging factors with the help of the "stationary phase aging" model, the cell kinetics model and assessment of colony-forming ability, have shown that in very many cases the factors studied have no beneficial effect on the viability of cultured cells, although they prolong life in experimental animals and increase the well-being of humans. This allowed us to assume that, in many cases, the anti-aging agent action appears only at the organism level, and is not limited to just improving the viability of some of its constituent cells.

THE FIRST EDITION OF MUTAGENESIS BY CRISPR/CAS IN THE EXTREME DESICCATION TOLERANT CULTURED CELL

T. Kikawada[1, 2]*, Y. Miyata[1, 3], Y. Sogame[1, 4], T. Furusawa[1], S. Kikuta[5], R. Cornette[1], O. Gusev[6, 7]

[1] *Institute of Agrobiological Sciences, NARO, Japan*
[2] *Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Japan*
[3] *Center for Biological Resources and Informatics, Tokyo Institute of Technology*
[4] *JSPS Research Fellow*
[5] *Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan*
[6] *Institute of Fundamental Medicine and Biology, Kazan Federal University,Russia*
[7] *Preventive Medicine & Diagnosis Innovation Program (PMI), RIKEN, Japan*
* *Corresponding author: kikawada@affrc.go.jp*

Key words: *anhydrobiosis, cell culture, genome editing, CRISPR-Cas*

Motivation and Aim: We have accomplished draft-genome analysis of the anhydrobiotic midge Polypedilum vanderplanki. From the comparative genomics, several specific genomic regions are likely to be involved in evolutional acquisition of the anhydrobiosis in P. vanderplanki; however, the molecular mechanisms underlying the anhydrobiosis remains to be elucidated. As with the larvae of P. vanderplanki, Pv11 cells, a cultured cell line derived from embryo of P. vanderplanki possesses a capability of the desiccation tolerance. To efficiently screen the responsible genes to the anhydrobiosis, we attempt to optimize the genome editing system, CRISPR-Cas for Pv11 cells.
Methods and Algorithms: Using effective promoters isolated from genome sequence of P. vanderplanki, Cas9 and single guide RNA (sgRNA) expression vectors were constructed. After transfection of the vectors, their expressions were evaluated with Western blot and Real-Time PCR. Finally, we checked phenotypes of the in-del mutated Pv11 cells after sorting the mutant cells.
Results: We have already established a stable Pv11 cells expressing GFP (Pv-KH cells). Exogenous Cas9 and sgRNA expressions in Pv-KH cells were confirmed. To validate knock-out efficiency of P. vanderplanki-optimized CRISPR/Cas9 system, we observed the loss of function in the Pv-KH cells co-transfected with expression vectors for Cas9 and sgRNA that recognized GFP gene. As a result, a small percent of the cells was completely loss of their fluorescence, indicating that CRISPR/Cas9 system could be worked in the anhydrobiotic midge.
Conclusion: The genome editing, CRISPR/Cas9 system can be applicable in P. vanderplanki.
Acknowledgements: This work was financially supported by JSPS KAKENHI (16K18827, 16K15073 and 25252060).

ANHYDRO-PRESERVATION OF EXOGENOUSLY-EXPRESSED DESICCATION-SENSITIVE ENZYME LUCIFERASE USING INSECT CELLS

S. Kikuta[1]*, S. Watanabe[1], O. Gusev[2, 3], Y. Sogame[4], R. Cornette[4], T. Kikawada[4]
[1] *Tokyo University of Agriculture and Technology, Tokyo, Japan*
[2] *Kazan Federal University, Kazan, Russia*
[3] *Riken Division of Genomic Technologies, Japan*
[4] *National Agriculture and Food Research Organization, Japan*
* *Corresponding author: singo@cc.tuat.ac.jp*

Key words: *dehydration, trehalose, luminescence, dry-preservation, anhydrobiosis*

Motivation and Aim: An insect cell line (Pv11) derived from the desiccation tolerant insect, Polypedilum venderplanki can withstand without water. Upon rehydration, the desiccated cells begin to re-start their metabolisms and proliferate. The results indicate that the intracellular enzymes involved in the metabolism should be preserved under the dry state. We would like to know whether Pv11 can preserve exogenously expressed enzymes in dehydrated state., i.e. be a model of new generation of dry preservation technology for enzymes.
Methods: As a representative exogenous enzyme, luciferase in firefly is suitable to examine the activity without taking account of intrinsic proteins. We established Emerald Luciferase-expressing Pv11 (ELuc-Pv11) by DNA electroporation and antibiotics treatment. We measured the luminescence by luciferase of dehydrated ELuc-Pv11 at 1h after rehydration.
Results: The luminescence by luciferase was clearly shown at 1h after rehydration. We concerned about the luciferase activities after rehydration. The activities were a potential to show de novo synthesis in rehydrated Pv11 re-started metabolism. Using protein translation inhibitors, the luminescence was also detected after rehydration. These results indicate that the conformation of the enzyme in the cells would be stably kept under the dehydrated state.
Conclusion: Pv11 can preserve the exogenously-expressed enzyme under dehydration.
Availability: Using this system, we can keep enzymes of interest without a deep freezer.
This work was partially supported by Ministry of Science and Education of RF, research project identification number: RFMEFI58414X0002.

MOLECULAR DYNAMICS CHARACTERIZATION OF GLYCYRRHIZIN INTERACTION WITH LIPID MEMBRANES

A.V. Kim*, E.A. Shelepova, N.N. Medvedev
Institute of Chemical Kinetics and Combustion SB RAS, Novosibirsk, Russia
** Corresponding author kim@kinetics.nsc.ru*

Key words: *glycyrrhizic acid, lipid bilayer, potential of mean force, molecular dynamics*

Motivation and Aim: Glycyrrhizic acid (GA) is a triterpene glycoside extracted from licorice root. It has a wide range of therapeutic activity with a minor amount of side effects. There are also many experimental evidences of its ability to enhance the bioavailability of another drug molecules when used together. But the mechanism of GA biological activity remains unknown. Computer simulation can shed the light on understanding the processes considered, at a molecular level.
Methods and Algorithms: All simulations were performed using the GROMACS 5.0 [1] molecular dynamics package. We used Berger's lipids model for DOPC, POPC and DPPC bilayers. GA parameters were generated by ATBuilder on the basis of gromos53a6 forcefield. Free energy calculations were performed using umbrella sampling approach for potential of mean force (PMF) estimation. 30 windows, spaced by 0.2 nm were employed; each window contained 150 ns of production run, covering a total of 4.5 μs per system.
Results: The series of 10 independent simulations of 200 ns for each of the three lipids were performed. The characteristic behavior of GA nearby and inside the model membrane was revealed. It was found, that GA being placed in the water in random orientation, first diffuses to the membrane surface. Then after moving over the surface for 20-80 ns it meets a suitable cavity and penetrates into the membrane, occupying the region under the lipid heads. In the case of DOPC bilayer GA stays in its first half-layer, not passing through the midplane. To estimate the energy barrier, the PMF calculation was performed for the process of GA penetration through the lipid bilayer. There are two energy barriers observed: the one is at the membrane surface and the other one is in the middle of the bilayer. The midplane barrier is about 3 Kcal/mol, which is approximately 5 times RT. The thermal energy is not enough for GA to pass to the next half-layer of membrane. The partial density profile of GA in DOPC bilayer shows a good agreement with PMF. A steep energy downhill from water to bilayer surface is about 8 Kcal/mol, so GA readily attaches to the membrane and penetrates in it, but does not able to escape it.
Conclusion: GA easily incorporates in membranes and locates under its surface near central parts of lipids tails. It is in a good agreement with experimental NMR studies, where GA demonstrates an influence on the mobility of both the polar heads and the central parts of lipid's hydrophobic tails. PMF shows a preference of GA to stay in the first half-layer of DOPC membrane due to an energy barrier of about 8 Kcal/mol. This fact is also in a good agreement with NMR results.
References:

1. S. Pronk et al. (2014) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit, Bioinformatics, 29(7):845–854.

GENOMEASIA 100K INITIATIVE ANNOUNCED TO SEQUENCE 100,000 GENOMES IN SOUTH, NORTH AND EAST ASIA

Hie Lim Kim, Elena S. Gusareva*, S.C. Schuster
Singapore Centre on Environmental Life Sciences Engineering, Singapore
Nanyang Technological University (NTU), Singapore
** Corresponding author: egusareva@ntu.edu.sg*

Key words: *GenomeAsia 100K, genetic diversity, Asian populations*

With recent insights into the genome diversity of Asian ethnicities, it is very important to understand the biology of disease in the currently under-studied Asian populations that represent 40% of mankind. The unique genetic diversity prevalent in South, North and East Asia can potentially provide a valuable source of clinical insights that should enhance our understanding of several rare and inherited diseases, as well as complex diseases such as cancer, diabetes and cardiovascular disease.

The non-profit consortium, GenomeAsia 100K, announced an ambitious plan to sequence 100,000 individuals from Afghanistan, Bangladesh, Bhutan, Burma, Brunei, Cambodia, China, India, Indonesia, Japan, Kazakhstan, Korea, Kyrgyzstan, Laos, Maldives, Malaysia, Mongolia, Pakistan, Philippines, Russia, Sri Lanka, Taiwan, Tajikistan, Thailand, Timor-Leste, Turkmenistan, Uzbekistan, and Vietnam. In the first phase, the project will focus on creating reference genomes for each of diverse Asian ethnic groups representing a major step forward in understanding the population history and substructure of the region. The initiative aims 1) to generate ethnicity-specific reference genomes to comprehend Asian genomes and identify rare genetic variants and structural variations with high accuracy in the populations, 2) to develop a database of allele frequency with defined ethnicities to share with academic communities, and 3) to develop insights into disease and human health by combining information from genome, clinical history and population substructure. Focus on inherited diseases, common diseases, oncology, metabolic disorders, neurology, and aging to promote precision medicine research.

The GenomeAsia 100K hosted at Nanyang Technological University in Singapore. To date, 50,000 DNA samples have been collected through blood or saliva samples from a network of clinics across Asia with the help of two of the consortium's founding members, genomics companies Macrogen in South Korea and MedGenome in India.

The GenomeAsia 100K is looking for significant partners and supporters in Asian countries and beyond.

References:

1. GenomeAsia 100K Initiative Announced to Sequence 100,000 Genomes in South, North and East Asia. Feb 11, 2016 News from Emerge Ventures.

THE SIGNIFICANCE OF DISSOCIATIVE NUCLEOTIDE CHANGES ACCUMULATION RATE IN THE GENOTYPE VARIABILITY OF TICK-BORNE ENCEPHALITIS VIRUS FOR GENE E

D.O. Kiselev[1]*, S.Ju. Bukin[2], A.I. Paramonov[3], Ju.P. Dzhioev[1], V.I. Zlobin[1]
[1] *Irkutsk State Medical University, Irkutsk, Russia*
[2] *Limnological Institute, Irkutsk, Russia*
[3] *Scientific Centre for Human Reproduction and family health problems, Irkutsk, Russia*
* *Corresponding author: Rancevich89@mail.ru*

Key words: *tick-borne encephalitis virus, molecular epidemiology, molecular clock*

Motivation and Aim: The mechanism of genetic molecular clock forms structural and functional features of all living systems, including viruses. The aim of research is the study of molecular clock in the evolutionary variability of the tick-borne encephalitis virus (TBEV) for gene E by methods of molecular genetics and bioinformatics that can be used to obtain the important information about the variability and evolution of virus. The material of the study is based on data about sequences of gene E of TBEV (55 strains), most of which were isolated in the Siberian region, as well as strains isolated in other regions.
Methods and Algorithms: MEGA6 – algorithms (neighbor-joining trees, Tajima Relative Rate Test of Molecular Clock) [1,2]
Results: Two phylogenetic trees are presented in the study (for 1-3 and 1-2 positions of codons of gene E). Significant differences of phylogeny structure have been found between two trees. The verification results of the strict molecular clock hypothesis efficiency are presented in the second part of study for selected TBEV strains. Selected strains have been divided into 2 subgroups with different geographical origin of strains according to the rate of nucleotide substitutions accumulation.
Conclusion: Results of the study demonstrate the high significance of the nucleotide substitutions accumulation in the 3-rd position of codons in the evolutionary history of TBEV, making significant adjustments to the process of studying the phylogeny and phylogeography of the pathogen.
References:
1. Tamura K. et al. (2013) Molecular Biology and Evolution 30:2725-2729
2. Tajima, F. (1993) Simple methods for testing molecular clock hypothesis. Genetics, 135, 599–607.

FUNCTIONAL AND STRUCTURAL CHARACTERISATION OF *PPD-B1* PHOTOPERIOD INSENSITIVE ALLELE

A.A. Kiseleva*, E.A. Salina
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: antkiseleva@bionet.nsc.ru*

Key words: *wheat, Triticum aestivum, photoperiod sensitivity, heading date, Ppd-B1*

Motivation and Aim: Photoperiod sensitivity is an important agronomic trait that influences wheat heading date. *Ppd-1* genes are significant regulators of this process. *Ppd-1* photoperiod insensitive alleles induce early heading of wheat. *Ppd-B1* is the only *Ppd-1* gene which dominant photoperiod insensitive allele is determined by the copy number variation. However, little is known about mechanisms determining its misexpression. The aim of our investigation was to reveal possible mechanisms of the photoperiod pathways involved *Ppd-B1*, to characterize functional specifications of *Ppd-B1* photoperiod insensitive allele and its interaction with other photoperiod genes.
Methods and Algorithms: PlantPAN 2.0 database (http://plantpan2.itps.ncku.edu.tw) was used to determine putative plant transcription factor binding sites. PCR analysis, molecular cloning and sequencing were used for the characterization of *Ppd-B1* sequence. To analyze diurnal expression of genes regulating heading date Real-time PCR was performed.
Results: We identified probable transcription factors involved in *Ppd-B1* regulation and factors common for the promoters of all *Ppd-1* homeologous genes. Promoter regions of such important genes regulating heading date as *TaFT1*, *PhyC* and *Vrn-1* were analyzed too. Using two pairs of Near Isogenic Lines (NILs) with dominant or recessive allele of *Ppd-B1* and different in their photoperiod sensitivity we investigated structure of *Ppd-B1* distinct copies and detected some SNPs confirmed difference between NILs and their sibs, the indel in promoter region distinguished the lines under investigation from other alleles with copy number increment, but revealed no polymorphisms between *Ppd-B1* gene copies. Then we analyzed diurnal expression of *Ppd-B1, Ppd-D1, Ppd-A1, Vrn-A1, TaFT1, PhyC* and some other genes important for the heading date.
Conclusion: We detected some TFBSs specific to the *Ppd-B1* promoter region but not *Ppd-D1, Ppd-A1* allows suggesting different regulation of this genes. Taken together our data about transcription factor binding sites in the promoter regions of genes controlled heading date and the analysis of their diurnal expression suggest hypothetic scheme of their interaction and the impact of *Ppd-B1* photoperiod allele on heading promotion.
Acknowledgements: This study is supported by Russian Scientific Foundation (14-14-00161)

PHAGE INFECTION SLOWS DOWN SPECIATION CAUSED BY GENE LOSS AND HORIZONTAL GENE TRANSFER OF METABOLIC GENES IN MODELS OF SPATIALLY DISTRIBUTED BACTERIAL COMMUNITIES

A.I. Klimenko*, Yu.G. Matushkin, N.A. Kolchanov, S.A. Lashin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: klimenko@bionet.nsc.ru*

Key words: *modeling, microbial communities, horizontal gene transfer*

Motivation and Aim: Bacteriophages are known to be one of the driving forces of bacterial evolution. Besides promoting horizontal gene transfer between cells, they may induce directional selection (for instance, according to more or less resistance of cells to phage infection). However, the impact of phages on metabolic evolution and formation of community trophic structure remains obscure. Spatial organization of the environment is another factor that is crucial for a bulk of processes in the corresponding microbial community including its evolution and infection patterns. We have simulated and analyzed a series of computer models of microbial communities evolving in spatially distributed habitats under the pressure of phage infection.
Methods and Algorithms: We used a multilayer simulation tool HEC 3D [1] taking into account genetic, metabolic, cellular, population, and ecological levels of community organization. It simulates both high-level (cellular chemotaxis and diffusion, substrates flow and diffusion) and low-level (mutations, horizontal gene transfer, gene regulation and metabolism) processes allowing combining various mathematical modeling approaches (agent-based modeling, differential equations, automata etc.) in one model.
Results: We modeled evolving microbial communities living in spatially distributed aquatic habitats characterized by a nutrient gradient. We varied time and location of initial phage infestation as well as switched chemotaxis on and off. Simulations have shown that phage infection decreases the speciation rate by more than one order as far as intensified selection blocks the origin of novel viable populations/species, which could carve out potential ecological niches. The dependence of speciation rate on the invasion node location varied on the invasion time corresponding to different stages of community formation.
Conclusion: Our study has shown that phage infection affects evolution of microbial community slowing down speciation caused by gene loss and horizontal gene transfer of metabolic genes and stabilizing the system as a whole [1]. This influence varied in its magnitude depending on spatially-ecological factors as well as community state at the moment of phage invasion.
Acknowledgements: The study has been partially funded by the RFBR grant 150703879 and Budget Project 0324-2015-0003.
References:
1. Klimenko A.I., Matushkin Yu.G., Kolchanov N.A., Lashin S.A. Bacteriophages affect evolution of bacterial communities in spatially distributed habitats: a simulation study // BMC Microbiology, 2016, Vol. 16 (Suppl. 1), No.1, p. 31

HAPLOID EVOLUTIONARY CONSTRUCTOR 3D: A FRAMEWORK FOR MULTILAYER MODELING OF SPATIALLY DISTRIBUTED MICROBIAL COMMUNITIES

A.I. Klimenko, Yu.G. Matushkin, Z.S. Mustafin, A.D. Chekantsev, R.K. Zudin, S.A. Lashin*

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: lashin@bionet.nsc.ru*

Key words: *modeling, microbial communities*

Motivation and Aim: Next-generation ecological modeling approaches are based on the principles of emergence, predictability and structural realism [1]. Nowadays it becomes clear that to make reasonable ecological models it is necessary to take into account not only species interrelations but also their microevolution and organisms' response to environment heterogeneity and its dynamic nature. It also tends to use functional types rather than taxonomical units in ecological models. We find microbial communities to be promising objects to manifest the power of such next-generation approaches. On the one hand, a community is a strictly localized spatially distributed structure where spatial localization of cells of different functional types determines the variety its metabolic functions. On the other hand, microevolution of microbes runs relatively fast and may be observed in both natural and experimental conditions [2].
Methods and Algorithms: In this study, we present a software package Haploid Evolutionary Constructor 3D (HEC 3D) [3] designed for modeling and simulating spatially distributed multispecies microbial communities. The HEC 3D specifies a model on several layers of biological organization, namely, on genetic, metabolic, cellular, population, and ecological ones. Using HEC 3D one may combine various mathematical modeling approaches (agent-based modeling, differential equations, automata etc.) in one model to simulate such processes as cellular chemotaxis, substrates flow and diffusion, mutations, horizontal gene transfer, gene regulation, reproduction and metabolism. We provide the import of SBML models into the HEC 3D model (metabolism layer) as well as the graphical user interface and high-performance calculations on various platforms.
Results: Several already published HEC 3D models (for example, in [3]) show the close interrelations between evolutionary and ecological processes occurred in microbial communities, where spatial organization of a community may predetermine the evolutionary scenarios manifesting during its whole lifetime.
Conclusion: HEC 3D allows building comprehensive models of microbial communities taking into account both high-level and low-level processes.
Availability: http://evol-constructor.bionet.nsc.ru
Acknowledgements: The study has been partially funded by the RFBR grant 150703879 and Budget Project 0324-2015-0003.
References:
1. V. Grimm, U. Berger (2016) Structural realism, emergence, and predictions in next-generation ecological modelling: Synthesis from a special issue, Ecological modelling, 326: 177-187.
2. R. Kassen (2014) Experimental evolution and the nature of biodiversity (Roberts).
3. A.I. Klimenko et al. (2016) Bacteriophages affect evolution of bacterial communities in spatially distributed habitats: a simulation study, BMC Microbiology, 16: 31.

MONOTROPA HYPOPITYS WHOLE GENOME AND TRANSCRIPTOME SEQUENCING DATA

E.Z. Kochieva*, E.V. Gruzdev, A.V. Beletsky, A.M. Mazur, A.V. Shchennikova, O.V. Shulga, M.A. Filyushin, V.V. Kadnikov, A.V. Mardanov, N.V. Ravin, K.G. Skryabin
Institute of Bioengineering, Research Center of Biotechnology RAS, Moscow, Russia
** Corresponding author: ekochieva@yandex.ru*

Key words: *parasitic plants, pinesap, genome sequencing, transcriptome sequencing, NGS*

Monotropa hypopitys L. (Monotropaceae), commonly known as pinesap, is a flowering perennial non-photosynthetic mycoheterotrophic plant depending on carbon compounds obtained via fungus linkages to autotrophic host plants. *M. hypopitys* can be a plant model to study the transition to parasitism that is accompanied by complex morphological and developmental changes, including loss of photosynthetic ability and the emergence of specific functions required for host interaction.

Here we present the data on the whole-genome and transcriptome sequencing using high throughput GS FLX pyrosequencing and Illumina technology. *M. hypopiys* genome sequencing result in 281 million reads (86 GB), after primer trimming, quality trimming and paired-end reads merging. The estimated genome size is about 2,5 Gb. *M. hypopitys* transcriptome sequencing result in a total of 103 million high quality sequencing reads after primer and quality trimming. Transcriptome was assembled into 98350 unigenes ranging from 201 to 12993bp in length. 37977 unigenes were annotated in the TrEMBL protein database using predicted protein sequences and 38419 unigenes were annotated in the Swiss-Port database using blastx. Out of these, 34385 unigenes were assigned to gene ontology categories.

M. hypopitys chloroplast genome reveals dramatic reduction to 34,8 kbp in size and extensive gene order rearrangement. cpDNA contains only 40 intact genes of ribosomal apparatus and tRNAs. All genes related to photosynthetic activity are absent or became pseudogenes.

Genome-wide characterization of micro RNA resulted in identification of 55 members belonging to 40 families of known miRNAs and 17 putative novel miRNAs unique for *M. hypopitys*. Computational screening revealed 206 potential mRNA targets for known miRNAs and 31 potential mRNA targets for novel miRNAs. The predicted target genes were found to be involved in a broad range of metabolic and regulatory pathways. The identification of novel M. hypopitys-specific miRNAs suggests their recent evolutionary origin and participation in highly specialized regulatory mechanisms fundamental for non-photosynthetic biology of *M. hypopitys*.

This work was supported by the Russian Science Foundation under Grant 14-24-00175.

THE INTERACTION BETWEEN ANAEROBIC RESPIRATORY COMPLEX II AND THE FLAGELLAR MOTOR

A. Koganitsky[1]*, T. Dadosh[2], V. Kiss[1], M. Eisenbach[1]
[1] *Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel*
[2] *Chemical Research Support, Weizmann Institute of Science, Rehovot, Israel*
** Corresponding author: anna.koganitsky@weizmann.ac.il*

Key words: *modeling, flagellar motor*

E. coli is well equipped for living in rapidly changing conditions of its natural ecosystem: it has flagella that enable it to move and navigate in nutrient-poor environments, and when oxygen is limited, it has the ability of switching to anaerobic respiration, using fumarate reduction for making energy. Recently it was discovered that there is a crosstalk between the fumarate reduction system (FRD) and flagellar motility: during fumarate respiration the flagellar motor operates in different mode. Our results reveal that the two systems communicate via physical interaction between their protein components and shed light on the molecular mechanism by which FRD controls the output of the flagellar motor. Moreover, super-resolution microscopy enabled imaging and quantitative analysis of FRD proteins interacting with the flagellar motors *in-vivo*.

VIRTUAL BIOLOGY – THE FOUNDATION

F.A. Kolpakov[1, 2]
[1] *Institute of Systems Biology Ltd., Novosibirsk, Russia*
[2] *Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
** Corresponding author: fedor@biouml.org*

Key words: *systems biology, virtual biology*

Motivation and Aim: Mathematical models of many biological systems and their parts are being intensively developed now. Titles for many of many of them (or embracing projects) highlight they implication to virtual world: virtual cell, virtual liver, virtual physiological rat, virtual physiological human, virtual patient, etc. Whether we can try to cover all these projects by one term?

Results: Yes, virtual biology is a good candidate for this purpose. So here we introduce the concept of virtual biology. It's ultimate goal is building of virtual world inhabited with mathematical models providing precise portrait modeling of biochemistry and physiology of living organisms of real world.

To define precise meaning of virtual biology we are developing a top level ontology using MediaWiki technology. Wiki technology allows a community to refine and maintain the ontology to represent a consensus view.

The core of all systems -ology sciences are mathematical models that should reproduce dynamics of corresponding biological systems. Thus virtual biology can be considered as an umbrella for systems -ology sciences: systems biology, systems physiology, systems medicine, systems pharmacology. Roughly they correspond:

- systems biology - virtual cell;
- systems physiology - virtual human and physiological models of other high eukaryote organism, for example virtual rat;
- systems medicine, systems pharmacology - virtual patient.

One of the key points of virtual biology is automated building of mathematical models using information from specialized biological databases and high throughput technologies. The use of multiple ontologies for defining components and subcomponents of models could allow them to be compared and integrated to form composite models in an automated manner.

The most suitable approach to simulate virtual organisms in virtual world is agent based modeling.

BioUML platform provides the most essential possibilities needed for virtual biology:

- automated building of mathematical models (for example virtual cell) using information from specialized biological databases and high throughput experiments;
- different simulation approaches (ODE, ADE, stochastic modeling) as well as composite modeling;
- parameters estimation to fit some model to experimental data, particularly it allows to personalize physiological models and create physiological avatars for virtual world;
- agent based modeling to simulate virtual organisms in virtual world.

Availability: http://wiki.biouml.org/index.php/Category:Virtual_biology

References:

1. Z.S. Syed et. al. (2008) Wikitology: Using wikipedia as an ontology. In: *Proc. of 2 international conference on Weblogs and Social Media*.
2. BioUML platform – http://wiki.biouml.org
3. http://virtual-biology.org

WHEATDB2: PLANT TRAIT DATABASE AND INFORMATION SYSTEM BASED ON CROP ONTOLOGY TERMS

E.G. Komyshev[1]*, M.A. Genaev[1], A.V. Akushkina[2], D.A. Afonnikov[1, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State Agrarian University, Novosibirsk, Russia*
[3] *Novosibirsk National Research State University, Novosibirsk, Russia*
[*] *Corresponding author: komyshev@bionet.nsc.ru*

Key words: *extreme climatic conditions, database, plant selection, crop, ontology*

Motivation and Aim: Extreme weather conditions such as hot summer or cold winter frequent in many regions of Russia and have a negative impact on crop yields. Increasing grain yields can be achieved by identifying those plant varieties which are most suitable for cultivation in certain target areas. Thereby a systematic collection of data on the genotypes and phenotypes of plants in different climatic conditions is required. This requires developing integrated databases that collect large scale data on phenotype, genotype and environment. An effective approach to solve this problem is to use ontology terms together with a detailed documentation of an experimental protocol to fix important information such as a condition and location of the experiment, used measurement tools. We present a database of specific plant traits based on Crop Ontology ontology terms [1] with a reference to genetic data: varieties/lines, genetic markers, sequences.
Methods and Algorithms: The basic structural unit of the database structure is the "experiment" that combines information on the protocol of the experiment, its conditions, measurement tools, methods and participants. Experiment data types can be bound to the terms of ontology Crop Ontology. As a database management system was selected the MongoDB [2] that allow to represent a variety of flexible data structure in single database.
Results: A user can save the data in structures in a way which is convenient for him. The data can be selected from a number of various experiments and used for comparative analysis to identify the most suitable crop for cultivation in certain climatic conditions.
Conclusion: A proposed database structure is suitable for storing information such as a condition and location of the plant breeding experiment, used measurement tools.
Availability: Wheatdb2 system is available at http://pixie.bionet.nsc.ru/wheatdb2
Acknowledgements: The work supported by RFBR grants №16-37-00304, №16-34-00688 and budget project № 0324-2015-0003.
References:
1. Shrestha R. et al. (2010) Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature, *AoB plants,* 2010: plq008.
2. Chodorow K. (2013) MongoDB: the definitive guide, In: *O'Reilly Media, Inc.*

ASSESSMENT OF TRANSLATIONAL IMPORTANCE OF MAMMALIAN MRNA SEQUENCE FEATURES BASED ON RIBO- AND MRNA-SEQ DATA

Yu.V. Kondrakhin[1, 2], R.N. Sharipov[1, 2], O.A. Volkova[3]*
[1] *Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
[2] *BIOSOFT.RU, Ltd, Novosibirsk, Russia*
[3] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
* *Corresponding author: ov@bionet.nsc.ru*

Key words: *mammalian mRNAs; translation efficiency; ribosome profiling; Ribo-Seq; mRNA-Seq; lncRNAs; regression analysis; discriminant analysis*

Motivation and Aim: Ribosome profiling technology (Ribo-Seq) allowed to highlight more details of mRNA translation in cell and get additional information on importance of mRNA sequence features for this process. But in spite of the progress in study of translation process, there are still many open questions regarding to mRNA and non-coding RNA nucleotide sequence features and influence of that features on translation efficiency (TE). The goal of our study was to assess the influence of mRNA sequence features on translation process. In particular, (1) their ability to discriminate between translated RNAs and lncRNAs; (2) their influence on discrimination between high- and low-translated mRNAs; (3) their relationships with TE.
Methods and Algorithms: We assessed the translational importance of mRNA sequence features with the help of statistical analysis of Ribo- and mRNA-Seq data derived from public databases. 14 translationally important features (10 known from literature as well as four proposed by the authors) were used in analysis. TE values for each transcript were computed as lg(TE) = lg((R1/L1)/(R2/L2)), where R1 – Ribo-Seq read counts for transcript, L1 – length of transcript CDS, R2 – mRNA-Seq read counts for transcript, L2 – length of transcript; R1, R2 $\geq$ 10. To convert TE values onto a common scale for all considered subsamples, we used the following normalization: lg(TE) - mean(lg(TE)), where mean(lg(TE)) was arithmetical mean within the individual subsample. Discriminant analysis has been used for comparison of protein-coding RNAs and lincRNAs as well as for comparison of protein-coding RNAs with high and low TE. Discriminant and regression analyses have been used to study relationships between these mRNA features and TE.
Results: (1) Statistical classification models that discriminated protein-coding and long non-coding RNAs with high accuracy were built. (2) The studied mRNA features were able to discriminate high- and low-translated mRNAs with good accuracy. (3) Built regression models demonstrated that all features were significantly related to TE. (4) Four novel mRNA features proposed by us, namely complementarity index (CI) [1, 20], CI[-15, 1], motif CAAGAA in [3, 103], and motif CCGCCA in [-100, 0] were found to be useful along with known features for discrimination between mRNAs and lincRNAs as well as for description of TE.
Availability: Tool for analysis is available in the frames of the BioUML platform (http://www.biouml.org).
Acknowledgements: Supported by the Russian Foundation for Basic Research (№14-04-01284).

MACROEVOLUTIONARY AND EXPERIMENTAL ASSAYS OF FITNESS LANDSCAPES

F. Kondrashov
Centre for Genomic Regulation (CRG), Barcelona, Spain
** Corresponding author: fkondrashov@gmail.com*

Key words: *evolution, fitness landscape, genotype*

Why do some genotypes lead to fit phenotypes while others do not? More importantly, is it possible to create a set of rules that can predict which genotypes will confer phenotypes of high fitness? The answer to these questions may lie in our ability to experimentally describe the fitness landscape, which is a representation of the fitness state of all genotypes. In my talk I will focus on three things. First, I will review the evidence obtained by numerous computational studies of the character of the fitness landscape of individual genes that span macroevolutionary timescales. Second, I will review recent data from our lab, as well as from other researchers, on the experimental quantification of local fitness landscapes, or the fitnesses of very similar genotypes. Finally, time permitting, I will describe recent results of a long-term experiment that documents the fitness states of genotypes of a specific gene from baker's yeast incorporating various amino acid states that match states found in orthologous sequences.

IT ANALYSIS OF CORNEA ENDOTHELIUM TRANSPORT ABILITY IN CORNEAL TRANSPLANTS AFTER HYPOTHERMIC CONSERVATION

A.A. Konev[1], I.G. Palchikova[1], I.A. Iskakov[2], L.E. Katkova[3], G.S. Baturina[3], E.I. Solenov[3]*

[1] *Technological Design Institute of Scientific Instrument Engineering SB RAS*

[2] *Multidisciplinary Science and Technology Complex "Eye Microsurgery" named after S.N. Fyodorov Federal State Institution, Novosibirsk Branch, Novosibirsk, Russia*

[3] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

* *Corresponding author: eugsol@bionet.nsc.ru*

Key words: *analitic software; cornea endothelium transport; hypothermic conservation*

Background: Despite advances in surgery, corneal edema is a usual complication after cornea transplantation. Removing of extra fluid from the swelled cornea tissue is function of corneal endothelium. Little is known about the influence of hypothermic conservation on ion and water transport in corneal endothelial cells. Cell transport mechanisms that provide cell osmotic balance are determined, in a great extent, by Na,K-ATPase pumping capacity. The intracellular concentration of sodium ([Na+]i) is result of balance between sodium entering the cell and outward current determined by Na,K-ATPase. The main purpose of hypothermic conservation is to save transport ability of endothelial cells in corneal transplants. The methods of transplants conservation are subject of intense studies nowadays.

Aims: The aim of the present work is development of effective method for analysis of corneal endothelial cells transport function by creating and using experimental protocol for measuring of [Na+]i and analytical software for automatic computer analysis of experimental data.

Methods: Intracellular sodium concentration was measured using as specific intracellular fluorescent probe for sodium Sodium Green (MolProbes, USA) by the method developed in the group of cell molecular physiology (Institute of Cytology and Genetics SB RAS). Microscopic fluorescent images were captured with AxioCam HSm (Zeiss, Germany) and stored on PC. The intensity of cell fluorescence was evaluated by analysis of microscopic images of cells using software developed in TDISIE, SB RAS: program "CytoDynamics" (state registration number 2016612766).

Conclusions: The kinetics of intracellular sodium concentration was measured in endothelial cells from isolated eyes of rat, pig and human. Experimental protocol was developed by the group of cell molecular physiology (Institute of Cytology and Genetics SB RAS). The software and user friendly interface was developed for automatic outline fluorescent images of cells as AOI and analysis of fluorescence intensity. The software was registered as "CytoDynamics" (state register number 2016612766). The program analyses consequences of microscopic images of living cells that undergone experimental protocol and calculates the values of [Na+]i. The results of calculations reflect the dynamics of [Na+]i in individual cells. Program uses libraries OpenCV / JavaCV, it accepts colors or monochrome images and presents the results as Excel tables and plots.

Complex approach includes experimental protocol for experimental measurement of [Na+]i in endothelial cells and analytical software enables analyze large scale experimental data sets. The program "CytoDynamics" could be useful as analytical tool for transplantology and drug discovery.

INVASIVE ENTOMO-MYCOLOGICAL ASSOCIATION OF *P. PROXIMYS* AND ITS PHYTOPATOGENIC SYMBIONT IN SIBERIA AND EUROPEAN PART OF RUSSIA

A. Kononov[1]*, A. Blinov[1], N. Pashenova[2], N. Percova[2], Y. Baranchikov[2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *V.N. Sukachev Institute of Forest SB RAS, Krasnoyarsk, Russia*
* *Corresponding author: kononov.nov@mail.ru*

Key words: *Polygraphus proximus; Ophiostoma; co-evolution; phytopathogens*

Motivation and Aim: Polygraphus proximus is a common and economically significant beetle-pest. This species has a close symbiotic relationship with a number of plant pathogenic fungi. The fungi are a cause of tree death. Members of Ophyostomataceae family are one of the most common fungi associated with xylophagous. *P. proximus* have a symbiotic relationship with the 11 fungi species of this family on the territory of its original range, in particularly with aggressive species such as *Grosmannia aoshima*. Recently *P. proximus* significantly expanded its range from relatively small areas in the eastern part of Eurasia to the European continent. In this work we studied fungal symbionts of *P.proximus* in the original and invasive populations.
Methods and Algorithms: The mycelium of fungi were collected in the original and invasive populations. For samples ITS2 and LSU markers were amplified by PCR using specific primer pairs, then sequencing was done. Multiple sequence alignment was performed by MUSCLE algorithm. Analysis of phylogeny by maximum-likelihood was carried out.
Results: 63 samples were investigated from the east, central and western part of the Russia. All the fungi involved in the morphological and genetic analysis. It revealed that there are *Ophyostoma. nikkoense*, *O. piceae*, *O. microcarpum*, *Leptographium taigense*, *G. penicillata* and *G. aoshimae* in the Far Eastern populations of beetles. *L. procerum*, *O. subalpinum*, *G. aoshimae* species were found in Siberian populations. *G. aoshimae* have been found in the European populations. In addition, 10 previously undescribed fungi were found. Obtained sequences took part in the phylogenetic analysis. According to the phylogenetic data fungi belonging to the *Ophiostoma* and *Grosmannia* genera of Ophyostomataceae family, fungi of Bionectriaceae family and even Basidiomycetes were found among indeterminate samples.
Conclusion: Significant diversity of fungal symbionts of *P.proximus* was confirmed in the original range. Association of *P.proximus* with aggressive phytopathogen - *G. aoshima* was established both in the original population and the recently occupied territories. Genetic uniformity of *G. aoshima* throughout the range shows that *P.proximus* transferred this symbiote in the invasive populations and saved the association with *G. aoshima* on the new territories. Based on the received information the phylogenetic relationships of the samples and their position on the phylogenetic tree were established. The number of previously undescribed organisms was studied.
The conclusions obtained in this study show us *P.proximus* and its symbionts as a united biological system wherein each element is closely associated with other elements. Bark beetle linked to fungi biologically, ecologically and even evolutionarily. Co-evolution of these species is reflected in a similar population dynamics.
Availability: All sequences will be available in the genebank NCBI.

GENETIC DIVERSITY AMONG EIGHT *DENDROLIMUS* SPECIES IN EURASIA (LEPIDOPTERA: LASIOCAMPIDAE) INFERRED FROM MITOCHONDRIAL COI AND COII, AND NUCLEAR ITS2 MARKERS

A. Kononov[1]*, K. Ustyantsev[1], B. Wang[2], V. Mastro[2], V. Fet[3], A. Blinov[1], Y. Baranchikov[4]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *USDA-APHIS-PPQ CPHST, Otis Laboratory, Otis Air National Guard Base, Buzzards Bay, USA*
[3] *Department of Biological Sciences, Marshall University, Huntington, USA*
[4] *V.N. Sukachev Institute of Forest SB RAS, Krasnoyarsk, Russia*
** Corresponding author: kononov.nov@mail.ru*

Key words: *Dendrolimus; phylogeny; pests; interspecific hybridization; divergence*

Motivation and Aim: Moths of genus *Dendrolimus* (Lepidoptera: Lasiocampidae) are among the major pests of coniferous forests worldwide. Taxonomy and nomenclature of this genus are not entirely established, and there are many species with a controversial taxonomic position. We present a comparative evolutionary analysis of the most economically important *Dendrolimus* species in Eurasia.
Methods and Algorithms: Larvae and moths of *Dendrolimus* were collected in the natural populations across Asian Russia. For samples COI and COII markers, about 1400 and 600 bp long, correspondingly, and ITS2 sequences (400 bp) were amplified by PCR using specific primer pairs, then PCR purification and sequencing was done. Multiple sequence alignment was performed by MUSCLE algorithm. Analysis of phylogeny by maximum-likelihood was carried out in PhyML 3.0 program with default settings and with the aLRT as a topology estimation method.
Results: Our analysis was based on the nucleotide sequences of COI and COII mitochondrial genes and ITS2 spacer of nuclear ribosomal genes. All known sequences were extracted from GenBank. Additional 112 new sequences were identified for 28 specimens of *D. sibiricus*, *D. pini*, and *D. superans* from five regions of Siberia and the Russian Far East to be able to compare the disparate data from all previous studies. In total, 528 sequences were used in phylogenetic analysis. Two clusters of closely related species in *Dendrolimus* were found. The first cluster includes *D. pini*, *D. sibiricus*, and *D. superans*; and the second, *D. spectabilis*, *D. punctatus*, and *D. tabulaeformis*. Species *D. houi* and *D. kikuchii* appear to be the most basal in the genus.
Conclusion: Genetic difference among the second cluster species is very low in contrast to the first cluster species. Phylogenetic position *D. tabulaeformis* as a subspecies was supported. It was found that *D. sibiricus* recently separated from *D. superan*. Integration of *D. sibiricus* mitochondrial DNA sequences and the spread of this species to the west of Eurasia have been established as the cause of the unjustified allocation of a new species: *D. kilmez*. Our study further clarifies taxonomic problems in the genus and gives more complete information on the genetic structure of *D. pini*, *D. sibiricus*, and *D. superans*.
Availability: The nucleotide sequence from the present work published in the GeneBank under the accession numbers: KJ007736 - KJ007819.

VRN1 GENES VARIABILITY IN TETRAPLOID WHEAT SPECIES WITH A SPRING GROWTH HABIT

I. Konopatskaia[1, 2]*, V. Vavilova[1], E.Ya. Kondratenko[1], A. Blinov[1], N.P. Goncharov[1, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Novosibirsk State Agrarian University, Novosibirsk, Russia*
* *Corresponding author: sormacheva@bionet.nsc.ru*

Key words: *evolution, growth habit, Triticum, vernalization, VRN1 gene, wheat*

Motivation and Aim: Vernalization genes *VRN1* play a major role in the transition from vegetative to reproductive growth in wheat. The floral activator *VRN1* encodes a MADS-box transcription factor that is required for the initiation of reproductive development at the shoot apical meristem. The expression of *VRN1* occurs at a low basal level but a measurable increase is seen during prolonged treatment with low temperatures. In di-, tetra- and hexaploid wheats the presence of a dominant allele of at least one *VRN1* gene homologue (*Vrn-A1*, *Vrn-B1*, *Vrn-G1* or *Vrn-D1*) determines the spring growth habit. Allelic variation between the *Vrn-1* and *vrn-1* alleles relies on mutations in the promoter region or the first intron. The evolution of spring cultivars of wheats from winter ancestors is a key event in the post-domestication spread of wheat. However, studies of the major vernalization gene *VRN1* are mostly limited to the analysis of di- and hexaploid wheat species. Questions concerning the origin and variability of the dominant *VRN1* alleles, determining the spring growth habit, remain unclear for tetraploid wheat species.
Methods and Algorithms: We used a combination of bioinformatical tools and molecular biology techniques to analyze diversity and evolution of *VRN1* genes from tetraploid wheats.
Results: We analyzed the growth habit of 230 accessions of tetraploid wheat species and the promoter and first intron regions of *VRN1* genes in spring accessions. The growth habit of most studied spring accessions was determined by known dominant alleles of *VRN1* genes. Two novel alleles were discovered and designated as *Vrn-A1dic* and *Vrn-B1dic*. *Vrn-A1dic* was widely distributed across the accessions of *T. dicoccoides* and had deletions 20 bp and 32 bp in length in the promoter region when compared to the recessive allele of *VRN-A1*. The dominant mode of inheritance of *Vrn-A1dic* allele was shown in the genetic experiments. *Vrn-B1dic* was identified in *T. dicoccoides* IG46225 and had 11 % sequence dissimilarity in comparison to the promoter of *vrn-B1*. The presence of *Vrn-A1dic* and *Vrn-B1dic* alleles is a predicted cause of the spring growth habit of studied accessions of T. dicoccoides. Three spring accessions *T. aethiopicum* K-19059, *T. turanicum* K-31693 and *T. turgidum* cv. Blancal possess recessive alleles of both *VRN-A1* and *VRN-B1* genes. Spring growth habit of these accessions seems to be determined by other vernalization genes.
Conclusion: Growth habit of 230 accessions of tetraploid wheat species was analyzed. Variability of *VRN1* genes was investigated for 59 spring accessions of tetraploid wheats and two novel dominant alleles *Vrn-A1dic* and *Vrn-B1dic* were detected.
Availability: The sequences of the promoter region and the first intron of *VRN-A1*, *VRN-B1* and *VRN-G1* genes were deposited in GenBank.

GENOME OF BLACK GARDEN ANT: DEFENSE AGAINST VIRUS INVASION?

E.A. Konorov, V.A. Scobeyeva, M.A. Nikitin, S.N. Lysenkov, S. Nuzhdin
Moscow State University, Moscow, Russia

Key words: *evolution, sequencing, ant genome*

More than 7 ant genomes are sequenced to date, and evolutionary history and genetic traits of this group are thoroughly described in many papers [1]. Most researchers focus on ant traits, associated with eusociality – complicated communication, extended longevity (mostly in queens) and cast differentiation. Search for genetic basis of eusociality determined the main stream of these studies – odorant and immune system and cast-specific developmental genes were examined. Nevertheless little attention was paid for genomic basis of ants environmental success. Ant successful adaptation to different environments can be understood only by detailed examination of each species biological peculiarities, especially population number and distribution.

We took Lasius niger for our analysis for its tremendous ability to stay against antropogenic pressure [2]. 50-80 workers were collected from 3 urban and 3 suburban populations, genomic DNA was isolated separately from each population and sequenced by Illumina. Reads from all populations were assembled with SPAdes 3.5 and annotated with Augustus, blast and blast2go. All bacterial and fungal sequences were removed from the assembly. Genes of CYP450, odorant receptors, neuropeptides and some other were manually annotated with reciprocal blastp. Reads from different populations were mapped on assembled genome with tophat and compared with popoolation2.

Genes of P450 CYP9 family were found amplified in *Lasius niger. O*ur previous data on protein modeling suggest that CYP9 proteins detoxify mycotoxins of Fusarium, fungus, highly contaminating our ant genome. Lasius niger has less genes of olfactory system than other sequenced ants, we suggest that there can be trade-off between olfactory and detoxification systems. Olfactory system reduction in Lasius niger could also facilitate polyfagy.

We found 4 times more integrated retroviral sequences in Lasius niger genome than in other ant's genomes. Comparison of different GO abundance between ant species revealed significant increase in GOs "nucleic acid binding" and "DNA repair" and also singificant decrease in GO "odorant binding". We can hypothesize that Lasius niger has high level of viral infection and amplifyed genes of DNA reparation system to reduce the damage from foreign DNA. Urban and suburban populations differ in integrated viral genes and some viral genes have signatures of selection.

1. Gadau J. et al. The genomic impact of 100 million years of social evolution in seven ant species // Trends in Genetics. – 2012. – T. 28. – №. 1. – C. 14-21.
2. Putiatina T. S. Effect of recreational pressure on ant communities of open biocenoses in Moscow // Moscow University Biological Sciences Bulletin. – 2011. – T. 66. – №. 1. – C. 42-45.

MOLECULAR EVOLUTION ANALYSIS OF GENES RELATED TO PLANT ROOT HAIR AND TRICHOME DEVELOPMENT

D.K. Konstantinov[1], A.V. Doroshkov[2]*
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: ad@bionet.nsc.ru*

Key words: *molecular evolution, epidermal cells, trichomes, root hairs*

Motivation and Aim: Specialized cell formation is an fruitful model system for analyzing the molecular mechanisms of plant cell differentiation, including cell fate choices, cell cycle control, and cell morphogenesis (Yang, C., & Ye, Z. 2013). In plants, epidermal cells are easily accessible and allow in vivo study. *Arabidopsis thaliana* trichomes were used as a model system for identification of the activator–inhibitor and the activator–depletion pattern formation models, studying the interplay between cell cycle and cell differentiation and numerous of genes involved in these processes were found. However, the evolution of specialized epidermal cell formation genetic network remains unclear. In this study, we analyze the phylogenetic relationships of genes associated with the formation of trichomes and root hairs from various species of flowering plants (monocotyledonous and dicotyledonous).
Methods and Algorithms: Extracting sets of homological sequences presets from databases was carried out using the reciprocal BLAST search. Multiple sequence alignment was conducted with MAFFT algorithm. The PhyML maximum likelihood algorithm was used to reconstruct the phylogeny and bootstrap resampling technique was used for testing the topology. Genetic networks containing target genes were reconstructed using Cytoscape and Pathway Studio software.
Results and conclusion: Our results argue that there is a large fraction of genes involved in the formation of trichomes and root hairs and do not reveal a direct correspondence between monocotyledonous and dicotyledonous plants. This fact is in the agreement with the latest data confirming that the closest homolog of *Arabidopsis thaliana* trichome and root hair-related gene in rice do not affect the rice pubescence (Zheng et.al., 2016). Also, this facts justify that part of the cellular morphogenesis mechanisms evolved independently in dicots and monocots. At the same time, we observe a good correspondence between studied genes of cell morphogenesis inside dicotyledonous as well as monocotyledonous clades. This allowed us to find orthologous genes in wheat genomic sequences and predict its chromosome localization to compare with known leaf hairiness QTLs.
Acknowledgements: This study was funded by Russian Science Foundation grant №14-14-00734.
References:
1. Yang, C., & Ye, Z. (2013). Trichomes as models for studying plant cell differentiation. Cellular and molecular life sciences, 70(11), 1937-1948.
2. Zheng, K., Tian, H., Hu, Q., Guo, H., Yang, L., Cai, L. & Wang, S. (2016). Ectopic expression of R3 MYB transcription factor gene OsTCL1 in Arabidopsis, but not rice, affects trichome and root hair formation. Scientific reports, 6, 19254.

TWO CONGENIC STRAINS PROVE EFFECTS ON CATARACT AND RETINOPATHY BUT NOT ON BRAIN NEURODEGENERATION IN SENESCENCE-ACCELERATED OXYS RATS

E.E. Korbolina*, A.O. Vitovtov, N.G. Kolosova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: lungry@bionet.nsc.ru*

Key words: *genetic architecture of complex trait, OXYS rats, quantitative trait locus, congenic strain, Alzheimer's disease, age-related macular degeneration*

Motivation and Aim. There has been a considerable interest in discovering the genetic architecture of complex traits, particularly those genetic components which are involved in age-related neurodegenerative disorders. To predict disease risk and to understand the underlying genetic basis in human, it is essential to study animal models. We assumed that mutations affecting the development of early cataract in senescent-accelerated OXYS rats also have effects on early manifestation of other age-related disorders. With this strategy, we identified two quantitative trait loci (QTLs) associated with early cataract, AMD-like retinopathy, and certain behavioral signs of brain neurodegeneration on rat chromosome 1 (RNO1). These loci were mapped as partially being within the introgressed RNO1 segments of WAG/OXYS-1.1 and WAG/OXYS-1.2 congenic strains, spanning totally 81.1 Mbp from 8900000 to 105000000 bp and from 178 000 000 to 275 000 000 bp, respectively, with defined breaks. The retina transcriptome profiles of congenic and OXYS rats were analyzed at 20 days by means of RNA-seq, since congenic rats carrying OXYS alleles displayed retinopathy. The present study is aimed to dissect the susceptibility genes and builds upon our prior analyses in several ways, including further characterization of congenic rat's phenotype.

Methods and Algorithms. Standart techniques were used to analyze rat's behavior (open field, plus maze) and learning capacity (8-arm radial maze); Beta direct ophthalmoscope – for ophtalmoscopic examinations; MRI tomography with MSME technique - for evaluating of brain morphology; Statistica 6.0 and free R software – for data analyses and visualization; RGD and WebGestalt online tools - for annotation.

*Results. The n*eurodegenerative brain alterations typical for both OXYS rats and patients diagnosed with Alzheimer's disease were not observed in congenic rats. So, we argue, that the accelerated-senescence phenotype in OXYS rats can be genetically dissected. The progression of retinopathy with age differ in WAG/OXYS-1.1, WAG/OXYS-1.2 and OXYS rats both clinically and histologically. Based on the results we suggested a certain similarity in molecular-genetic mechanisms underlying retinopathy development and progression in WAG/OXYS-1.1 and WAG/OXYS-1.2 rats. A comprehensive analysis of candidate genes within previously defined QTLs and congenic segments was carried out. Moreover, the set of genes differentially expressed in retina of OXYS and WAG/OXYS-1.2 rats at 20 days was found significantly enriched with targets for Jun and Stat3 transcriptional factors, widely involved in retinopathy progression. Thereby, we revealed some synergy in causing early cataract and retinopathy in OXYS rats, yet the relationship between the variance in a trait and allele frequency should be investigated further. *This study was supported by the Russian Foundation for Basic Research (Grant # 15-04-06066 A).*

ASSOCIATION OF MATRIX METALLOPROTEINASES GENE POLYMORPHISM WITH THE RISK OF DEVELOPING EXTRA-ARTICULAR SYMPTOMS OF RHEUMATOID ARTHRITIS

M.A. Korolev*, Y.B. Ubshaeva, E.A. Letyagina, A.V. Shevchenko, V.F. Prokof'yev, V.I. Konenkov

Scientific Institute of clinical and experimental lymphology SB RAS, Novosibirsk, Russia

** Corresponding author: kormax@bk.ru*

Key words: *rheumatoid nodules, MMP2-1306, MMP3-1171, MMP9-1562, SNP*

Motivation and Aim: Rheumatoid nodule is considered specific extra-articular symptom of rheumatoid arthritis (RA) and one of the diagnostic criteria associated with severe disease. Matrix metalloproteinases (MMPs) are seen as major proteolysis enzymes involved in the implementation of destructive processes in RA and in a number of data correlate with disease activity. MMPs can play a key role in the development of extra-articular symptoms of RA. The aim of this study was to investigate MMP gene polymorphism in Caucasoid RA patients with and without rheumatoid nodules. The control group included healthy women.

Methods and Algorithms: 162 women with RA satisfied (ACR, 1987) criteria were included in this study, rheumatoid nodules were detected in 38 of them (23,5%). We have studied single nucleotide polymorphisms of the promoter region of genes MMP2-1306 (rs243865), MMP3-1171 (rs3025058), MMP9-1562 (rs3918242). Genotyping was performed by method of restriction fragment length polymorphism. In the statistical analysis of the results of the studies we used indicators such as the frequency of genes, genotypes and their combinations, specificity, odds ratio (OR) with the calculation of 95% confidence interval (OR 95% CI). Calculation of the OR was performed by the method of Woolf-Haldane. Significant difference in the distribution of the studied traits in alternative groups was determined by criterion $\chi 2$ with Yates correction.

Results: Differences in the frequencies of MMP genes between groups of patients with and without rheumatoid nodule were not detected, may be due to the small group of patients with nodules. However, the relatively healthy women RA patients had a reduction in the frequency of the homozygous genotype MMP36A6A (OR =0,29, P 0,0062 in group with the presence of rheumatoid nodules and OR =0,52, P 0,0295 in group without nodules). Also in both groups of relatively healthy patients we observed increased incidence of MMP2-1306 TT homozygous gene variant (OR =2,94 P 0,0498 in group with rheumatoid nodules and OR =3,08 P 0,0030 in group without nodules).

Conclusions: Polymorphism at promoter region of MMP genes may determine the development of the disease and the presence of extra-articular symptoms of RA.

GENERALISING BETTER: APPLYING DEEP-LEARNING TO INTEGRATE DELETERIOUSNESS PREDICTION SCORES FOR WHOLE-EXOME SNV STUDIES

I.O. Korvigo[1], A.A. Afanasyev[1]*
[1] *Moscow Institute of Physics and Technology, Moscow, Russia*
* *Corresponding author: andrei.afanasiev@gmail.com*

Key words: *deep learning, SNV, in silico prediction, ANN, MLP, ReLU, sDA*

Motivation and Aim: Over the past fifteen years researchers have developed a plethora of individual deleteriousness scoring systems, such as PolyPhen and SIFT. Lately, the focus has been shifting from creating novel standalone tools towards combining available scoring systems into ensembles or meta-scores. The shortage of reference data, combined with the bias towards intensively studied model systems, calls for improved generalisation ability.
*Methods and Algorithms:*We applied several deep-learning strategies and compared the performance with all ACMG-recommended scoring systems as well as two novel scores, employing deep-learning. We dumped the entire Uniprot and ClinVar variation databases and trained an out-of-the bag classifier to filter all trustworthy records. We handpicked three hundred random records and manually classified them into trustworthy and dubious based on the number of referenced publications, publication key-words and record descriptions. After removing duplicates and SNPs, found in the testing datasets, we had 90k SNPs left for training purposes. Although artificial neural networks, such as the Multilayer Perceptron (MLP), are not a novelty per se, it was not until recently that advances in their design, training algorithms and hardware accelerated computations, made it possible to effectively train deep nets, pushing them to the forefront in many fields demanding precise classification of noisy and highly polymorphic input data. Here we applied three designs: the standard MLP with sigmoid activations, the MLP with a rectifier activation function (ReLU) and stochastic dropout training procedure, and the MLP pretrained as stacked Denoising Autoencoders (sDA). Since each network has several parameters to configure, such as the hidden-part shape, the learning rate, the momentum applied and the dropout rate, we used a genetic algorithm approach to find the best set of parameters for each design.
Results: Our ReLU network showed the highest performance of all. The best model has surpassed MetaLR on both the VariBench dataset (Average accuracy 0.9 vs 0.86) and the Nature dataset (Average accuracy 0.91 vs 0.84) by statistically significant margins. Surprisingly the performance skyrocketed to 0.97 average accuracy when all missing records were removed from the second testing dataset and dramatically decreased on the Varibench set, hence the network got overfitted to missing data
Conclusion: The stacked Autoecoder-based network, though not as good as the dropout MLP, has been equally stable at predicting both datasets given full and missing records, hence it really does generalise better.

KU ANTIGEN DISPLAYS THE AP LYASE ACTIVITY ON A CERTAIN TYPE OF DUPLEX DNA

A.A. Kosova*, S.N. Khodyreva, O.I. Lavrik
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
** Corresponding author: kosova.anastasiya@gmail.com*

Key words: *Ku antigen, DNA repair, AP sites, AP lyase, Schiff base*

Motivation and Aim: In the search for proteins reactive to apurinic/apyrimidinic (AP) sites, it has been earlier found that proteins of human cell extracts formed the Schiff-base-dependent covalent adduct with an apparent molecular mass of 100 kDa with a partial DNA duplex containing an AP site and 5′- and 3′-protruding e
nds (DDE-AP DNA). The adduct of such electrophoretic mobility was characteristic of only DDE-AP DNA [1].
Methods: The Schiff-base-dependent cross-linking of proteins with AP DNA (borohydride trapping) in combination with gel electrophoresis and MALDI-TOF MS was used to identify the protein. Chromatography was used to enrich the cell extract in the target protein and to purify the protein.
Results: The protein forming the unusual 100-kDa adduct with DDE-AP DNA was identified as the Ku80 subunit of Ku antigen by peptide mass mapping based on MALDI-TOF MS data [2]. Then the interaction of Ku with DDE-AP DNA was studied in details. Purified Ku (the Ku80 subunit) was shown to form the 100-kDa adduct highly specific for AP DNA with a certain length of protruding ends, base opposite the AP site and AP site location. Ku is capable of AP site cleavage in DDE-AP DNA unlike in analogous AP DNA with blunt ends. Ku cleaves AP sites via β-elimination and prefers apurinic sites over apyrimidinic ones. The AP site in DDE-DNA can be repaired via the successive action of Ku (cleavage of the AP site), tyrosyl-DNA phosphodiesterase 1 (removal of the 3′-deoxyribose residue), polynucleotide kinase 3′-phosphatase (removal of the 3′-phosphate), DNA polymerase β (incorporation of dNMP), and DNA ligase (sealing the nick) [3].
Conclusion: This scenario may be considered as an APE1-independent backup pathway of the AP site repair. It can be realized under the pharmacological inhibition of APE1 activity used to potentiate the activity of chemotherapeutic alkylating agents.
Acknowledgements: The work was supported by the Russian Scientific Foundation project no. 14-14-00501 and the program of the Russian Academy of Sciences "Molecular and Cellular Biology".
References:
1. E.S. Ilina, O.I. Lavrik, S.N. Khodyreva. (2008) Ku antigen interacts with abasic sites, *Biochim. Biophys. Acta,* 1784: 1777–1785.
2. A.A. Kosova, S.N. Khodyreva, O.I. Lavrik (2014) Ku80 interaction with apurinic/apyrimidinic sites depends on the structure of DNA ends, Biopolym. Cell, 30: 42–46.
3. A.A. Kosova, S.N. Khodyreva, O.I. Lavrik (2016) Ku antigen displays the AP lyase activity on a certain type of duplex DNA, *Biochim. Biophys. Acta,* doi: 10.1016/j.bbapap.2016.04.011

INVOLVEMENT OF VARIOUS CELL DEATH MODALITIES IN CYTOTOXIC ACTIVITY OF LACTAPTIN ANALOG

O.A. Koval[1, 2]*, G.V. Kochneva[1, 3], A.V. Tkachenko[1], O.S. Troitskaya[1, 2], G.F. Sivolobova[1, 3], E.V. Kuligina[1], A.Y. Yunusova[1], V.A. Richter[1]

[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *State Research Center of Virology and Biotechnology "Vector", Koltsovo, Russia*

* *Corresponding author: o.koval@niboch.nsc.ru*

Key words: *lactaptin, apoptosis, autophagy, immunogenic cell death, chemoresistant tumor, oncolytic viruses*

Motivation and Aim: Lactaptin has previously been shown to act as a tumor suppressor in mouse hepatoma as well as MDA-MB-231 human adenocarcinoma cells grafted into SCID mice [1, 2]. Although lactaptin has been demonstrated to induce cancer cell death, the mode of cell death remains increasingly confusing, and the data as a whole suggest apoptosis is not the only mechanism that is responsible for lactaptin-induced tumor regression. The objective of this study was to investigate the involvement of molecular markers of autophagy and immunogenic cell death in cytotoxic activity of lactaptin analogs. *Methods and Algorithms:* Flow cytometry, Western blot analysis, RT PCR analysis and ELISA were performed to analyze the markers of lactaptin-induced cell death *in vitro*. The inhibitors and inducers of autophagy were used for improve cytotoxic activity of lactaptin. Transition electron microscopy was used for the imaging of the autophagosome formation. A double recombinant vaccinia virus (VV-GMCSF-Lact) that expresses exogenous proteins: the antitumor protein lactaptin and human granulocyte-macrophage colony-stimulating factor (GM-CSF) was also generated to enhance antitumor activity of vaccinia virus. *Results:* We observed that combination treatment of the cells with lactaptin and chloroquine (CQ) increased the death of cancer cells. Combination therapy with lactaptin and autophagy inhibitor CQ significantly suppressed the growth of sensitive and chemoresistant tumors. The major DAMPs of immunogenic cell death were estimated after the treatment of cancer cells with lactaptin analog. Cell surface expression of calreticulin, the decreasing of the cellular HMGB1 and increasing of extracellular ATP were detected after the treatment with lactaptin analog. We also investigated the biological effect of recombinant vaccinia virus VV-GMCSF-Lact on the growth properties and apoptosis of human cancer cells and its antitumor activity against various tumors. *Conclusion:* These results demonstrate the involvement of molecular markers of autophagy and immunogenic cell death in cytotoxic activity of lactaptin analogs. Taken together, our findings confirm that lactaptin analogs are perspective molecules for a further design of anticancer drugs.

Acknowledgements This work was partly supported by the RFBR grant # 16-04-01232.

References:

1. O.A. Koval, A.V. Tkachenko et al. (2014) A.Uthor, O.N.E.More. (2003) Lactaptin induces p53-independent cell death associated with features of apoptosis and autophagy and delays growth of breast cancer cells in mouse xenografts. *PLoS One*. **9**:e93921.
2. O.A. Koval, A.S. Fomin, et al. (2012) A novel pro-apoptotic effector lactaptin inhibits tumor growth in mice models. *Biochimie*. **94**:2467-2474.

INCONGRUENT NUCLEAR AND MITOCHONDRIAL GENETIC STRUCTURE IN BAIKALIAN AMPHIPODS GMELINOIDES FASCIATUS

M.V. Kovalenkova*, Zh.V. Petunina, D.Yu. Sherbakov
Limnological Institute SB RAS, Irkutsk, Russia
** Corresponding author: kovalenkovam@mail.ru*

Key words: *Gmelinoides fasciatus, mito-nuclear discordance, neutrality*

Motivation and Aim: Gmelinoides fasciatus is rather ancient baikalian amphipod species which easily invades new habitats currently. It's population structure in Baikal can provide new information about the details of microevolutionary processes in this species. We used fast evolving single intron of a nuclear gene coding for the ATPase beta-subunit along with the usual bar-code mitochondrial COI gene segment in order to elucidate population structure of the baikalian *G. fasciatus*.
Methods and Algorithms: Bayesian clustering analysis as implemented in GENELAND 3.2.4 was used to investigate spatial distribution of genetic clusters. To test population subdivision for isolation by distance we performed the Mantel Test. Pairwise Fst values and test their statistical significance were estimated using the ARLEQUIN v. 3.11. We tested both datasets for neutrality using the DNAsp v. 5. A median-joining network analysis was performed using NETWORK v. 4.5.1.6.
Results: The mtDNA bayesian analysis indicates that there are four distinct *G. fasciatus* populations in Baikal, while the nDNA data suggest only three populations. Moreover there is only one population boundary coincided as inferred from the two data sets. All populations have high significant pairwise Fst values and separate clades on the networks. We highlight two main incongruent geographical clusters: North Basin seems to be homogenous in case of the mitochondrial marker, but splits into East and West parts if intron allele frequencies are considered. In contrast to the introns sequences based subdivision, mtDNA sequences split the South Basin group into two dramatically distinct communities previously designated as Southeastern and Southwestern populations [1]. We haven't found significantly strong correlation between genetic and geographic distances with both markers. Neutrality tests for the nuclear data provide considerable evidence of the population expansion in Southern group, whereas other neutrality tests indicate that purifying selection have shaped the mitochondrial polymorphism in the Southwestern group.
Conclusion: Described discordances are due to recent introgression observed in nDNA and selective pressures affecting mtDNA. Applying of mt and nuclear molecular markers let uncover different evolutionary processes acting on the genomes. The effect may be further enhanced by the distinctions in effective population numbers as estimated from mitochondrial and nuclear data.
Acknowledgements: The study was partly supported by the RFBR grants Nos. 15-04-03848 and 14-44-04138.
References:
1. G.V. Gomanenko et al. (2005) Population structure of baikalian amphipoda *Gmelinoides fasciatus* (Stebbing), *Genetika* (In rus), 41: 1108-1114.

MATHEMATICAL MODELING A RECIPROCAL INTERACTIONS BETWEEN AUXIN AND ITS PIN TRANSPORTERS IN THE ROOT TIP OF A. THALINA L

V.V. Kovriznykh[1, 2]*, F.V. Kazantsev[1], N.A. Omelyanchuk[1, 2], V.V. Mironova[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *Corresponding author: vasilinakovr@gmail.com*

Key words: *auxin, Arabidopsis thaliana, PIN-FORMED (PIN), mathematical modeling*

Motivation and Aim: Plant hormone auxin is the main regulator of plant growth and development. In the root, auxin is distributed nonuniformly, with concentration gradients and a maximum in the quiescent center (QC), which are necessary for the meristem maintaining. The PIN-FORMED family proteins (PIN1-PIN4, PIN7) asymmetrically locate on the cell membrane and by this form directed auxin flows in the root tissue. Auxin regulates the speed of its efflux from the cell by mediating PINs expression. Comparing the experimental data on auxin response distribution and PIN proteins localizations, we hypothesized that different auxin concentrations might be optimal for expression of various PINs. This assumption was verified by mathematical modeling.
Methods and Algorithms: 2D cellular automaton was developed on the basis of the model [1] to study auxin distribution along longitudinal direction of the *A. thaliana* root tip. The cell layout consists of inner and outer layers, which differ by the set of expressed PINs. Auxin flows into the upper cell raw of the inner layers, then it is distributed across the cell layout by diffusion and PIN-mediated polar transport. In accordance with the assumption, PINs expression is established by auxin concentration in a cell in a dose-response manner.
Results: We started the model calculation from zero initial condition, which at the beginning of calculation led to generation of PIN2 expression in the outer layers with shootward polarization. As auxin entered to the cell layout, a PIN3, PIN4, PIN7 and then PIN1 began to be expressed with rootward polarization. All these events resulted in formation of auxin flow towards the root tip, which reached the root end and accumulated there. Auxin accumulation in columella cells mediated the changes in PINs expression there from rootward to non-polar. Gradually the auxin concentration maximum shifted to fifth cells from the root end. Finally, auxin concentration gradients were formed at both sides of the concentration maximum.
Conclusion: As a result of the numerical simulation, from the uniform auxin distribution in a tissue we got a steady-state solution, in which the auxin distribution organized itself simultaneously with the pattern of PIN proteins. The results of *in silico* experiments fitted well the *in vivo* data on auxin and PIN proteins pattern in wild-type. Consequently, our assumptions confirmed that auxin provides positional information to root tip cells.
Acknowledgements: The work was supported by RSF grant 14-14-00734.
1. Mironova V.V. et.al. (2012). Annals of Botany, 110(2), 349-360.

TWO MODELS OF THE DROSOPHILA GAP GENE NETWORK WITH VARIATION OF MATHERNAL INPUT

K.N. Kozlov[1]*, A.V. Svichkarev[1], V.V. Gursky[1, 2], I.V. Kulakovskiy[3], S.Y. Surkova[1], M.G. Samsonova[1]

[1] *Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

[2] *Ioffe Institute, St. Petersburg, Russia*

[3] *Engelhardt Institute of Molecular Biology RAS, Moscow, Russia*

* *Corresponding author: kozlov_kn@spbstu.ru*

Key words: *morphogen, transcription, thermodynamics, reaction-diffusion, drosophila*

Motivation and Aim: Bicoid (Bcd) as a classic morphogen plays a key role in the expression of segmentation genes in Drosophila embryo. The segmentation gene network controls the determination of the fruit fly segments during the blastoderm stage, just before the onset of gastrulation. We applied the systems-level approach to understand the spatial dynamics of gap gene expression domains under different Bicoid dosages.

Methods and Algorithms: We considered two modeling approaches. The first phenomenological model [1] uses the reaction-diffusion differential equations and the matrix of regulatory coefficients characterizing the action of regulators on their targets. The second sequence-based model [2, 3] uses thermodynamic approach to model target gene expression at the RNA level and two sets of reaction-diffusion differential equations for mRNA and protein concentrations to describe the dynamics of the system. We model the expression of 4 gap genes – *hb*, *Kr*, *gt*, and *kni* – under control of 11 transcription factors (TF) – the products of *hb*, *Kr*, *gt*, *kni*, *bcd*, *tll*, *cad*, *hkb*, *cic*, *slp,* and *run* genes. We predicted TF binding sites in the potential regulatory region from 12Kbp upstream to 6Kbp downstream of transcription start site for each gene using enhanced dinucleotide positional weight matrices. The unknown model parameters were obtained with the DEEP method by fitting the model solutions to both expression patterns of gap genes and data on the *hb* anterior domain shifts in embryos with varying Bcd concentration.

Results: Both models successfully reproduce the characteristic features of experimental data. The sequence-based model reproduces the spatial dynamics of the *hb* anterior expression domain more precisely.

Availability: The software is available from authors upon request.

Acknowledgments: The study was supported by the RFBR grant 14-04-01522 and the "5-100-2020" Program of the Ministry of Education and Science of the Russian Federation.

References:

1. Andreev, S. A., Samsonova, M. G., and Gursky, V. V. (2015). Modeling of the Drosophila gap-gene network with the variation of the Bcd morphogen. Biophysics, 60(2), 173-180.
2. K. Kozlov, V. Gursky, I. Kulakovskiy, A. Dymova and M. Samsonova, Analysis of functional importance of binding sites in the Drosophila gap gene network model. BMC Genomics 2015, 16(Suppl 13):S7.
3. K. Kozlov, V. Gursky, I. Kulakovskiy, M. Samsonova (2014). Sequence-Based Model of Gap Gene Regulatory Network. BMC Genomics, 15(Suppl 12):S6.

COMPARATIVE GENOMICS AND TRANSCRIPTOMICS OF *CHIRONOMIDAE* MIDGES UNDER EXTREME CONDITIONS

O.S. Kozlova[1]*, A.V. Cherkasov[1], R.M. Devyatiarov[1], M.D. Logacheva[2, 1], R. Cornette[3], T. Kikawada[3], A.A. Przhiboro[4], O.A. Gusev[5, 1]

[1] *Extreme Biology Laboratory, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[2] *Laboratory of evolutionary genomics, Faculty of bioengineering and bioinformatics, Moscow State University, Moscow, Russia*

[3] *National Institute of Agrobiological Sciences, Tsukuba, Japan*

[4] *Zoological Institute, Russian Academy of Sciences, Moscow, Russia*

[5] *Preventive Medicine & Diagnosis Innovation Program (PMI), Division of Genomic Technologies, RIKEN, Yokohama, Japan*

** Corresponding author: olga-sphinx@yandex.ru*

Key words: *Chironomidae, stress exposure, heat shock proteins*

Motivation and Aim: Chironomidae are considered to be among the most accommodated to abiotic stresses species of insects. Detailed understanding of underlying mechanisms of their tolerance or resistance to extreme conditions remains the hotspot of evolutionary molecular study of arthropods. In this research we compared genomics and transcriptomics properties of six chironomids midges from different extreme ecosystems to reveal a background of their extraordinary resistance to abiotic stresses.

Methods and Algorithms: We have used previously published genomes of three chironomids species as well as three newly sequenced and annotated genomes and transcriptomes. Since *de-novo* genome assembly of eukaryotic species presents significant challenges (even despite relatively small genome size of chironomids), in many cases it was easier to perform general transcriptome analysis according to Trinity-associated protocol with Transdecoder, Trinotate and bowtie-RSEM-edgeR for differential expression.

Results and Conclusion: Detailed analysis of expansion/contraction of functional groups of stress-responsible genes showed that high ability to withstand stress is not always necessary linked to changes in number of such genes. Different behavior of orthologous genes or transcripts upon stress exposure makes it reasonable to conjecture that often accommodation to abiotic stress goes on the way of regulation, as it was shown for such conservative group of stress-inducible agents as heat shock proteins.

Acknowledgements: The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

COMPARATIVE TRANSCRIPTOMICS PROVIDES NEW INSIGHTS INTO ORIGIN OF EXTRAORDINARY RESISTANCE TO DESICCATION IN AUSTRALIAN MIDGE *PARABORNIELLA TONNOIRI (CHIRONOMIDAE)*

O.S. Kozlova[1]*, E.I. Shagimardanova[1], L.Kh. Shigapova[1], R.M. Devyatiarov[1], M.D. Logacheva[2, 1], R. Cornette[3], T. Kikawada[3], O.A. Gusev[4, 1]

[1] *Extreme Biology Laboratory, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[2] *Laboratory of evolutionary genomics, Faculty of bioengineering and bioinformatics, Moscow State University, Moscow, Russia*

[3] *National Institute of Agrobiological Sciences, Tsukuba, Japan*

[4] *Preventive Medicine & Diagnosis Innovation Program (PMI), Division of Genomic Technologies, RIKEN, Yokohama, Japan*

** Corresponding author: olga-sphinx@yandex.ru*

Key words: *desiccation, transcriptome analysis, Chironomidae, Paraborniella tonnoiri Motivation and Aim*

Many *Chironomidae* species are known for their ability to successfully combat abiotic stresses using wide range of behavioral, morphological and biochemical features. Larvae of Australian midge *Paraborniella tonnoiri* presents example of evolving mechanism of extraordinary resistance to water loss. Remarkably, morphological, physiological and behavior patterns here are different from those used by truly anhydrobiotic African chironomid *Polypedilum vandrerplanki*. The aim of our study was to reveal new molecular mechanisms, which help the given species to retain necessary amount of water inside its body and thus resist partial desiccation in laboratory conditions, as well as in nature.

Methods and Algorithms: Since *P. tonnoiri* remains previously unsequenced and does not have its genome assembled and annotated, we performed high throughput mRNA sequencing using control, desiccated and heat shocked (to estimate general response to heat) groups of larvae to get the pool of 13 libraries with paired-end Illumina reads. These libraries were used for *de-novo* draft transcriptome assembly using Trinity, after that ORF prediction and functional annotation were carried out according to Trinity-associated approach with Transdecoder and Trinotate.

Results and Conclusion: Among general pool of transcripts, more than 2 000 showed themselves as differentially expressed in response to desiccation and/or heat shock with at least 2-fold change in expression (FDR$\leq$0.01). The most interesting cluster of transcripts stands for specific response to desiccation. Comparing the data with those of truly anhydrobiotic chironomids suggest that genetic mechanism of ability to resist severe desiccation in *P. tonnoiri* evolved in completely independent from truly anhydrobiosis evolutionary way.

Acknowledgements: The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

TRAJECTORIES OF THE DNA KINKS IN THE SEQUENCES CONTAINING CDS REGIONS

L.A. Krasnobaeva[1, 2]*, L.V. Yakushevich[3]
[1] *Tomsk State University, Tomsk, Russia*
[2] *Siberian State Medical University, Tomsk, Russia*
[3] *Institute of Cell Biophysics RAS, Pushchino, Moscow region, Russia*
** Corresponding author: kla1983@mail.ru*

Key words: *DNA dynamics, kink trajectories, gene energy profile, CDS regions, interferon alpha 17*

Motivation and Aim: In our previous works [1-2] to model the DNA kinks, we used a simple model based on the sine-Gordon equation with parameters that were averaged over all length of the gene sequence. However, this approach does not allow us to take into account the effect of the internal structure of the DNA sequence, and, in particular, the presence of the CDS regions, on the DNA kinks dynamics. In this paper, we just consider this problem and solve it for the gene encoding interferon alpha 17, the sequence of which consists of three regions: the coding (CDS) region (50..619) and the two regions (1..49 and 620..980) with unknown functional significance.
Methods and Algorithms: To solve the problem, we use several methods: the method of McLaughlin and Scott, the average field approximation and the block method where the parameters of the model equation are averaged separately for each of the three regions. To analyze the DNA kinks dynamics, we use the physical approach which includes the calculation of the energy profile of the sequence and the construction of the trajectory of the movement of the DNA kink in the potential with this profile.
Results: We have obtained the energy profile of the sequence of the gene coding interferon-alpha 17. It was shown that the CDS region corresponds to the region of the energy barrier. The minimum value of the kink initial velocity required to overcome the barrier was estimated. The trajectories with different initial kink velocities were constructed. The trajectories were calculated both with and without dissipative effects. It was shown that with the increasing of the initial kink velocity the trajectories became more independent on the inhomogeneity of the sequence. We suggest that the proposed approach can be applied to analyze the movement of transcription bubbles through the CDS regions of the DNA sequences.

References:

1. L.A. Krasnobaeva, L.V. Yakushevich (2015) Rotational dynamics of bases in the gene coding interferon alpha 17 (IFNA17), *Journal of Bioinformatics and Computational Biology*, **13:** №1, 1540002 (13 pages).
2. L.V. Yakushevich, L.A. Krasnobaeva (2016) Forced oscillations of DNA bases, *Biophysics,* **61:** № 2, 286-296.

INMETHYL: A TOOL FOR DESIGN OF SPECIFIC PRIMERS FOR METHYLATION PROFILING OF COMPLETE CPG ISLANDS

G.S. Krasnov*, A.V. Kudryavtseva, N.V. Melnikova, A.A. Dmitriev
Engelhardt Institute of Molecular Biology RAS, Moscow, Russia
** Corresponding author: gskrasnov@mail.ru*

Key words: *DNA methylation, CpG island, primer design, bisulfite conversion*

Motivation and Aim: The non-specificity of PCR amplification after bisulfite conversion is one of the most common issues in gene methylation studies. In fact, bisulfite treatment leads to the reduction of 4-letter alphabet (ATGC) to 3-letter (ATG, except methylated cytosines) that dramatically increases a possibility of mispriming. The second issue of the promoter region studies is coming from the features of CpG islands sequence: low-complexity, polyN-rich, and CG-rich.
Methods and Algorithms: InMethyl is a Python-based application. It uses bowtie high-throughput aligner to identify potential mispriming sites and undesirable PCR products in the bisulfite treated or intact genome. Primer selection is based on calculating scoring factor that takes into account primer pair specificity, nucleotide composition (sequence complexity), thermodynamic features (melting temperature, dimers dG, *etc.*), presence of CpG sites and other parameters. Users are intended to customize desired or limit ranges of these values as well as penalties for out-of-bounds values.
Results: We developed the InMethyl software, a novel application enabling the design of target-specific primers for amplification of CpG islands and other hard-to-study genomic regions. A key feature of the tool is the balance between various characteristics that allows to pick up primers in the arduous genomic regions. Moreover, InMethyl software allows users to optimize combination of PCR primer pairs to perform the amplification of large genomic regions, e.g. CpG islands.
Conclusion: InMethyl is a novel powerful tool for the design of specific primers for DNA methylation profiling even of complete CpG islands.
Availability: InMethyl software is freely available at https://sourceforge.net/projects/inmethyl/.
Acknowledgements: This work was financially supported by the Russian Foundation for Basic Research (grants 15-04-08731 and 15-34-70055 mol_a_mos) and RAS Presidium Program “Molecular and Cellular Biology” (grant for AAD).

RTRANS: ANALYSIS OF RNA-SEQ DIFFERENTIAL EXPRESSION USING GLM APPROACH AND UNCOVERING ITS BIOLOGICAL BACKGROUND

G.S. Krasnov*, A.V. Snezhkina, N.V. Melnikova, A.A. Dmitriev, A.V. Kudryavtseva
Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia
** Corresponding author: gskrasnov@mail.ru*

Key words: *transcriptomic data, RNA-Seq analysis, RTrans, Bioconductor*

Motivation and Aim: RNA-Seq analysis is one of the most informative approaches for the discovering alterations in cell mechanisms standing behind adaptation, differentiation, development, response to stress and other essential biological processes. The analysis of transcriptomic data quite often is a problem since it is labor intensive and needs high qualification of a specialist.
Methods, algorithms, results: We present RTrans, an R script aimed at comprehensive analysis of RNA-Seq data, from identification of differentially expressed genes (including GLM multivariate testing), creating PCA plots and heatmaps to the gene set enrichment analysis (GSEA) based on Gene Ontology and KEGG databases as well as visualization of alterations in KEGG pathways (e.g. MAPK, PI3K/mTOR, p53, etc.). RTrans uses read-counts-per-gene files, which are generated with HT-Seq, featureCount tools or PPLine, our previously developed automated pipeline for the analysis of transcriptome or exome sequencing data. RTrans is based on several Bioconductor packages: edgeR, topGO, clusterProfiler and pathview.
Conclusion: Thus, RTrans is flexible R tool that provide a way to analyze RNA-Seq data and understand its biological background by cost of spending only several minutes to create sample description sheet and select conditions or GLM models to test.
Availability: RTrans is freely available at https://sourceforge.net/projects/rtrans/
Acknowledgements: This work was financially supported by the Russian Foundation for Basic Research (grants 15-04-08731, 16-16-00114 and 15-34-70055) and RAS Presidium Program “Molecular and Cellular Biology”.

DIFFERENTIAL EXPRESSION OF ALTERNATIVELY SPLICED TRANSCRIPTS RELATED TO ENERGY METABOLISM IN COLORECTAL CANCER

G.S. Krasnov[1], A.V. Snezhkina[1], I.Y. Karpova[1], O.L. Kardymon[1]*, M.S. Fedorova[1], A.A. Moskalev[1], A.F. Sadritdinova[1, 2], K.M. Nyushko[2], N.V. Melnikova[1], D.V. Kalinin[3], A.A. Belova[1, 2], M.A. Chernichenko[2], K.M. Klimina[4], D.V. Sidorov[2], A.Y. Popov[2], A.A. Dmitriev[1], A.V. Kudryavtseva[1, 2]

[1] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*

[2] *Herzen Moscow Cancer Research Institute, Ministry of Health of the Russian Federation, Moscow, Russia*

[3] *A.V. Vishnevsky Institute of Surgery, Moscow, Russia*

[4] *Vavilov Institute of General Genetics, Russian Academy of Sciences Moscow, Russia*

** Corresponding author: rhizamoeba@mail.ru*

Key words: *alternative splicing, energy metabolism, tumor-specific mRNA isoforms, colorectal cancer, adenocarcinoma*

Colorectal cancer (CRC) is one of the most common malignant tumors worldwide. CRC has several susceptibility factors such as hereditary, smoking, diet, microbiota, and chronic inflammation. CRC molecular pathogenesis is heterogeneous and may be followed by mutations in oncogenes and tumor suppressor genes, chromosomal and microsatellite instability, hypermethylation of CpG islands, oxidative stress, impairment of different signaling pathways, and energy metabolism alterations. In the present study, we used CrossHub software to evaluate the expression of alternatively spliced transcripts related to energy metabolism in CRC. Using the analysis of The Cancer Genome Atlas (TCGA) RNA-Seq datasets derived from colorectal cancer and adjacent normal tissue we examined the expression of 1014 alternative mRNA isoforms involved in cell energy metabolism. We revealed 11 genes with differentially expressed alternative transcripts whereas overall gene expression was not significantly altered in CRC. A set of 15 differentially expressed mRNA isoforms of interest has been validated by qPCR. Thirteen mRNA isoforms were overexpressed in colorectal tumors (OGDH: uc011kby.1 and uc011kbz.1; COL6A3: uc002vwo.2; ICAM1: uc010xle.1; PHPT1: uc004cjq.3; PPP2R5D: uc010jyd.2; HKDC1: uc001jpf.3 and uc009xqb.2; SLC29A1: uc003owz.1; MYBBP1A: uc002fxz.3; SRI: uc003ujq.1; TRIB3: uc002wdm.2 and uc002wdn.2) which is in concordance with the bioinformatics data. Three genes revealed negative correlation between the expression level of their alternative mRNA isoforms (MYBBP1A: uc002fxz.3 and uc002fyb.3; SRI: uc003ujq.1 and uc003ujr.1; TRIB3: uc002wdm.2 and uc002wdn.2). Six of thirteen isoforms were also strongly overexpressed in breast, lung, prostate and kidney tumors. Thus, these alternative mRNA isoforms could be involved in the development of cancers through altered energy metabolism. Our results suggest a set of six tumor-specific mRNA isoforms that may be used for cancer diagnosis methods development.

This work was financially supported by grant 14-15-01083 from the Russian Science Foundation. Part of this work was performed using the equipment of EIMB RAS "Genome" center (http://www.eimb.ru/rus/ckp/ccu_genome_c.php).

FAIRDOM: DATA AND MODEL MANAGEMENT FOR SYSTEMS BIOLOGY PROJECTS

O. Krebs[1]*, R. Kuzyakiv[5], M. Golebiewski[1], S. Owen[2], Q. Nguyen[1], N. Stanford[2], K. Wolstencroft[4], J.L. Snoep[2, 3], B. Rinn[5], W. Mueller[1], C. Goble[2]
[1] Heidelberg Institute for Theoretical Studies, Germany
[2] School of Computer Science, University of Manchester, UK
[3] Department of Biochemistry, University of Stellenbosch, South Africa
[4] Leiden Institute of Advanced Computer Science, Leiden University, NL
[5] ETH Zurich, Swiss
** Corresponding author:olga.krebs@h-its.org*

Key words: *Data management, Databases, SEEK, FAIRDOM, Data integration*

Motivation and Aim: Systems Biologists need a data management infrastructure that enables collaborating researchers to share and exchange information and data as and when it is produced, throughout the entire iterative cycle of experimentation and modelling.
Method: We develop and offer integrated data management support for systems biology research within and across research consortia comprising a whole package of solutions. This is applied to large-scale research initiatives in which we are responsible for the scientific data management, like the German Virtual Liver Network (http://www.virtual-liver.de/) and European research networks like ERASysAPP (ERA-Net for Systems Biology Applications), SysMO (Systems Biology of Microorganisms) or NMTrypI (New Medicines for Trypanosomatidic Infections), and Synthetic Biology Centres at Manchester (SynBioChem) and Edinburgh (SynthSys).
Results: Our data management concept consists of 4 major pillars:
1) Infrastructure backbone: The SEEK platform as registry and a commons for data, models, processes and resulting publications and presentations, at the same time yellow pages for projects, people and events
2) Terminology: Tailored use of controlled vocabularies and ontologies to describe the data
3) Modelling support: Seamless handling and simulation of models by integrated modelling platforms (JWS-Online, SYCAMORE, Cytoscape)
4) Social support: Data management advocates within the projects for gathering requirements and dissemination
The data management concept that we have developed is not only applied in the research consortia that we are responsible for, but also used by other systems biology projects.
Conclusion: Unlike the majority of data management systems, we specifically support the interaction between modelling and experimentation. Datasets can be associated with models and/or workflows, and model simulations can be compared with experimental data.

THE MITOCHONDRIA-TARGETED PLASTOQUINONE SKQ1 AFFECTS *DROSOPHILA MELANOGASTER* LIFESPAN IN VARIOUS ENVIRONMENTS

A.V. Krementsova[1]*, N.V. Roshina[2], E.A. Tsybul'ko[2], O.Y. Rybina[2], A.V. Symonenko[2], E.G. Pasyukova[2]

[1] *Emmanuel Institute of Biochemical Physics RAS, Moscow, Russia*

[2] *Institute of Molecular Genetics RAS, Moscow, Russia*

* *Corresponding author: krementsova@sky.chph.ras.ru*

Key words: *Antioxidant, Life span, Drosophila*

It was previously shown that mitochondria-targeted plastoquinone derivative SkQ1 (10-(6′-plastoquinonyl) decyltriphenylphosphonium) in extremely low nanomolar concentrations is able to prolong Drosophila melanogaster w^{1118} line male and female mean lifespan by about 10% [1]. To examine again if the effect of SkQ1 depends on the baseline life span and to determine whether SkQ1 affects life span of other, genetically unrelated lines of flies, three wild type lines, R340 (extremely low life span), Canton S (medium lifespan) and Oregon RC (high lifespan), which have no genetic similarity to the w^{1118} line. A significant increase in female and male survival was observed in experiments with lines R340 and Canton S, the effect of about 10% was similar to that observed previously for the w^{1118} line, but no significant increase in female and male survival was observed in experiments with the Oregon RC line.

We also assessed the effectiveness of SkQ1 treatment when severely changing environmental factors: temperature (18ºC and 8ºC), day and night darkened and diet (starvation, 12.5% and 25% of regular food supply) [2]. The action of these factors leads to a slower metabolism and thus to a reduction in generation of reactive oxygen species (ROS) in the cell and to increase longevity. It has been shown that the combined effect of several factors did not lead to a synergistic effect. Adding to the feed flies antioxidant SkQ1 not only led to a further increase in life expectancy of flies, but in some cases led to its reduction. Antioxidant SkQ1 proved to be quite effective (20%) with some deterioration of environmental conditions, but showed its low efficiency in extremely harsh stress conditions.

SkQ1 positively affected life span of individuals with different wild type genotypes living in a variety of environments; it demonstrated properties of a promising life-prolonging drug unsusceptible to fluctuations in the mean lifespan of recipients, methods of preparation and administration of the drug, seasons, or calendar years.

References

1. Krementsova AV, Roshina NV, Tsybul'ko EA, Rybina OY, Symonenko AV, Pasyukova EG: Reproducible effects of the mitochondria-targeted plastoquinone derivative SkQ1 on Drosophila melanogaster lifespan under different experimental scenarios. (2012) *Biogerontology*, 13(6):595-607.
2. Tsybul'ko EA Krementsova A, Symonenko AV, Rybina OY, Roshina NV, Pasyukova EG (2016) The mitochondria-targeted plastoquinone derivative SkQ1 promotes health and increases *Drosophila melanogaster* longevity in various environments. *J Gerontol A Biol Sci Med Sci* in press.

REGULATION OF BASE EXCISION REPAIR – CANONICAL AND NON-CANONICAL PROCESSING OF GENOMIC URACIL

H.E. Krokan*, H.S. Pettersen, R. Mjelle, S.A. Hegre, P. Sætrom, F. Drabløs, A. Sarno, A. Galashevskaya, P.A. Aas, N.B. Liabakk, B. Doseth, G. Slupphaug, B. Kavli
Norwegian University of Science and Technology, Trondheim, Norway
** Corresponding author: hans.krokan@ntnu.no*

Key words: *DNA repair, database, genomic uracil*

Generally, the molecular events in successive steps of DNA repair are much better understood than their regulation. Base excision repair (BER) is regulated at several levels, including posttranslational modifications, protein interactions, protein stability and cell cycle expression. We have focused our research on genomic uracil and it's repair. Genomic uracil may result from spontaneous or enzymatic deamination of cytosine, giving rise to mutagenic U:G mismatches, or from incorporation of dUMP during replication, giving rise to U:A pairs that are potentially mutagenic through errors in BER. Three major DNA glycosylases, UNG, TDG and SMUG1, initiate BER of genomic uracil. These proteins have different functions, as reflected by their catalytic properties, interactions and cell cycle regulation [1,2]. We recently established a database on the cell cycle regulation of all known DNA repair and chromatin remodeling proteins (www.dnarepairgenes.com) [3] and demonstrated that DNA glycosylases are differentially expressed. In most cells, BER supports essentially error-free repair of genomic uracil. However, in B-cells enzymatic deamination of cytosine to uracil by AID in Ig-genes is required for adaptive immunity, in which UNG has a non-canonical mutagenic role. Although essential as an immunological defense mechanism, this is also a risky process. Thus, we found that genomic uracil is significantly higher in B-cell lymphoma cell lines compared to non-lymphoma cancer cell lines and normal circulating lymphocytes, suggesting off-target deamination of cytosine to uracil by AID [4]. The genomic uracil levels correlated with AID mRNA and protein expression, but not with expression of other APOBECs. Accordingly, AID knockdown significantly reduced genomic uracil content. B-cells stimulated to express endogenous AID and undergo class switch recombination displayed a several-fold increase in total genomic uracil, indicating that B cells may undergo widespread cytosine deamination after stimulation. In line with this, we found that clustered mutations (*kataegis)* in B-cell lymphoma and chronic lymphocytic leukemia predominantly carry AID-hotspot mutational signatures. Moreover, we observed an inverse correlation of genomic uracil with uracil excision activity and expression of the uracil-DNA glycosylases UNG and SMUG1[4]. In conclusion, AID-induced mutagenic U:G mismatches in DNA may be a fundamental and common cause of mutations in B-cell malignancies.

1. Krokan H.E. and Bjørås M. (2013) Base Excision Repair. Cold Spring Harb. Perspect. Biol. 2013 Apr 1;5(4). a012583.
2. Krokan H.E., Sætrom P., Aas P.A., Pettersen,H.S., Kavli B. and Slupphaug G. (2014) Error-free versus mutagenic processing of genomic uracil – relevance to cancer. DNA Repair (Amst.)19:38-47.
3. Mjelle, R., Hegre, S.A., Aas, P.A. Slupphaug, G., Drabløs, F., Sætrom, P. and , Krokan, H.E. (2015) Cell cycle regulation of human DNA repair and chromatin remodeling genes. DNA Repair (Amst) 30:53-67.
4. Pettersen HS, Galashevskaya A., Doseth B., Sousa1 M.M.L., Sarno A., Visnes T., Aas P.A., Liabakk N.B., Slupphaug G., Sætrom P., Kavli B. and Krokan H.E. (2015) AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature. DNA Repair (Amst.) 25:60-71.

EFFECT OF LENTIVIRUS-MEDIATED SHRNA INACTIVATION OF HK1, HK2, AND HK3 GENES IN COLORECTAL CANCER AND MELANOMA CELLS

A.V. Kudryavtseva[1, 2]*, M.S. Fedorova[1], O.L. Kardymon[1], A.A. Dmitriev[1], A.I. Afremova[1], D.V. Kochetkov[1], A.V. Lipatova[1], A.F. Sadritdinova[1, 2], I.Y. Karpova[1], K.M. Nyushko[2], D.V. Kalinin[3], N.N. Volchenko[2], N.V. Melnikova[1], A.A. Belova[1, 2], M.A. Chernichenko[2], K.M. Klimina[4], N.V. Nasedkina[1], A.S. Zasedatelev[1], D.V. Sidorov[2], A.Y. Popov[2], G.S. Krasnov[1], A.V. Snezhkina[1]

[1] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

[2] Herzen Moscow Cancer Research Institute, Ministry of Health of the Russian Federation, Moscow, Russia

[3] A.V. Vishnevsky Institute of Surgery, Moscow, Russia

[4] Vavilov Institute of General Genetics, Russian Academy of Sciences Moscow, Russia

** Corresponding author: rhizamoeba@mail.ru*

Key words: *Warburg effect, hexokinases, shRNA, glycolysis, melanoma, colorectal cancer*

The switch from oxidative phosphorylation to glycolysis in proliferating cancer cell even under aerobic conditions has been shown first in 1926 by Otto Warburg. Today "Warburg effect" is known as a metabolic phenotype, a hallmark of malignant tumors. This shift is associated with alterations in signaling pathways involved in energy metabolism, including glucose uptake and fermentation, and regulation of mitochondrial functions. It has been shown that many genes encoding glycolytic enzymes are dysregulated in cancer. Hexokinases (HKs), which catalyze the first step of glycolysis, have been identified to play a role in tumorigenesis of human colorectal cancer (CRC) and melanoma. However, the mechanisms of HKs in the promotion of tumor growth remain elusive. The purpose of this study is to investigate the effect of silencing hexokinase genes (HK1, HK2, and HK3) in colon cancer (HT29, RKO, HCT116, CW480, and HCT15) and melanoma (Ksen, Kor, Z, and Cher) cells using short hairpin RNA (shRNA) lentiviral vectors. shRNA lentiviral plasmid vectors PLSLP-HK1, PLSLP-HK2, and PLSLP-HK3 were constructed and then transfected separately or co-transfected into cells. The results indicated that shRNA-mediated attenuation of HK2 and HK3 separately, as well as one together led to increased apoptosis rates of cancer cells and decreased glucose metabolism. HK1 gene inactivation did not result to significant changes in apoptotic rate or cells growth. Co-transfection by shRNA vectors against HK1, HK2, and HK3 together resulted in a rapid cell death by apoptosis. Thus, our results suggest that HK2 and HK3 genes are the key therapeutic targets for reducing aerobic glycolysis. This study shows a promising strategy for colorectal and melanoma cancer therapy.

This work was financially supported by grant 14-15-01083 from the Russian Science Foundation. Part of this work was performed using the equipment of EIMB RAS "Genome" center (http://www.eimb.ru/rus/ckp/ccu_genome_c.php).

COMPUTER SOFTWARE FOR STATISTICAL ANALYSIS OF GENES LOCATION RELATIVE TO CHROMOSOME CONTACTS REVEALED BY CHIA-PET

E.V. Kulakova*, A.M. Spitsina
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: kylakovaekaterina@gmail.com*

Key words: *sequencing, ChIA-PET, Hi-C, chromosome contacts, genome, CTCF sites*

Motivation and Aim: Several technologies based on chromatin immunoprecipitation (ChIP) have been developed to study the binding of transcription factors (TF) to genomic DNA including microarray (ChIP-chip), ChIP-PET and ChIP-Seq [1]. The challenge is to define whether such distal binding sites are functional, i.e. physically proximal to target gene promoters via chromosome loops attracting RNA polymerase II complex for gene transcription [2]. Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) method fits these demands still requiring development of specialized high-throughput software for data integration [2]. The aim of the work was to develop a computer program for statistical data analysis and test it on CTCF (CCCTC-binding factor) binding sites, genes and spatial topological domains.
Methods and Algorithms: We used data on the location of CTCF binding sites clusters obtained by ChIA-PET as well as obtained experimentally by methods Hi-C, ChIA-PET [2]. Gene annotation was obtained from UCSC Genome Browser (http://genome.ucsc.edu).
Results: In result has been developed computer software for statistical analysis and visualization of results for experimental data obtained by ChIA-PET and Hi-C. The program was developed in Java language that calls modules based on R and Matlab environment using library such as Rserve and MatlabControl. The program has graphical user interface. This tool has function such as identification gene location near to domains boundary; near to binding sites of transcription factor; visualization of heatmap based on pairs of CTCF binding sites; distributions of human genes relative CTCF binding sites and a randomly generated list of such sites.
Conclusion: We considered a model the location of genes relative chromosome loops and binding sites. Genes of RefSeq are located inside the loop between the sites accounted for half of the total. It was revealed that most of the genes in the chromosomal loops are arranged individually
Availability: Software is available from the author upon request
Acknowledgements: The work is supported by ICG budget project 0324-2015-0003.
References:
1. E.V. Ignatieva et al. (2015) Regulatory Genomics: Combined Experimental and Computational Approaches*, Russian Journal of Genetics*, V. 51, 4: 334-352.
2. G. Li et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation, Cell V.148. 1-2: 84-98.

SYSTEMIC ROLE OF ALLELIC VARIANTS IN A 2Q22 REGION IN MAJOR AGE-RELATED DISEASES AND LIFESPAN

A.M. Kulminski*, L. He, I. Culminskaya, Y. Loika, Y. Kernogitski, K.G. Arbeev, E. Loiko, L. Arbeeva, O. Bagley, M. Duan, A. Yashkin, F. Fang, M. Kovtun, S.V. Ukraintseva, D. Wu, A.I. Yashin

Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, USA

** Corresponding author: kulminsk@duke.edu*

Key words: *healthspan, lifespan, aging, geroscience*

Motivation and Aim: Gaining insights into genetic influences on age-related diseases and lifespan is a challenging task complicated by the elusive role of evolution in these phenotypes. Combining approaches from genome-wide and candidate-gene studies may be beneficial.

Methods and Algorithms: Genome-wide scan of participants of the Atherosclerosis Risk in Communities (ARIC) Study (N = 9,573) was used to pre-select promising loci. Candidate-gene methods were used to comprehensively analyze associations of novel uncommon variants in Caucasians (minor allele frequency~2.5%) located in band 2q22.3 with risks of coronary heart disease (CHD), congestive heart failure (CHF), stroke, diabetes, cancer, neurodegenerative diseases (ND), and mortality in the ARIC study, the Framingham Heart Study (N = 4,434), and the Health and Retirement Study (N = 9,676). We leveraged the analyses of pleiotropy, age-related heterogeneity, and causal inferences.

Results: Meta-analysis of the results from these comprehensive analyses shows that the minor allele increases risks of death by about 50% ($p = 4.6\times10^{-9}$), CHD by 35% ($p = 8.9\times10^{-6}$), CHF by 55% ($p = 9.7\times10^{-5}$), stroke by 25% ($p = 4.0\times10^{-2}$), and ND by 100% ($p = 1.3\times10^{-3}$). This allele also antagonistically influences each of two diseases, diabetes and cancer, in different populations. Combined significance of the pleiotropic effects was $p = 6.6\times10^{-21}$. Causal mediation analyses show that endophenotypes explained only small fractions of these effects. This locus harbors an evolutionary conserved gene-desert region with non-coding intergenic sequences likely involved in regulation of protein-coding flanking genes *ZEB2* and *ACVR2A*. This region is intensively studied for mutations causing severe developmental/genetic disorders.

Conclusion: Our analyses indicate a promising target region for interventions aimed to reduce risks of many major human diseases and mortality.

COMPARATIVE EXPRESSION LANDSCAPES IN REPLICATIVE AND STRESS INDUCED PREMATURE SENESCENCE

K.C. Kural[1], N. Tandon[2], O.V. Kel-Margoulis[2], A. Baranova[1, 3, 4]*
[1] *School of Systems Biology, George Mason University, Fairfax, VA USA*
[2] *geneXplain, Wolfenbüttel Germany*
[3] *Federal State Budgetary Institution "Research Centre for Medical Genetics", Moscow, Russia*
[4] *ATLAS Biomed Group, Moscow, Russia*
* *Corresponding author: abaranov@gmu.edu; aancha@gmail.com*

Key words: *senescence, ageing, expression landscapes, distance analysis*

Motivation and Aim: Senescence is the phenomenon accompanying the cessation of cell division in population of normal diploid cells. It can happen naturally in form of replicative senescence or may be a consequence of external challenges such as oxidative or radiation stress. It is widely believed that the senescence provides a protection against possible tumor formation. This research aimed at identifying differentially expressed gene signatures chracteristic for replicative and stress induced senescence in human fibroblast cells.
Methods and Algorithms: We employed geneXplain bioinformatics software platform. Our approach was to classify the differentially expressed genes by their functions to dissect the involvement of these genes in various cell cycle processes and the pathways involved in senescence. The list of genes which are exclusively up and down regulated in stress induced senescence and replicative senescence were compared. The genes unique for each of these two types of senescence were then subjected to promoter analysis. Potential transcription factor (TF) binding sites were used to deduce master regulators that control the activity of these downstream genes.
Results: Our analysis suggests that the Aurora-A kinase pathway is the master regulator central to both types of cell senescence. IL-1α, GM-CSF, MKP-2, MKK3, GDNF, TRIM36, MLTK are identified as specific contributors to replicative senescence, while BTEB-2, SHIP-110, SPK, and CDP play a role in stress induced senescence.
Conclusion: Our study outlines converging pathways for stress induced and replicative senescence, and connects mitochondrial stress with replicative cellular aging through accumulating failures of the centrosomal homeostasis.
Availability: on request.

TRANSCRIPTOMICS OF THE CRYOBIOTIC LEECH OZOBRANCHUS JANTSEANUS

S.V. Kuznecova[1], D. Suzuki[2], M.D. Logacheva[3], O.S. Kozlova[1], T. Kikawada[4], R.M. Sabirov[1], O.A. Gusev[1]

[1] *Kazan Federal University, Kazan, Russia*

[2] *Tokyo University of Marine Science and Technology, Tokyo, Japan*

[3] *M.V. Lomonosov Moscow State University, Moscow, Russia*

[4] *National Institute of Agrobiological Sciences, Tsukuba, Japan*

* *Corresponding author: svtkuzn@gmail.com*

Key words: *transcriptome, cyobiosis, turtle leech, Ozobranchus jantseanus*

Motivation and Aim: The *Ozobranchus jantseanus* is the turtle leech capable to survive after exposure to super-low temperature.[1] In contrast to known cryotolerant invertebrates the turtle leeches does not require prolonged pre-conditioning and resist immediate freezing. However, in natural habitat leeches are never exposed to such low temperatures. Further understanding of the genetic mechanisms underlying such a unique ability would contribute to development of new technologies in non-toxic cryopreservation of living cells and tissues of animals.

Methods and Algorithms: In the order to address the processes underlying cryotolerance ability we analyzed transcriptional changes in the leech on different stages of exposure to low temperature. We analyzed of mRNA expression in 4 groups of leeches: active animals, immediately after freezing, and further recovery for 3,5h and 24h after freezing.

Results: De-novo assembly of the transcriptome resulted in 99 624 coding transcripts. Among them, we observed dramatically changes in expression between control and 3,5h group. In 3,5h group we selected a cluster with 143 most up-regulated transcripts. Surprisingly that among them more than 25% are orphan genes and they weren't found in known organisms in data base. We observed increasing in RPKM value for transcripts coding tyrosinase, coadhesin, thrombospondin type 1, destabilase, properdin, brain-specific angiogenesis inhibitor, alpha-2-macroglobulin and etc. A significant increase in the expression value of transcripts of heat shock protein is not detected.

Conclusion: Among the most up-regulated by freezing transcripts expressed in the leech, 56% are represented by products of unknown genes. Furthermore, 25% of them are orphan, Ozobranchus-specific genes. We suppose that among them, there are some new candidate uncharacterized cryoprotectors and further in vitro studies will be done for deeper understanding of this phenomenon.

Acknowledgements: The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University

References:

1. Dai Suzuki, Tomoko Miyamoto, Takahiro Kikawada, Manabu Watanabe, Toru Suzuki (2014) A Leech Capable of Surviving Exposure to Extremely Low Temperatures, PLoS ONE 9(1): e86807. doi:10.1371/journal.pone.0086807 2.

A MATHEMATICAL MODEL FOR PREDICTING OF IGD–CD27+B LYMPHOCYTES LEVELS IN DONORS' BLOOD

S.R. Kuznetsov[1]*, I.V. Kudryavtsev[2], A.V. Orekhov[1], A.V. Polevshchikov[2], M.K. Serebriakova[2], V.I. Shishkin[1]

[1] *St. Petersburg State University, St. Petersburg, Russia*

[2] *Institute of Experimental Medicine, St. Petersburg, Russia*

** Corresponding author: abxy01@yandex.ru*

Key words: *mathematical model, systems immunology, flow cytometry, T cells, B cells*

Motivation and Aim: Previously we proposed to use McKendrick–von Foerster age-structured population model to describe proliferation and differentiation processes of T and B-lymphocytes in the immune response [1, 2]. In this work McKendrick–von Foerster equation is used to define parameters of IgD+CD27+ lymphocytes proliferation in a healthy person. As a result, a formula was built, allowing to predict the level of IgD– CD27+ lymphocytes from the data of IgD+CD27+ B lymphocytes level and CD3+CD4+CXCR5+CXCR3–CCR6–CCR4+ Tfh2 lymphocytes level.

Methods and Algorithms: We supposed that levels of IgD+CD27+, IgD–CD27+ and Tfh2 cells of a healthy person do not changes over the time, and it is possible to use a stationary solution of McKendrick–von Foerster equation. Optimal parameters of the model were determined by the Monte Carlo method, using the data on the level of lymphocytes in the blood of 30 donors. For verification of the model, there was used data from 19 donors. Based on verifying data (in particular concentrations of IgD+CD27+ B cells and Tfh2) there were made predictions of the level of IgD–CD27+ B cells for each of 19 donors; the result was compared with experimental data.

Results: The formula has been obtained to determine the level of IgD–CD27+B-lymphocytes from the data of IgD+CD27+ and Tfh2 lymphocytes levels in the venousblood of donors. Verification revealed the comparison of experimental data with prediction made by the model; the correlation coefficient for 19 donors was 0.65.

Conclusion: The proposed model reliably describes the process of IgD+CD27+B-lymphocytes proliferation in a healthy person.

Availability: The derived computational scheme for predicting the level of IgD–CD27+B cells can be applied in a clinical practice.

References:

1. S.R. Kuznetsov (2015) Mathematical model of the immune response, Vestnik S.-Petersburg Univ. Ser. 10. Prikl. Mat. Inform. Prots. Upr., 4: 72–87. (In Russian).
2. S.R. Kuznetsov (2015) Mathematical Model of the Humoral Immune Response: Focusing On Thl7 Autoimmunity, Math. Biol. Bioinf., 10(2):455-472. doi: 10.17537/2015.10.45

TARGET ENRICHMENT TECHNOLOGIES FOR APPLIED RESEARCH

D.A. Kwon
Agilent Technologies Russia, Moscow, Russia
** Corresponding author: dmitry.kwon@agilent.com*

Key words: *NGS, target enrichment, amplification, hybridization, custom*

Motivation and Aim: (text): NGS as an approach gave a powerful breakthrough in genomics, transcriptomics and epigenetics research and provides a tool for evaluation of wide range of principles, regularities and even mechanisms of genome functioning. The current stage of NGS is an applied research for clinical and biotechnological fields. Where not whole genome\transcriptome but more discrete objectives appears. Target enrichment – is a group of methods for library preparation for NGS to study genome\transcriptome in any scale – from single gene up to whole functionally significant part of.
Methods and Algorithms: (text): Agilent Technologies, Inc. provides wide menu of target enrichment kits, using two technologies – amplification (Haloplex technology) for small amounts of starting material and hybridization (SureSelect and OneSeq). These kits are able to provide high quality NGS libraries with excellent *on target* index for wide range of experiments – from small number of genes up to whole-exome and NGS-based cytogenetics for comprehensive, all-in-one detection of genome-wide CNVs, copy-neutral LOH (cnLOH), SNPs, and indels analysis in one target enrichment capture.
Important to mention that any of the kits above can be used with custom design for any object.
Also Agilent Technologies, Inc. provides wide range of NGS-related products for

- *In silico* Design of custom NGS Target Enrichment panels
- NGS data analysis software
- NGS libraries quality control
- NGS workflow automation

CHARACTERIZATION OF NOVEL ALKANE-DEGRADING AND BIOSURFACTANT-PRODUCING STRAIN *TSUKAMURELLA TYROSINOSOLVENS* PS2

A.V. Laikov[1]*, E.A. Boulygina[1], V.A. Romanova[1], T.V. Grigorieva[1]
[1] *Kazan Federal University, Kazan, Russia*
* *Corresponding author: a.v.laikov@gmail.com*

Key words: *Tsukamurella tyrosinosolvens, biosurfactants, alkane, trehalose lipids*

Motivation and Aim: Crude oil is a necessary condition for the existence of the world economy and industrial growth. The volume of its exploration and production is constantly increased, which results in massive concomitant environmental pollution. Among other methods of oil contaminants removal from soil and water the use of hydrocarbon oxidizing bacteria takes a special place. Novel microorganisms with enhanced oxidizing activity or valuable biotechnological features (such as ability to produce biosurfactants) are being searched permanently. Biosurfactants, revealed by bacteria, allow microbial cells to make a contant with hydrophobic hydrocarbon substrate, increasing their biological availability both for biosurfactant-producing bacteria and for other members of the community [1]. The aim of our study is characterization of biosurfactants and finding its genetic determinants.

Methods and Algorithms: From solid chemical waste the bacterial isolate, which was able to utilize alkanes as the sole carbon and energy source, was extracted. According to the classification by 16S rRNA sequence, this isolate was identified as *Tsukamurella tyrosinosolvens str.* PS2. The strain was sequenced on the MiSeq (Illumina) platform, assembled and annotated. Decreasing of surface tension was evaluated on the area of culture medium droplets on the hydrophobic surface. Emulsifying activity was assessed by the optical density of the emulsion [2].

Results: Culture fluid of *T. tyrosinosolvens* PS2 formed a stable emulsion while being mixed with hexadecane. The medium with hexadecane appeared to have the highest biosurfactants output in comparison with the medium containing glucose or sucrose. It was found that biosurfactants of this bacteria are of non-ionogenic nature.

The genome sequencing of *T. tyrosinosolvens* PS2 revealed genes of 2 trehalose synthesis pathways (trehalose-6-phosphate synthase, trehalose-6-phosphate phosphatase, maltooligosyl trehalose synthase and malto-oligosyltrehalose trehalohydrolase). In addition, mycolyltransferase genes were found. Based on this, we assume that biosurfactants of *T. tyrosinosolvens* PS2 belong to the class of trehalose lipids [3].

Conclusion: New strain *T. tyrosinosolvens* PS2 is able to simultaneously oxidize alkanes and produce biosurfactants into culture fluid. Some properties of these compounds have been characterized. Genomic approach allowed us to identify possible ways of biosurfactant synthesis, with target genes for further research.

References:

1. R.S. Reis *et al*. (2013). Biosurfactants: Production and Applications, Biodegradation - Life of Science, Dr. Rolando Chamy (Ed.), 31-61 (InTech).
2. P.U. Mahalingam *et al*. (2014) Isolation, characterization and identification of bacterial biosurfactant. *Europ. J. Exp. Biol*. 4(6): 59-64.
3. B. Tuleva *et al*. (2008). Production and structural elucidation of trehalose tetraesters (biosurfactants) from a novel alkanothrophic *Rhodococcus wratislaviensis* strain. *J. Appl. Microbiol.* 104(6): 1703-1710.

3D MAP OF PROLIFERATION ACTIVITY IN *ARABIDOPSIS THALIANA* ROOT TIPS: TRANSITION DOMAIN BOUNDARIES AND ITS BILATERAL SYMMETRY

V.V. Lavrekha[1, 2]*, T. Pasternak[3], N.A. Omelyanchuk[1, 2], V.B. Ivanov[4], V.V. Mironova[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *LCTEB, Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Biology II/Molecular Plant Physiology, Centre for BioSystems Analysis, BIOSS Centre for Biological Signalling Studies University of Freiburg, Germany*
[4] *Timiryazev Institute of Plant Physiology, Russian Academy of Sciences, Moscow, Russia*
** Corresponding author: vvl@bionet.nsc.ru*

Key words: *mitosis, root apical meristem, transition domain, bilateral root symmetry, Arabidopsis thaliana*

Background and Aims: The plant root due to its near regular arrangement of cells in files is one of the most suitable experimental system to study growth and developmental processes. Total root length reflects both cell proliferation and cell elongation rates. Cell proliferation occurs in the root apical meristem (RAM) located in the end of the root tip. Cells specification in the end of the RAM leads to the cell cycle arrest. Up to date the RAM structure (length of the proliferation domain, transition to elongation, arrangement of specific cell lineages) was mainly analyzed on 2D images, and this approach has some weaknesses considering bilateral symmetry of the root.
Methods: Recently, iRoCS toolbox was developed [1] for annotation of the root tip organization in three dimensions. Together with the refined experimental procedure for detection of cell cycle progression [2], it provides an unprecedented potential to study the mechanism of cell cycle regulation in an entire organ. Here, using these techniques, we analysed the distributions of the key cell cycle events – DNA replication and mitosis in the root tips of *Arabidopsis thaliana*. Seeds were sterilized and grown on AM medium in plates with 22 °C temperature under 16/8 light cycle conditions. 5-th day old seeds were incubated in 200 μg/l DAPI/10 μM EdU and 1 mg/ml colchicine for 90 minutes. Then, the root tips were investigated using a confocal laser scanning microscope (LSM 510 Duo Live).
Results: Annotation of the confocal images with EdU and DAPI labelling using iRoCS toolbox allowed us to give a new insight on the root tip zonation. Namely, we quantitatively showed that the proliferation activity differs for distinct cell types and files. The differences were associated with bilateral and radial symmetries of the root. In all cell files DNA replication events occurred after the last mitosis and before transition to rapid cell growth.
Conclusion: As a result, the concept of the transition zone emerged which in *A. thaliana* root meristem locates between the last mitoses and last DNA replication events in the different cell files.
Acknowledgments: This research was supported in by the Budget project № 0324-2015-0003 and RSF 14-14-00734.

1. Schmidt, T., Pasternak, T., Liu et al (2014). The iRoCS Toolbox–3D analysis of the plant root apical meristem at cellular resolution. The Plant Journal, 77(5), 806-814.
2. Pasternak, T., Tietz, O., Rapp, K., Begheldo, M., Nitschke, R., Ruperti, B., & Palme, K. (2015). Protocol: an improved and universal procedure for whole-mount immunolocalization in plants. Plant methods, 11(1), 1

POLY(ADP-RIBOSE) POLYMERASE 1 AND REGULATION OF DNA REPAIR

O.I. Lavrik[1, 2]
[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: lavrik@niboch.nsc.ru*

Key words: *PARP1, PARP2, poly(ADP-ribose), BER, protein-protein interaction*

Motivation and aim: The phenomenon of nicotinamide adenine dinucleotide (NAD+)-dependent poly(ADP-ribosyl)ation catalyzed with PARP1 was discovered long time ago, but it is still unclear how this post-translational modification governs a multitude of cellular processes including DNA repair. When interacting with the damaged DNA, PARP1 catalyzes the synthesis of a long branched poly (ADP-ribose) polymer (PAR) by using NAD^+ as a substrate. PAR can be attached to the acceptor amino acid residues of nuclear proteins or to PARP1 itself. This process leads to reorganization of the functional protein complexes involved in base excision repair (BER) and other key processes in cell. The aim of the present research was to investigate the role of poly (ADP-ribosyl)ation in regulation of BER and to search new targets of PARylation catalyzed with PARP1 and PARP2. The role of DNA breaks in PARylation was investigated. The protein–protein interactions in BER were analyzed and quantified in the presence of BER DNA intermediates.
Methods: Fluorescence titration methods, atomic force microscopy (AFM), light-scattering technique, biochemical and immunochemical approaches.
Results: PARP1 interacts with BER proteins as well as with DNA intermediates of BER containing breaks or apurinic/apyrimidinic (AP-sites) which appear in BER process. PARP1 interacting with the AP sites shows AP lyase and 5'-dRP lyase activities. Protein-protein interactions of PARP1 with APE1, Pol beta, XRCC1, tyrosyl-DNA-phosphodiesterase 1 and other components of BER machine were investigated quantitatively by various methods. The strength of protein-protein interactions in BER was influenced by structure of DNA repair intermediates. The specificity of PARP1 and PARP2 interaction with various DNA structures as well as the active role of DNA in PARylation was approved.
Conclusion: The results obtained show that PARP1 and PARP2 interact specifically with different kinds of damaged DNA and DNA breaks play an active role in poly(ADP-ribosyl)ation. The protein-protein interactions and its regulation were estimated quantitatively at the various stages of BER.
Acknowledgements: This work was supported by grant from RSF (14-24-00038).

TOWARDS UNDERSTANDING THE DYNAMICS OF DEATH RECEPTOR NETWORKS

I.N. Lavrik[1, 2]
[1] *Translational Inflammation Research, Medical Faculty, Center of Dynamic Systems, Otto von Guericke University, Magdeburg, Germany*
[2] *Laboratory of systems biology of programmed cell death, Institute of Cytology and Genetics, Novosibirsk, Russia*

Key words: *gene networks, modeling, programmed cell death*

Our studies are devoted to the analysis of death receptor networks by systems biology approaches. The members of the death receptor family control programmed cell death (PCD) and proliferative pathways. PCD is essential for regulation of homeostasis and elimination of unneeded, damaged, or infected cells in multicellular organisms. PCD deregulation contributes to cancer, as well as neurodegenerative and autoimmune diseases. Creation of mathematical models of death receptor signaling led to an enormous progress in the quantitative understanding of the network regulation and provided fascinating insights into the mechanisms of death receptor control. The key step in the initiation of the death receptor-induced apoptosis is the activation of caspase-8 at the death receptor complex. To understand the dynamics of caspase-8 activation we have developed an agent-based model (Schleich et al, 2012). Interestingly, this mathematical model supported by quantitative mass-spectrometry and western blot experimental data allowed to find out that different stimulation strength results in the distinct composition of the death receptor complexes that contribute to apoptosis induction in a different manner. Using the advanced version of this agent-based model we have discovered recently a negative feedback-loop in procaspase-8 activation (Schleich et al., 2016). Furthermore, the model has allowed to delineate the dynamics of caspase-8 activation events at the death receptor activation complex and suggest the new targets for the development of small molecules, which in turn might be used for the development of new therapies. Overall, these findings provide new insights into caspase-8 activation and apoptosis initiation and underline the power of systems biology in analyzing complex apoptotic networks.

References:

1. Schleich K, Warnken U, Fricker N, Oztürk S, Richter P, Kammerer K, Schnölzer M, Krammer PH, Lavrik IN. Stoichiometry of the CD95 death-inducing signaling complex: experimental and modeling evidence for a death effector domain chain model. Mol Cell. 2012 Jul 27;47(2):306-19.
2. Schleich K, Buchbinder JH, Pietkiewicz S, Kähne T, Warnken U, Öztürk S, Schnölzer M, Naumann M, Krammer PH, Lavrik IN. Molecular architecture of the DED chains at the DISC: regulation of procaspase-8 activation by short DED proteins c-FLIP and procaspase-8 prodomain. Cell Death Differ. 2016 Apr;23(4):681-94.

Support

The work is supported by Russian Science Foundation 14-44-00011.

DOES THYROID DIVERGENCE SERVE AS A DRIVER OF SPECIATION IN CYPRINID FISHES OF THE GENUS *BALLERUS* (TELEOSTEI)?

B.A. Levin[1]*, A.A. Bolotovskiy[1], M.A. Levina[1], A.V. Nedoluzhko[2], K.G. Skryabin[2, 3, 4], S.M. Rastorguev[2], E.B. Prokhortchouk[3, 4]

[1] *Institute of Biology of Inland Waters RAS, Borok, Russia*

[2] *National Research Center Kurchatov Institute, Moscow, Russia*

[3] *Institute of Bioengineering, Federal Research Center "Fundamentals of Biotechnology" RAS, Moscow, Russia*

[4] *Lomonosov Moscow State University, Faculty of Biology, Moscow, Russia*

** Corresponding author: borislyovin@mail.ru*

Key words: *thyroid hormones, fish, phenotype, RNA-seq, evolution, speciation*

Motivation and Aim: Regulatory evolution is one of the most important mechanisms for origination of novel morphological complexity and evolutionary innovations. Thyroid hormones (TH) regulate development of many phenotypic traits in fishes. We detected sister species of the genus *Ballerus* (Cyprinidae), which were naturally diverged in TH level and are different in phenotypic traits, which morphogenesis is regulated by TH. The aim of this study is to testify the hypothesis about involvement of TH level divergence in evolution of sister species via experimental manipulation of TH level during early development and analysis of derived phenotypes and gene expression.

Methods and Algorithms: The progeny from naturally TH deficient species *B. ballerus* was treated by T_3 during early ontogeny [1]. Enzyme-linked immunosorbent assay (ELISA). Phenotype analysis (counts and measurements). RNA-seq by Illumina GAIIx. Liver and brain, totally 12 cDNA-libraries from control and TH-treated fish. Transcriptome *de-novo* by Trinity. Differential expression by edgeR. GO-enrichment analysis. Details [2].

Results: The phenotype of T_3-treated fish of the TH-deficient species *B. ballerus* was approached to sister species with naturally higher TH level, *B. sapa*, in the numbers of lateral line scales and gill rakers, and in eye size. RNA-seq revealed more than 1200 differentially expressed genes (DEGs) between control and T_3-treated fish. Many DEGs were involved in determination of morphological traits. Dozen DEGs were homeobox.

Conclusion: Pronounced effect of TH on regulation of both phenotype and gene expression was found. Natural divergence in thyroid level might be a trigger for speciation in *Ballerus*.

Acknoweldgements: Study was supported by Russian Foundation for Basic Research (project nos. 15-04-3586 and 15-34-20416), the transcriptome sequencing was supported by the Russian Science Foundation (project no. 14-24-00175).

References:

1. B.A. Levin, M.A. Levina. (2014) Poly-and oligomerization of scales in the blue bream Ballerus ballerus (Cyprinidae) as a consequence of thyroid status regulation. J. Appl. Ichthyol., 12: 809-813.
2. S.M. Rastorguev et al. (2016) Pleiotropic effect of thyroid hormones on gene expression in fish as exemplified from the blue bream Ballerus ballerus (Cyprinidae): Results of transcriptomic analysis. Dokl. Biol. Sci., 467: 124-127.

DIFFERENTIAL ANALYSIS OF THREE-DIMENSIONAL (3D) GENOMICS DATA

G. Li[1]*
[1] *Huazhong Agricultural University, Wuhan, China*
** Corresponding author: guoliang.li@mail.hzau.edu.cn*

Key words: *gene expression, chromatin, 3D chromosome structures, chromosome contacts, CTCF sites, ChIA-PET*

Motivation and Aim: Studying of 3D chromosome structure is important problem of molecular biology challenging sequencing technologies. ChIA-PET (Chromatin Interaction Analysis with Paired-End-Tag) technology allows detect interactions between pairs of DNA sites affecting gene regulation. Fullwood et al. [1] used ChIA-PET technology to construct chromatin interaction network bound by estrogen receptor alpha (ER) from human breast cancer cell line (MCF-7) and found long-range ER binding sites are mostly located at promoter regions. CTCF-mediated interactions found in mouse embryonic pluripotent stem cells and human cell lines [2].
Methods and Algorithms: We developed computer programs for 3D genomics data analysis. The data have been obtained experimentally by using Hi-C and ChIA-PET methods [3].
Results: Five distinct chromatin domains revealed by CTCF ChIA-PET raised a new model of CTCF function for chromosome structure organization and linking enhancers to promoters for gene transcription regulation.
Conclusion: Chromatin interaction network will be discussed.

References:

1. M.J.Fullwood et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome, *Nature*, 462(7269):58-64.
2. G.Li et al. (2014) Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application, *BMC Genomics*, 15(Suppl 12):S11.
3. Z. Tang et al. (2015) CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription, *Cell,* 163(7):1611-27.

GENOTYPE DISTRIBUTION IN PATIENTS WITH CHRONIC HEPATITIS C ANALYSIS USING MULTIFACTOR DIMENSIONALITY REDUCTION METHOD

A.D. Liaudanski[1]*, M.S. Rodzkin[1], V.S. Pankratov[1], D.E. Danilau[2], I.A. Karpov[2], O.G. Davydenko[1]

[1] *Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Belarus*

[2] *Belarusian State Medical University, Minsk, Belarus*

** Corresponding author: 666555@tut.by*

Key words: *chronic hepatitis C, gene polymorphism, MDR method*

Motivation and aim: hepatitis C virus is one of the major causes of chronic liver pathologies. Near 2% of world population suffer from chronic hepatitis C, which determine high relevance of the studying of spontaneous elimination and therapy efficiency influencing factors [1]. The prevalent type of chronic hepatitis C therapy includes pegylated interferon in combination with ribavirin, and IL28b gene polymorphism has been announced to be the most informative predictor of such therapy efficiency [2,3]. Taking in account the accuracy of the therapy efficiency prediction based on IL28b gene polymorphism (up to 50% in some population) it is still actual to find other genetic markers to enhance the precision of prognostication.

Methods and Algorithms: 100 patients with chronic hepatitis C, who had taken pegylated interferon and ribavirin therapy with different (67 unsuccessful and 33 successful) outcomes were analyzed. Polymorphism of IL28b, TNFα, CCR5 and CCL5 genes was defined using PCR or PCR-RFLP methods. Multifactor dimensionality reduction method was used to find the best model for therapy efficiency prediction (MDR ver.3.0.2 (build 2)).

Results: IL28b was found to be the best single marker, as expected (accuracy=0,61, CV consistency 10/10, p=0,0003). But the best accuracy was demonstrated by two-factor model, including IL28b and CCL5 genotypes (accuracy=0,77, CV consistency 10/10, p<0,0001). CCL5 genotypes distribution didn't vary significantly in patients with different therapy outcome, but including this in the prognosis model improve the prediction significantly. Entropy distribution demonstrated more than 1,8-fold value increase when comparing IL28b and CCL5 together with IL28b alone: IL28b entropy was 11.57%, CCL5 – 1.75%, IL28b and CCL5 interaction – 7,56% (20,88% in total). Three- and four-factor models were found to be much less informative.

Conclusion: multifactor dimensionality reduction proved to be an effective method for genotype distribution in patients with chronic hepatitis C analysis. Two-factor model including IL28b and CCL5 genotypes was found to be the most accurate model for prediction of the chronic hepatitis C therapy efficiency. Taking in account CCL5 genotype of the patient together with IL28b could help to improve the prognosis precision of such therapy outcome.

References:

1. C.W. Shepard et al. (2005) Global epidemiology of hepatitis C virus infection, Lancet Infect. Dis., 5: 558-567.
2. D. Ge et al. (2009) Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance, Nature, 17: 399-401
3. C.N. Hayes et al. (2012) Genetics of IL28B and HCV--response to infection and treatment, Nat Rev Gastroenterol. Hepatol., 9:406-417

POSTGENOME MEDICINE AS N-OF-ONE SCIENCE

A.V. Lisitsa, E.V. Kolker, H. Chen, V.E. Frankevich
IBMC, Moscow, Russia

Key words: *postgenome medicine, diagnostic patterns, molecular profiling*

N-of-one science is destined to serve a particular individual who is considered as an "uber-client" but not a "patient". Normally we observe n-of-one science in connection with gadgets and Internet-services for non-obtrusive life couching. For example, number of steps taken a day is counted by owners of the wearable devices. The postgenome medicine rests upon the application of multi-omics technologies to observe human physiology at the molecular level. Instead of searching a singe-molecule biomarkers of disease, the postgenome medicine operates with an ensembles of molecules, which comprise the diagnostic patterns of health and disease. By definition the molecular signatures are surrogate, as the specific disease-related biomarkers cannot be revealed due to analytical limit of detection. We observed that molecular profiling could specifically characterize the human health in case it is collected sequentially over a period of time. That provides a technology challenge to collect these profiles 100 times faster and at least 10 times cheaper in comparison to the biochemical blood-tests. The methods of direct mass-spectrometry coupled with novelties in technology of ionization and ion separation are described as a possible solution of n-of-one challenge for the sake of the postgenome medicine.

DNA DUPLEX STRUCTURE AND THERMODYNAMICS BY MOLECULAR DYNAMICS SIMULATION

A.A. Lomzov*, D.V. Pyshnyi
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
** Corresponding author: lomzov@niboch.nsc.ru*

Key words: *DNA, molecular dynamics simulation, hybridization thermodynamics*

Motivation and Aim: Prediction of the thermal stability of double stranded DNA is useful in basic research (e.g. protein – nucleic acid interaction) and in a number of applications (e.g. PCR, biosensing). The development of new derivatives and analogs of nucleic acid with improved characteristics for purposes of molecular biology, biotechnology and biomedicine is still of current interests. New derivatives not always have the same physico-chemical and molecular-biological properties as were proposed at development. The essential problem is the reliable prediction of hybridization properties of nucleic acids derivatives and analogues. The nearest-neighbor model is the only one which allows calculating enthalpy, entropy and Gibbs free energy changes only for experimentally well-studied native nucleic acid and for a limited number of their derivatives. The development of new tools for reliable prediction of nucleic acids complex properties is a serious challenge. Recently we have shown the possibility of high accuracy calculation of hybridization enthalpy for native fully complementary DNA duplex [1]. In this work we have evaluate the hybridization enthalpy and Gibbs free energy.
Methods and Algorithms: Molecular dynamics simulations of a set of oligonucleotides and its complexes were performed in implicit and explicit solvent shell using parmbsc0 force field in AMBER12 software package. Conformational contribution in hybridization entropy was evaluated using normal mode analysis and quasi-harmonic calculations.
Results: Calculated values of conformational entropy correlate well with experimental data and values calculated using nearest neighbor model [2]. The linear correlation between the values calculated using quasi-harmonic and nearest-neighbor models has a slope 1.05, intercept -84.0 cal/mol/K and R2=0.97 for 305 DNA complexes of different length (4-20 bp) and GC-content (0-100%).
Conclusion: The use of combination of molecular dynamic simulation in explicit solvent with MM/PBSA and Quasi-harmonic calculation allows good evaluation of thermodynamic parameters of formation perfectly matched DNA duplexes.
Availability: This is a first step in prognostic calculations of physico-chemical properties of nucleic acids derivatives and development of new analogs with predetermined characteristics.
Acknowledgements: This work was supported by RFBR #16-04-01029.
References:
1. A.A.Lomzov et al. (2015) Evaluation of Gibbs free energy changes and melting temperatures of DNA/DNA duplexes using hybridization enthalpy calculated by molecular dynamics simulation, The Journal of Physical Chemistry B, 119: 15221–15234.
2. A.A.Lomzov et al. (2006) Thermodynamic parameters for calculating the stability of complexes of bridged oligonucleotides, Doklady Biochemistry Biophysics, 409: 211-215.

EVOLUTION OF PHENOTYPIC CONTROL BY NEW GENES THROUGH INTEGRATING AND REWIRING OF ANCESTRAL EXPRESSION NETWORKS

M. Long*
Department of Ecology and Evolution, University of Chicago, Chicago, USA

Key words: *evolution, gene networks*

We have observed that new genes quickly evolved essential functions in mammals and fruit flies.These observations, while overturning the conventional notion of evolutionarily static protein coding gene functions, raised a new question of how a new gene acquires essential functions.We proposed a hypothesis that a new gene may be integrated into and reshape ancestral gene-gene interaction (GGI) networks and thus subsequently create a hub topological structure with essential genetic control of a phenotype distinct from the new gene absent species that maintain ancestral networks. We examined the network topological and functional evolution of new genes that originated at various stages in humans, mouse and fruit flies, by constructing and analyzing various GGI networks. We computationally and experimentally identified a large number of new genes in various topological positions in GGI networks, showing distinct evolutionary patterns in mammalian and Drosophila lineages. These genes experienced a stepwise integration process into GGI networks, starting on the network periphery and eventually becoming highly connected hubs, and acquiring pleiotropic and essential functions. We identified species-specific hub genes that have evolved critical or essential functions in development, brains and behaviors, supported by increasing mechanistic analyses of young genes in literatures. We explored the possible underlying mechanisms driving the GGI network evolution and the observed patterns of new gene integration process and identified certain mechanisms distinct from interpretations of general network scientific theories. We propose that the difference in effective population sizes in humans and fruit flies plays a significant role in shaping their strikingly different evolutionary patterns.

References:

1. Chen et al, 2013. Nature Reviews Genetics. 14: 645–60.
2. Zhang et al, 2015. Genome Biology. 16:202.
3. Krinsky et al, 2016. bioRxiv.dx.doi.org/10.1101/045385
4. Andrea R Gschwend

TARGETED SPATIAL GENOME MODIFICATION IN TOPOLO-GICALLY ASSOCIATING DOMAINS STRUCTUREIN MOUSE EMBRYONIC STEM CELLS

V.A. Lukyanchikova*, N.R. Battulin, O.L. Serov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: lukyanchikova_varvara@mail.ru*

Key words: *3D-genome organization, TADs, CTCF sites, genomic deletions, CRISPR/Cas9, mESCs*

Motivation and Aim: it was recently shown that genomes of higher eukaryotes are partitioned at a sub-megabase self-interacting units termed topologically associating domains (TADs) (Dixon et al., 2012). Molecular basis of TADs structure, the mechanism of its maintenance, and the role in regulation of gene expression are still important fundamental questions. Due to the latest reports, cohesin complex and CTCF protein are key players of TADs formation (Guo et al. 2015, Nichols et al., 2015), whilst dislocations in domain boundaries and disruptions of CTCF-binding sites can lead to faults in gene regulation and to activation of oncogenes (Lupiáñez et al., 2015, Hnisz et al., 2016). So, generation of cell lines and animal models with modified TADs boundaries can reveal the natural meaning of TADs integrity and mechanisms of TADs maintenance in general.
Methods and Algorithms: we were focused on the murine genomic locus between *Kdr* and *c-Kit*, chr5:76,240,000-76,280,000 (UCSC Genome Browser on Mouse July 2007 (NCBI37/mm9) Assembly). This region contains well-defined TAD boundary co-localized with CTCF-binding sites, while neighboring *Kdr* and *c-Kit* genes are mainly involved in embryonic development and have a detectable level of expression in mESCs. To remove the chosen region from mESC genome we utilized CRISPR/Cas9 editing tool (Mali et al., 2013b). Real-time PCR was used to characterize alterations in gene expression pattern.
Results and Conclusion: we have produced mESCs lines with genomic deletions of 1, 2 and 4 CTCF-binding sites in *Kdr*/*c-Kit*-region. Karyotype screening of all obtained cell clones has been performed, and gene expression from only ones which have balanced karyotype was analyzed with real-time PCR. Data analysis has demonstrated the significant shift in expression patterns of neighboring genes which can be evidence of gene regulation impairment due to deletion of TAD boundary.
Acknowledgements: This project was supported by the Russian Foundation for Basic Research (grant 16-04-01453 A).

References:

1. Dixon, Jesse R., et al. "Topological domains in mammalian genomes identified by analysis of chromatin interactions."*Nature* 485.7398 (2012): 376-380;
2. Guo, Ya, et al. "CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function." *Cell* 162.4 (2015): 900-910;
3. Nichols, Michael H., and Victor G. Corces. "A CTCF code for 3d genome architecture." *Cell* 162.4 (2015): 703-705;
4. Lupiáñez, Darío G., et al. "Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions." *Cell* 161.5 (2015): 1012-1025;
5. Hnisz, Denes, et al. "Activation of proto-oncogenes by disruption of chromosome neighborhoods."*Science* 351.6280 (2016): 1454-1458;
6. Mali, Prashant, et al. "RNA-guided human genome engineering via Cas9." *Science* 339.6121 (2013): 823-826.

THE DENSITY OF *WOLBACHIA* STRAIN *w*MELPOP IN *DROSOPHILA MELANOGASTER* BRAIN IS INVERSELY RELATED TO THE LEVEL OF *HSP67BC* GENE EXPRESSION

D.A. Malkeyeva[1, 2]*, E.V. Kiseleva[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author:malkeyeva@bionet.nsc.ru*

Key words: *Wolbachia, heat shock proteins, hsp67Bc*

Motivation and Aim: Wolbachia are the widespread endosymbionts of arthropods and nematodes. These bacteria are known for having a significant impact on hosts' reproductive functions, longevity and gene expression. *Wolbachia* down-regulates 41% of heat shock protein genes, including *hsp67Bc*, in *Drosophila melanogaster* S2 cells [1]. Hsp67Bc was shown to stimulate macroautophagy [2], a mechanism by which *Wolbachia* bacteria are removed from host cells [3]. In this study, we investigated the interrelation between the expression of *hsp67Bc* gene and the density of pathogenic *Wolbachia* strain *w*MelPop in *D. melanogaster* brain.

Methods and Algorithms: To obtain *D. melanogaster* with down-regulated *hsp67Bc* gene expression we used the imprecise excision of the P-element from *hsp67Bc* promoter region, which resulted in a deletion in *hsp67Bc* gene. In this work, flies with one (mutants) of two (control) copies of the gene were used. Overexpression of *hsp67Bc* in *D. melanogaster* was obtained using GAL4-driven expression of *UAS-hsp67Bc* construct. *Wolbachia* in fly brains were visualized by fluorescence *in situ* hybridization with Cy3-labeled W2 probes specific to *Wolbachia* 16S rRNA. The quantity of endosymbionts was measured as *Wolbachia*-occupied area/brain area (S_w/S_b) ratio on brain optical sections with the use of ImageJ 1.48 software.

Results: In our study, an inverse relation between *hsp67Bc* expression and the quantity of pathogenic *w*MelPop strain of *Wolbachia* bacteria was determined. In *D. melanogaster* with one copy of *hsp67Bc* the S_w/S_b ratio was 2.7 times higher than in control flies (4.82±1.69 % compared to 1.75±1.00 %), whereas in flies overexpressing *hsp67Bc* the S_w/S_b ratio was 2.4 times lower than in control animals (1.21±0.29 % in comparison with 2.97±0.65 %).

Conclusion: Deletion of one of two copies of the host's *hsp67Bc* gene results in the over-replication of *Wolbachia w*MelPop strain in *D. melanogaster* brain. Overexpression of *hsp67Bc* not only retains the endosymbionts' titer, but also reduces the titer of *Wolbachia* in the host's brain to below the native (control) level. These results suggest that *Wolbachia* strain *w*MelPop is recognized by the host organism as a pathogen.

Acknowledgements: This work was supported by RFBR grant № 15-04-08993 and the State project of ICG SB RAS № 0324-2015-0003.

References:

1. Z. Xi et al. (2008) Genome-wide analysis of the interaction between the endosymbiotic bacterium *Wolbachia* and its *Drosophila* host, *BMC Genomics*, 9:1.
2. S. Carra et al. (2010) Identification of the *Drosophila* ortholog of HSPB8: implication of HSPB8 loss of function in protein folding diseases, *J Biol Chem*, 285**:** 37811-37822.
3. D. Voronin et al. (2012) Autophagy regulates *Wolbachia* populations across diverse symbiotic associations, *Proc Natl Acad Sci USA*, 109: E1638-1646.

PREDICTION OF FUNCTIONAL EFFECTS OF REGULATORY SEQUENCE VARIATIONS

M. Malkowska[1]*, J. Zubek[2, 3], D. Plewczynski[3], L.S. Wyrwicz[1]
[1] *Maria Sklodowska-Curie Memorial Cancer Centre and Institute of Oncology, Warsaw, Poland*
[2] *Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland*
[3] *Center of New Technologies, University of Warsaw, Warsaw, Poland*
* *Corresponding author: bioinfo@coi.waw.pl*

Key words: *regulatory variation, function prediction, machine learning*

Motivation and Aim: In the era of huge amounts of data from high-throughput DNA sequencing is necessary to search for more effective methods of analysis of functional regions of genomes, in particular regulatory regions, that may be crucial in the search for single-nucleotide polymorphisms potentially responsible for the development of the diseases.
Our research aims to create and evaluate an integrative machine-learning model for regulatory variants identification within promoter regions of genes. We intend to utilize various gene annotations and DNA shape features, which have recently become available and determine the significance of these new components.
Methods and Algorithms: To improve the accuracy of predicting the functional consequences of regulatory variants we are implementing machine learning algorithms using newly available features. Our positive examples dataset was constructed using regulatory mutations from the Human Gene Mutation Database[1]. Negative examples were derived using single nuclear variations from the 1000 Genomes Project[2]. The annotation data include the local DNA 3D shape, phylogenetic conservation, transcriptomic and epigenetic measurements. The model prediction performance will be evaluated across human genome to determine potential, not reported, true functional variants and compared to existing methods.
Results: Currently, we are in process of model selection and optimization. There are strong indications that chosen features will improve the accuracy and precision of functional effects of regulatory sequence variations predictions.
Conclusion: The results of the project will improve our understanding of the molecular mechanisms of disease pathogenesis and complex traits.
Availability: The machine-learning model and variant predictions will be available through GitHub repository (project in progress).

References:

1. P.D. Stenson at el. (2014) The human gene mutation database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine, *Human Genetics*, 133:1-9.
2. 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes, *Nature*, 491(7422):56-65.

MOLECULAR EVOLUTION AND SYSTEMATICS OF FLAT LEECHES (HIRUDINEA: GLOSSIPHONIIDAE)

N.B. Mandzyak[1]*, I.A. Kaygorodova[1, 2]
[1] *Limnological Institute SB RAS, Irkutsk, Russia*
[2] *Irkutsk State University, Irkutsk, Russia*
* *Corresponding author: mandzyak@lin.irk.ru*

Key words: *species divergence, species delimitation, genetic distances*

Motivation and Aim: Species as a basic unit of evolution have to be identifiable. Meanwhile, the presence of cryptic species poses significant challenges for the accurate assessment of biodiversity. Increasingly, molecular data are being used to evaluate species boundaries and to assist in the identification of groups such as hirudinids in which morphological characters often do not distinguish species reliably [1]. However, the development of species-specific DNA barcodes for leeches has lagged behind that of other groups.
Methods and Algorithms: To study species-specific DNA barcodes, leech samples were collected during multiple expeditions in 2003-2015. Sequences of universal for Metazoan barcoding fragment (*CoxI*) of 73 leech specimens from different sites of Northern Eurasia were newly generated for the present study. Bioinformatic analyses were conducted in MEGA V6.06 [2]. The evolutionary distances were computed separately for both intraspecific and congeneric comparisons using the Kimura two-parameter (K2P) model in accordance with DNA barcode techniques [3].
Results: The molecular phylogenetic analysis of barcoding fragment of the *Glossiphonia* representatives allowed to reveal four evolutionary branches among glossifonias inhabiting Siberian and Far Eastern freshwater bodies. On the resulting phylogenetic tree, each group of nucleotide sequences belonging to *Glossiphonia* sp.1, *Glossiphonia* sp.2, *Glossiphonia* sp.3 and *Glossiphonia* sp.4 was clustered apart of each other and separately from homologous sequences of the rest world fauna representatives. Genetic distances between Siberian leeches and phylogenetically close lineages exceed significantly the threshold of 3% [3] that could indicate their independent taxonomic position and assume that four potentially new species exist. In addition, intraspecific genetic polymorphism varies within the range of 0-0.73% and confirms this supposition. Moreover, due to impressive barcoding gap (over 5%), a long branch attraction phenomenon observed in this group of parasitic organisms [4] becomes clear.
Conclusion: Thus, the use of barcoding thresholds bode well for delineating closely related species and taxonomically understudied groups within Hirudinea.

This work was supported by the Russian Foundation for Basic Research (14-04- 00345).
References:
1. A.E. Bely, D.A. Weiseblat (2006) *Evolution & Development*, 8: 491-501.
2. K. Tamura et al. (2013) *Molecular Biology and Evolution*, 30: 2725-2729.
3. P. Hebert et al. (2003) *Proc. Royal Soc. B: Biological Sciences*, 270:313–321.
4. P. Martin et al. (2000) *Mol. Phyl. Evol.*, 15: 355-368.

PREDICTION OF STRUCTURAL PROPERTIES OF UNCHARACTERIZED PROTEINS FROM THEIR POST-CLEAVAGE MASS SPECTRA BY A MULTIVARIATE STATISTICAL MODEL

O.A. Markelov[1]*, A.R. Kayumov[2], M.I. Bogachev[1]
[1] *St. Petersburg Electrotechnical University, St. Petersburg, Russia*
[2] *Kazan (Volga region) Federal University, Kazan, Russia*
** Corresponding author: OAMarkelov@etu.ru*

Key words: *protein structure prediction, mass spectrometry, multivariate regression model*

Motivation and Aim: Successful prediction of structural and structure-associated properties of uncharacterized proteins currently requires the knowledge of the sequence either of the studied protein itself or of its homologous counterparts. We aimed to characterize proteins directly from their post-cleavage mass distribution shapes without reconstruction of their primary sequences.
Methods and Algorithms: We used Kolmogorov statistics to compare the simulated post-cleavage mass distribution shapes of various proteins from the UniProtKB/Swiss-Prot, RSCB PDB, DisProt, Ideal, PDBTM and VFPB databases. To combine results for several cleavages, we suggest a solution based on the multivariate logistic regression.
Results: We found that various structural and structure-associated properties of proteins are explicitly reflected in the shapes of their post-cleavage mass distributions. In particular, proteins carrying presumably α-helixes in their secondary structure exhibit broad mass distributions, while for proteins with β-sheet structures mass distributions decay by a simple exponential, allowing to distinguish between them with 90% accuracy by using a single Thermolysin cleavage. For membrane-associated proteins with specific structural properties we could properly predict their location (membrane vs soluble) as well as localization in the membrane (monotopic vs transmembrane, single-pass vs multi-pass transmembrane) with 80% accuracy for each pairwise comparison by combining 3-4 different cleavages. Other prominent examples included 3 out of 4 successful predictions of intrinsically disordered proteins vs fixed-structure proteins as well as membrane-associated proteins by their functional group and host bacteria phylum. Remarkably, the best prediction accuracy of structural and localization properties was achieved by Thermolysin and GluC(phosphate)+Lys C cleavage simulations digesting the protein at the positions of amino acid residues from the inertial and external hydropathy groups, respectively [1].
Conclusion: To summarize, we have suggested a multivariate statistical model to successfully predict various structural and structure-driven properties of uncharacterized proteins from their post-cleavage mass spectrometry data.
Availability: Complete data and test examples at http://www.nature.com/articles/srep22286.
Acknowledgements: We acknowledge the financial support by the RSF grant 15-14-00046.
References:
1. Bogachev, M. I. *et al.* (2016) Statistical prediction of protein structural, localization and functional properties by the analysis of its fragment mass distributions after proteolytic cleavage, *Scientific Reports,* 6: 22286; doi: 10.1038/srep22286.

ON THE POSSIBLE IMPACT OF EXOGENOUS 8-OXO-2'-DEOXYGUANOSINE ON DNA SYNTHESIS, DAMAGE AND REPAIR IN AGING CELL CULTURES AND ORGANISM

N.V. Marmiy[1]*, G.V. Morgunova[2], D.S. Esipov[1], A.N. Khokhlov[2]
[1] *Division of Bioorganic Chemistry, Biology Department, Moscow State University, Moscow, Russia*
[2] *Evolutionary Cytogerontology Sector, Biology Department, Moscow State University, Moscow, Russia*
* *Corresponding author: marmiynv@gmail.com*

Key words: *8-oxo-2'-deoxyguanosine, "stationary aging", exogenous 8-oxo-dG, biomarker of aging*

8-Oxo-dG is one of the popular biomarkers of aging and oxidative stress. Its content in DNA increases in aging of humans, animals and cell cultures. The process takes place also in cancer, inflammatory, infectious, neurodegenerative diseases and in some stressful conditions related to smoking, diet modifications, athletic exercises, etc. According to a commonly held view, 8-oxo-dG is just a "waste product" removed from the damaged DNA. However, in the last years some data appeared which proved a possible role of 8-oxo-dG in the regulation of biological processes. In particular, it is capable of exerting anti-inflammatory and antiallergic effect, increasing survival of organisms by irradiation, hypoxia, and starvation. The objective of our research was to identify the possible biological effects of 8-oxo-dG in cell culture, in which they can not be due to immune and neurohumoral mechanisms, and are solely the result of direct exposure to a substance on a cellular level.

The experiments used culture of transformed Chinese hamster cells line B11-dii FAF28. To measure the content of 8-oxo-dG in DNA method of RP HPLC with EC detectionwas used.

During the experiments we have shown that the ratio of 8-oxo-dG / dG in DNA is increased by more than 6 times in the transition from log phase of cell culture growth to stationary thus 8-oxo-dG be one of the biomarkers "stationary aging" culture. By adding exogenous 8-oxo-dG in the culture medium it is intensively absorbed by the cells. Cytotoxicity of 8-oxo-dG, when added to cells in the logarithmic growth phase was not observed in the concentration range 10^{-6}-10^{-3} M. At the same time, exogenous 8-oxo-dG in the middle and high concentrations resulted in a significant and reproducible reduction in the content of 8-oxo-dG in DNA of cells of the stationary phase of growth. The ratio of 8-oxo-dG / dG falling under its influence in 5-6 times in comparison with control, approaching the level typical for the cells in logarithmic growth phase. At the same time, the addition of exogenous 8-oxo-dG in the early stages of cultivation (growth - the beginning stationary phase) had no effect on its content in the DNA.

Thus, the nature of exposure 8-oxo-dG cells strongly depends on its concentration in the medium and proliferative status of the culture. The experimental data clearly indicate the presence in 8-oxo-dG expressed biological effect. We assume that it can be associated with the activation of reparative and antioxidant systems without oxidative stress.

MOLECULAR EVOLUTION ANALYSIS OF RNA-BINDING NIP7 PROTEIN FROM DEEP- AND SHALLOW-WATER ARCHAEA

K.E. Medvedev*, D.A. Afonnikov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: albatros@bionet.nsc.ru*

Key words: *Nip7 protein, SDP, extremophiles, high pressure, archaea, adaptation*

Motivation and Aim: Pressure and temperature are important environmental factors that determine many of the processes occurring in living organisms. The greatest interest is caused by organisms-extremophiles, which inhabited the ecosystems where conditions are not compatible with the lives of most of other organisms. Mechanisms to ensure survival of cells under such conditions are still not clear. Their understanding will help to answer some fundamental questions related to the origin of life and evolution of microorganisms in its earlier stages and adaptation to conditions of different ecological systems. One of the effective approaches to study the adaptation of protein structures to extreme conditions is the comparative analysis of sequences. In current work we conduct a study of the molecular evolution of protein Nip7, which contributes to adaptation to life at high pressure and temperature.
Methods and Algorithms: We used two programs to identify specific substitutions: multi-Harmony [1] and Zebra [2].
Results: First of all should be considered the results of the identification of specific protein positions in relation to the depths of habitats of organisms. The number of significant positions that is common to both programs amounted to 19. In the case of the specificity to the temperature, the number of significant specific positions for multi-Harmony was 72, for Zebra – 64 positions. The number of detected specific position in this case is much larger than in the analysis of specificity to pressure. The intersection of the results of two programs is also larger for temperature.
Conclusion: It should be noted that the number of positions in which specific substitutions are associated with temperature more than three times exceeds the number of positions specific to pressure. Furthermore, their significance level (Z-statistics) for a significant part of these positions higher than the positions selected for specificity to the depths of habitat. The data obtained may reflect the fact that molecular adaptation to high temperatures in the Nip7 protein is more pronounced than the pressure, i.e. temperature is a factor in the selection to a greater extent than the pressure. Thus, the analysis showed that on the one hand, for positions in which substitutions can be specific way in relation to the depth habitats of the organisms for deep-sea organisms are more typical substitutions that increase the hydrophobicity of the residues from shallow-water organisms compared to deep-water.
References:

1. W. Pirovano, K.A. Feenstra, J. Heringa. (2006) Sequence comparison by sequence harmony identifies subtype-specific functional sites. *Nucl. Acids Res.*, 34:6540-6548.
2. D. Suplatov, E. Kirilin, V. Takhaveev, V. Švedas. (2014) Zebra: a web server for bioinformatic analysis of diverse protein families. *JBSD*, 32:1752-1758.

HIGH TEMPERATURE AND PRESSURE INFLUENCE ON INTERDOMAIN INTERFACE OF THE NIP7 PROTEINS FROM *P. ABYSSI* AND *P. FURIOSUS*: MD SIMULATION RESEARCH

K.E. Medvedev*, D.A. Afonnikov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: albatros@bionet.nsc.ru*

Key words: *Interdomain interface, domain motions, high temperature, adaptation, Nip7*

Motivation and Aim: Interaction between domains of the protein, as well as their relative positions and movement relative to each other is essential for the stability of protein structure and normal functioning of a protein molecule. The feature of the movement of the domains may define the mechanism of enzymatic reactions. Therefore, the description of this motion is an important task in the analysis of the structures and functions of multidomain proteins. In current work we provide the investigation of the influence of high pressure and temperature influence on change of two domains motion parameters of the Nip7 protein from deep-water (*P. abyssi*) and shallow-water (*P. furiosus*) archaea.
Methods and Algorithms: We used DynDom [1] to analyze the changes in the parameters of the mutual orientation of the domains of protein Nip7 during MD simulation.
Results: Obtained data showed that interdomain interfaces of *P. abyssi* and *P. furiosus* Nip7 proteins were formed by stable hydrophobic interactions. It is shown that increasing the pressure significantly influences the angle of rotation of the domains and increasing the temperature slightly reduces the value of the angle of rotation of the domains. The effect of temperature on translation along the axis has a different pattern for the two proteins. The analysis of the quantiles of the distribution showed that the increase of temperature shifts the distribution of the model parameter *P. abyssi* Nip7 in the direction of increasing, for the *P. furiosus* Nip7 in the direction of decreasing.
Conclusion: In current work we propose an approach that allows us to analyze the motion parameters of the protein domains during MD simulation. This approach implies a certain approximation that domains constitute a rigid structural subunit of protein structure. This allows you to concentrate on the study of changes occurring in the interdomain space which have a significant impact on the entire structure. Analysis of the direction of movement of the domains showed that the domains of deep-water and shallow-water organisms Nip7 protein move differently. In addition, it is suggested that the type of motion of domains under study is similar to the “shear motions”.

References:

1. S. Hayward R.A. Lee. (2002). Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J. Molecular Graphics Modelling*, 21(3):181-183.

SITEX 2.0: FUNCTIONAL SITES PROJECTION ON ALTERNATIVE SPLICED ISOFORMS AND HOMOLOGOUS GENES

I.V. Medvedeva*, P.S. Demenkov, V.A. Ivanisenko
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author:brukaro@bionet.nsc.ru*

Key words: *gene structure, protein functional sites, alternative splicing*

Motivation and Aim: The SitEx system stores information on projection of exon and domain boundaries, positions of functional sites on protein sequence and the coding sequence of the gene [1]. The functions of proteins and their domains are defined mostly by functional sites so they are highly conserved units of proteins. Protein functional sites variability is low that let us to project known protein functional site amino acids positions from one isoform to another, as well as to project them to orthologous and paralogous gene sequences. This information is applicable in the study of the structural–functional organization of the gene in evolutionary perspective and could help in design of novel proteins.
Methods and Algorithms: This work presents update of SitEx system. The information about isoforms and homologous genes was extracted from Ensembl 82 release using public MySQL database. For protein functional site projection on other isoforms was used exon-to-exon blast alignment. To indicate function site amino acid in homologous genes we applied Clustal Omega with further checking for functional site amino acid position sequence environment. Previous SitEx version contained the information about PDB entries with 40% sequence similarity. We included every possible PDB entry in current release. Found homologous sequences are presented using alignments. We also integrated information about SNP from Ensembl and 1000 genomes projects, annotated sites from Catalytic Site Atlas.
Results: Currently PDB contains about 120 000 structures. We obtained only 44 000 structures that have known protein functional sites from Eukaryota. These structures were connected to 6637 genes and 13034 transcripts from 41 species including animals, plants and fungi according to previously published pipeline [1]. There were discovered protein functional sites changes as the consequences of alternative splicing for only 5 genes.
Conclusion: The developed system allows analyzing protein functional site in alternative isoforms using 3D structures because the affinity of the sites could be changed. SitEx 2.0 also opens possibility to evaluate protein functional site preservation in homologous sequences.
Availability: http://www-bionet.sscc.ru/sitex/, MySQL dump of 2.0 release
References:

1. Medvedeva I.V., Demenkov P.S., Kolchanov N.A., Ivanisenko V.A. (2012). SitEx: a computer system for analysis of projections of protein functional sites on eukaryotic genes. Nucleic Acids Res. Database issue:D278-83.
2. Medvedeva I.V., Demenkov P.S., Ivanisenko V.A. (2015) Computer analysis of protein functional sites projection on exon structure of genes in Metazoa. BMC Genomics. 16(Suppl 13):S2

PROGRAM COMPLEX ICGENOMICS FOR ANALYSIS OF HIGH-THROUGHPUT SEQUENCING EXPERIMENTS

I.V. Medvedeva[1], A.O. Bragin[1], K.V. Gunbin[1], P.S. Demenkov[1], O.V. Vishnevsky[1], A.M. Spitsina[2], F.M. Naumenko[2], V.N. Babenko[1], N.L. Podkolodnyy[1], Y.L. Orlov[1]*
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: orlov@bionet.nsc.ru*

Key words: *computer genomics, sequencing, program complex, data integration*

Motivation and Aim: The program complex ICGenomics has been designed for storage, mining, and analysis of high-throughput sequencing experiments [1]. ICGenomics enables wet-lab biologists to perform high-quality processing of data in the fields of genomics, biomedicine, and biotechnology. They include novel methods of the processing of initial high-throughput sequencing data. Examples are: ChIP-seq analysis; functional annotation of gene regulatory regions in nucleotide sequences; prediction of nucleosome positioning; and structural and functional annotation of proteins, including prediction of their allergenicity parameters, as well as estimates of evolution changes in protein families. Applications of ICGenomics to the analysis of genomic sequences of the yeast, ChIP-seq data on the mouse and human are considered.
Methods and Algorithms: We developed set of computer programs and have integrated them in program complex. ICGenomics implements both standard and modern methods for processing, analyzing, and visualizing sequencing data and functional annotation of genome regions.
Results: The program complex ICGenomics allows to fulfill the following distinct functions: (1) processing of extended nucleotide sequences from next generation sequencing data including; (2) annotation of genomic sequences including exon search, and prediction of miRNA gene promoters using specific nucleotide structure motifs; (3) prediction protein allergenicity by their structural and functional properties using functional annotation of protein spatial structure; (4) research of evolution modes of protein coding genes, including reconstruction of evolutionary history of proteins on the basis of ortholog prediction in sequenced genomes; the phylogenetic analysis and investigation of selection modes.
Conclusion: Evolutionary history of proteins reconstruction is based on ortholog prediction in sequenced genomes. The component is realized in the form of the data processing pipe-line. New development of the system includes ChIP-seq analysis software [2].
Availability: The system is available at http://www-bionet.sscc.ru/icgenomics.

References:

1. YL. Orlov et al. (2012) ICGenomics: program complex for symbol sequence analysis in genomics, *Vavilov journal of genetics and breedi*ng, 16(4/1): 732-741. (In Russian).
2. A.M. Spitsina et al. (2015) Supercomputer analysis of genomics and transcriptomics data revealed by high-throughput DNA sequencing, Program systems: theory and applications, 6:1(24): 157–174. (In Russian).

NONTHERMAL IMPACT TERAHERTZ RADIATION ON THE LIVING SYSTEMS

I.A. Mescheryakova[1], E.V. Demidova[1], T.N. Goryachkovskaya[1], E.A. Demidov[1], A.V. Bryanskaya[1], S.V. Sergeeva[1], S.L. Kiselev[3], M.A. Lagarkova[3], G.N. Kulipanov[2], A.I. Semenov[2], N.A. Vinokurov[2], N.A. Kolchanov[1], V.M. Popik[2], S.E. Peltek[1]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Budker Institute of Nuclear Physics SB RAS, Novosibirsk, Russia*

[3] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

Key words: *evolution, genome, k-mer distribution*

Terahertz (THz) radiation was proposed recently for use in various applications, including medical imaging and security scanners.

We studied the impact of terahertz radiation on E. coli biosensor cells containing plasmids with promoters of stress-sensitive genes controlling the expression of GFP. GFP level was measured by fluorometry. The impact of terahertz radiation was nonthermal, i.e. special care was taken to keep specimen temperature at the 35±2 °C range during irradiation so that heat shock genes are not induced. We have found that terahertz radiation activates genes associated with oxidative stress response. Results of the Ames test and SOS-chromotest indicate that terahertz radiation produces neither mutagenic nor genotoxic effects.

The exposure of E. coli cells under terahertz radiation causes increased expression of 14 genes of rapid response. Among these genes was glutamine synthetase gene (glnA). By using the glnA gene promoter we have designed biosensor sensitive to the effects of the terahertz radiation.

Since human embryonic stem cells (hESCs) are extremely sensitive to environmental stimuli, we have therefore utilised this cell model to investigate the non-thermal effects of THz irradiation. We have studied DNA damage and transcriptome responses in hESCs exposed to the narrow-band THz radiation (2.3 THz) under strict temperature control.

This work was supported by a grant from the Ministry of Education and Science of the Russian Federation (agreement № 14.616.21.0053 (RFMEFI61615X0053)).

BIOMARKERS OF AGE IN THE "STATIONARY PHASE AGING" MODEL

G.V. Morgunova*, D.S. Esipov, M.V. Marmiy, A.N. Khokhlov
Evolutionary Cytogerontology Sector, School of Biology, Lomonosov Moscow State University, Moscow, Russia
** Corresponding author: morgunova@mail.bio.msu.ru*

Key words: *biomarkers of aging, senescence-associated β-galactosidase, cytogerontology, stationary phase aging, cell senescence, 8-oxo-2'-deoxyguanosine*

Motivation and Aim: Currently gerontologists search for biomarkers of aging (including cell aging) which can allow determining age of organism/cell quickly and easily. Senescence-associated beta-galactosidase (SA-β-Gal) remains the most popular biomarker of cell aging. Nowadays, another biomarker, 8-oxo-2'-deoxyguanosine (8-oxo-dG), is becoming popular in gerontology. We have investigated applicability of these markers to our "stationary phase aging" model, i.e. increase in the probability of dying for cultured cells upon retardation and subsequent complete cessation of their proliferation within one passage. *Methods and Algorithms:* Experiments were performed on transformed Chinese hamster cells (B11-dii FAF28 line, clone 237). The cells were cultivated for 14-15 days at 37°C in Carrel glass flasks using DMEM supplemented with 10% bovine serum and antibiotics. In the first series of experiments on the 4th, 8th, and 15th day contents of 8-oxo-dG and dG in DNA hydrolyzate were analyzed chromatographically using Beckman-Gold chromatograph (USA) at a wavelength of 254 nm. In the second series of experiments on the 7th and 14th day the cells were fixed for 3-5 minutes in 2% formaldehyde and 0.2% glutaraldehyde and incubated with X-Gal for 12-16 hours at 37°C. *Results:* It was found that the ratio of 8-oxo-dG/dG increased with the "age" of cell culture. On the 15th day the cells became to die and the ratio had significantly increased ($22.40 \cdot 10^{-5}$) compared to this index on the 4th ($6.26 \cdot 10^{-5}$) and on the 8th ($4.42 \cdot 10^{-5}$) days when the cells had reached monolayer and gone into the stationary phase of growth. It was also found that 14-day-old culture had much higher percentage of cells staining for SA-β-Gal than the «young» (7-day-old) cells. *Conclusion:* Thus, 8-oxo-dG accumulates in the stationary phase aging culture of Chinese hamster cells as evidenced by a significant increase in the ratio of 8-oxo-dG/dG in DNA of the cells on the 15th day. Consequently, it is possible to predict an increase in the probability of death in the cell culture evaluating expression of this biomarker. Furthermore, stationary phase aged cells express SA-β-Gal demonstrating a good correlation of this parameter with "age" of cell culture. We believe that both methods can be used to determine the biological age of cells in testing of new potential geroprotectors.

k-MER FREQUENCY DISTRIBUTION OF EUKARYOTIC PROTEOMES

A.A. Morozov[1]*
Limnological Institute SB RAS, Irkutsk, Russia
** Corresponding author: morozov@lin.irk.ru*

Motivation & aims: There is a body of work regarding k-mer frequency analysis in DNA sequences. It was shown earlier that such distribution is species-specific and can be used for short sequence classification *eg* in metagenomic projects. In this work I show that the same is true for aminoacid sequences, and attempt to assess some of the factors that could influence the specificity of k-mer distribution.
Methods & algorithms: Two datasets were used: the CEGMA collection of highly conservative homologous housekeeping genes of six model eukaryotes (*A. thaliana, C. elegans, D. melanogaster, H. sapiens, S. cerevisiae* and *S. pombe*), and complete proteomes of the same six species. UNIPROT annotations for structural and functional elements of proteins were used.
The sequences were classified using naïve Bayes classifier with k-mer frequencies as features. 90% of sequences in each dataset were randomly selected for training the classifier, and the remaining 10% were classified. Calculations were performed in pure Python. For comparing the distributions, euclidean distances were calculated using only the frequencies of *k*-mers present at least once in both distributions.
Results: The specificity of the analysis varied for the different genomes and values of *k*, but typically it was from 30% to 80%. For most of the genomes specificity peaked at *k* values of 5 or 6, unlike optimal *k* for DNA-based analyses which was shown to be between 10 and 15 nucleotides in various earlier works.
The next question is whether factors that shape the k-mer distribution specificity, whatever they may be, apply to the entire protein or are restricted only to some of its parts. If there indeed are such parts, their *k*-mer distributions will be more different from each other than those of sequences as a whole. If, on the opposite, some subsequences were under lesser influence of *k*-mer shaping factors, their distributions in different proteomes will be more similar. To test this hypothesis, I have build proteome-specific distributions for a series of features. Distances were calculated between these distributions and the universal distributions for the same features, built using all the species in the datasets. Distribution of distances is approximately the same for both complete protein sequences and structural features (helices, beta-strands, transmembrane domains). It allows to assume that these structural features have no specificity in terms of *k*-mer composition. The same is true for entire annotated domains and non-domain subsequences, which means that *k*-mer specificity is not, in general, affected by sequences' functional status.
Conclusion: k-mer distributions were shown to be proteome-specific. No hypotheses can be made regarding the reasons of this phenomenon, but it was shown that it applies to a significant portion of proteins within proteome and to the entire protein sequences.

CAN LONG ANTIPARALLEL OPEN READING FRAMES BE ENCODING ESSENTIAL GENES IN PROKARYOTIC GENOMES?

D.M. Moshensky*, A.V. Alexeevski
A.N. Belozersky Institute of Physico-Chemical Biology MSU, Moscow, Russia
** Corresponding author: moshenskydenis@gmail.com*

Key words: *overlapping genes, antiparallel open reading frame*

Motivation and Aim: The origin and evolution of genes that have common base pairs (overlapping genes) is a topic of several dozen modern researches. Co-directional overlaps longer than 60 bp seem to be erroneous except for a couple of cases [1]. So we decided to analyze antiparallel overlapping genes: whether open reading frames (ORFs) located opposite to genes (antiparallel ORFs) can be genes.
Methods and Algorithms: To analyze 968 reference prokaryotic genomes we have developed a program which finds ORFs on the complement strand of the annotated genes and calculates their P-value. Ka/Ks calculator 2.0 [2] has been used to estimate Ka/Ks ratio following the Yang and Nielsen method [3].
Results: First, we have found that long antiparallel ORFs are observed reliably more frequent than expected. There are 10472000 antiparallel ORFs in 968 analyzed genomes with overlap length more than 180 bp. Stop-codons on the opposite to the coding strand are avoided in 2898 cases with FDR P-value 0.01.
Second, we have discovered that antiparallel ORFs are subjected to positive selection. Demonstrative example is more than 1800 bp antiparallel ORF found opposite to extremely conserved *dnaK* genes. Translations of these ORFs were claimed 'glutamate dehydrogenases' (GDH) and even included into Pfam DB as third protein family PF10712 of GDH [4]. Ka/Ks analysis demonstrated that if these translations correspond to proteins then they are under positive selection while *dnaK* genes are under strong negative selection. Due to zero shift of the ORFs' codons each synonymous mutation in *dnaK* gene leads to amino acid substitution in the ORF translation. However, some antiparallel ORFs, in particular *dnaK* related, have been found to resist to synonymous changes in genes. It indicates possibility of their translation. We speculate that their occasional translations should avoid toxicity for a bacterial cell.
Conclusion: Essential genes are unlikely to be encoded by antiparallel ORFs in prokaryotic genomes. Nevertheless, some antiparallel ORFs might have biological significance associated with their translations.
Acknowledgements. The work was supported by RFBR grant 14-04-01693 (1st author) and RNF grant 14-50-00029 (2nd author)
References:
1. A. Pallejà, E. Harrington, P. Bork. (2008) Large gene overlaps in prokaryotic genomes: result of functional constraints or mispredictions? *BMC Genomics*, 9:335.
2. D. Wang et al. (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies, *Genomics Proteomics Bioinformatics*, 8:77-80.
3. Z. Yang, R. Nielsen. (2000) Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models, *Mol Biol Evol*, 17:32-43.
4. N. Walker, N. McEwan, R. Wallace. (1999) Overlapping Sequences with High Homology to Functional Proteins Coexist on Complementary Strands of DNA in the Rumen Bacterium *Prevotella albensis*, *Biochemical and Biophysical Research Communications*, 263:58-62.

GEROPROTECTOR AND CRITERIA FOR ITS EVALUATION

A. Moskalev[1-4]*, M. Shaposhnikov[1, 2], E. Proshkina[1, 2], V. Tsvetkov[4], A. Fedintsev[4], E. Chernyagina[4], A. Zhavironkov[4]

[1] *Institute of Biology of Komi Science Center of Ural Branch of RAS, Syktyvkar, Russia*

[2] *Syktyvkar State University, Syktyvkar, Russia*

[3] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*

[4] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

** Corresponding author: kateplus@mail.ru*

Key words: *aging, health, geroprotector*

To date, more than 200 substances that prolong the life of model organisms have been reported in the literature (http://geroprotectors.org). Reducing the cost and improving the efficiency with which increasingly large amounts of data from model organisms can be applied to humans will be critical to progress in the development of human geroprotectors. For this purpose, we have to come to an agreement what should be considered applicable to human geroprotectors. Primary selection criteria for potential geroprotector can be as following: 1) Increased lifespan in models or human. The increase in lifespan is not always accompanied by positive changes in the quality of life, and additional criteria for geroprotectors is needed, and discussed below. 2) Amelioration of human aging biomarkers (http://ageing-map.org). 3) Acceptable toxicity. 4) Minimal side effects. 5) Improving health-related quality of life. Secondary selection criteria are needed to reduce cost of investigations and increase efficacy for potential geroprotector: 6) Evolutionary conservatism of target or mechanism of action (http://agingchart.org). 7) Reproducibility of geroprotective effects on different model organisms. 8) Simultaneous influence on several aging-associated causes of death in mammals. 9) Increase of stress resistance.

Analysis of published data with the use of developed criteria did reveal candidates that fit all of the main criteria: (e.g., acarbose, deprenyl, d-glucosamine, dihydroergocristine methanesulfonate, ellagic acid, fenofibrate, glutathione, metformin, spermidine, tyrosol, and vinpocetine, telmisartan), and we suggest that they are tractable candidates for human interventions.

PHOSPHORYLATION OF AB-CRYSTALLIN: EFFECTS OF AGING AND CARDIOMYOPATHY

N.A. Muraleva[1]*, V.A. Devyatkin[1, 2], N.A. Kolosova[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: Myraleva@bionet.nsc.ru*

Key words: *αB-crystallin, phosphorylation, cardiomyopathy, OXYS rats*

Motivation and Aim: αB-crystallin (Cryab) is ubiquitous and has critical roles in several cellular processes that are modulated by phosphorylation (pCryab). Aging and age-related diseases lead to increased pCryab in several tissues such as eye lens, retina and skeletal muscle. Earlier we showed connection of cataract and retinopathy development in accelerated senescence OXYS rats with alteration of *Cryab* gene expression. Here we evaluated the changes of Cryab and its phosphorylation at serine positions 59 (pS59-Cryab) in myocardium in response to normal aging and development of cardiomyopathy in OXYS rats to estimate the contribution of these changes in the systemic manifestations of accelerated aging.
Methods: Electrocardiographic analysis, testing of blood pressure, western blot analysis, real-time PCR, histological and immunofluorescent assay. *Results:* We found that the clinical signs of cardiomyopathy manifested themselves in electrocardiograms of OXYS rats at the age of 12 months. Nevertheless, a slight functional alteration (tachycardia) was observed at 20-day-old rats and increased by the age of 3 months when OXYS rats developed arterial hypertension. In 12-month-old OXYS rats, a histological assay revealed hypertrophy of cardiomyocytes; substantial fibrosis of connective tissue was detected among the cardiomyocytes and around the coronary vessels; the vessel index (ratio of outer/inner diameter) was increased in comparison with Wistar rats (controls). The signs of atherosclerosis in the vessels OXYS rats were not detected. The development of cardiomyopathy was not associated with alterations in expression of Cryab (determined by Western blot) in OXYS rats, but the expression of its protein increased by the 24 months. Phosphorylation of Cryab in OXYS and Wistar rats increased by the age of 3 months, as a result there was a decrease level of protein Cryab. The different on pS59-Cryab expression from detergent-soluble fraction between OXYS and Wistar rats was not found. Moreover, the level of pS59-Cryab in detergent-insoluble fraction was increased in 3-month-old OXYS rats compared with the control animals. This was the result of translocation of the pS59-Cryab from detergent-soluble to detergent-insoluble fraction.
Conclusion: Our study confirmed that the development of secondary cardiomyopathy in 12-month-old OXYS rats associated with hypertension. The expression of Cryab increased in myocardium OXYS and Wistar rats with the aging and it had not links with development of cardiomyopathy. The accumulation pS59-Cryab in detergent-insoluble protein fraction of myocardium elevated translocated to striated sarcomeres with the aging in both strain. Moreover, it was faster in myocardium OXYS rats and preceded the development of cardiomyopathy. *This study was supported by the RFBR (Grant # 14-04-00376).*

ADVANCED CAPABILITIES OF VISUALIZATION AND ANALYSIS OF CULTURAL MODELS

E.R. Muslikhov, L.A. Strukova
Qvadros-Bio, LLC, Moscow, Russia
** Corresponding author: erm@qvadrosbio.ru*

Key words: *cell culture, experimental technology*

Cell culture - one of the most commonly used models in biology, which is applicable for the morphological, physiological, genetic, molecular biological, pharmacological and other studies. At the same time cell is a very complicated dynamic structure with a variety of biological and chemical processes occurring inside it. Many factors, both external and internal, at the same time can affect its state, which can very can be important for the results of the study. Through the development of photometry and microscopy, more and more attention is focusing on dynamic processes in the cell, but thanks to advances in technology, it is possible to make a continuous shooting in unacceptable for conventional microscopes and cameras conditions at high temperature and humidity directly inside the CO_2 incubator. Thus, it is possible to study the behavior and track the status of cells without having to create a complex of climatic chambers, which are much less effective than incubators.

"Quadros-Bio" company is represent IncuCyte Zoom automated microscopes and a number of equipment for shooting inside the incubator in Russia. The equipment is compatible with most types of cultural plastic, automatically focuses and takes pictures at specified intervals, processes the information and produces graphical results. It is now possible to observe the cells during hours, days, weeks and even months in real time, from anywhere in the world where you have a network connection.

In our report, we will review how based on continuous recording in real time measurement helping to get a true idea of the influence of various factors on the morphology, proliferation and cell health; on the nature of the interaction between the different cell types; on angiogenesis mechanisms of neurogenesis, and neurodegenerative disorders. This survey also allows to perform phenotypic characterization of genes associated with various events in the life of the cell, to assess the overall condition of the culture and the effectiveness of pharmacological agents.

References:

1. I. V. Arutyunyan, T. H. Fatkhudinov, A.V. El'chaninov, A.V. Makarov, E.Yu. Kananykhina, N.Yu. Usman, E.Sh. Raimova, D.V. Goldshtein, G.B. Bol'shakova. Effect of Endothelial Cells on Angiogenic Properties of Multipotent Stromal Cells from the Umbilical Cord during Angiogenesis Modeling in the Basement Membrane Matrix Cell Technologies in Biology and Medicine, (2016) 4: 575-583
2. V. Vakhrushev, A. S. Vdovin, L. A. Strukova*, K. N. Yarygin. Variability of the Phenotype and Proliferation and Migration Characteristics of Human Mesenchymal Stromal Cells Derived from the Deciduous Teeth Pulp of Different Donors. Cell Technologies in Biology and Medicine, (2016) 4: 525-529

ORTHOSCAPE: A CYTOSCAPE PLUGIN FOR EVOLUTIONARY ANALYSIS OF GENE NETWORKS

Z.S. Mustafin[1]*, D.A. Afonnikov[1, 2], K.V. Gunbin[1], Yu.G. Matushkin[1, 2], S.A. Lashin[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: mustafinzs@bionet.nsc.ru*

Key words: *Cytoscape plugin, ortholog, paralog, metabolic pathway, gene regulatory network, evolution, phylostratigraphy, evolution*

Motivation and Aim: There are a number of software intended to visualization and analysis of biological networks. Among them, Cytoscape (http://cytoscape.org/) is one of the most comprehensive tools. There are a lot of plugins extending the base functionality of Cytoscape. Nevertheless, there is still lack of evolution-oriented plugins: just three plugins tagged 'network evolution' in Cytoscape official app store and in literature. We have developed a new Cytoscape plugin Orthoscape aimed to perform evolutionary analysis of gene networks and visualize its results. *Methods and Algorithms:* We used KEGG (http://www.kegg.jp/) Pathway database to get gene networks, KEGG Orthology to get lists or homologs with identity and SW Score values, KEGG Genes to get protein domains, also nucleotide and amino acid sequences and KEGG Organism to get taxonomic information. The domain composition and sequence similarity approaches used to discriminate between paralogs and orthologs [1]. Last common taxonomic level overlapped with inferred orthology status were used as a phylostratigraphic data [2] about (sub)network origin/divergence. We also allow users to create new networks from gene sets using GeneMANIA or CyPath2 (for BioPax format support) plugins. In this case, we used KEGG sequence similarity search to associate proteins from new networks with data annotated in KEGG. Divergence index is calculated by KaKs calculator [3] using pairwise sequence comparisons for the taxa under analysis. The average value of this index allow us to discriminate between diversifying and stabilizing selection in orthologous groups. *Results:* Cytoscape plugin Orthoscape has been developed. The plugin allows users to analyse evolutionary information in the gene sets and networks: (1) the orthology relationships between genes; (2) the evolutionary origin of gene network components; (3) the evolutionary regime (diversifying or stabilizing, negative or positive selection) of orthologous groups in general and/or branch-oriented mode. The distinctive feature of Orthoscape is the ability to control all data analysis steps via user-friendly interface.

Conclusion: Orthoscape allows user to analyze gene networks or separated gene sets to know the evolutionary origin of genes (sub)networks and selective pressure. It also provides convenient visualization and data manipulation abilities on each data analysis step. *Availability:* Upon the requests to the authors. *Acknowledgements:* The study is supported by the RSF 14-24-00123 grant.

References:

1. T. Gabaldón, E.V. Koonin (2013) Functional and evolutionary implications of gene orthology. Nat Rev Genet. 14:360-6.
2. M. Quint et al. (2012) A transcriptomic hourglass in plant embryogenesis, Nature, 490: 98-101.
3. Z. Zhang et al. (2006) KaKs_Calculator: Calculating Ka and Ks through model selection and model averaging, Genomics, Proteomics & Bioinformatics, 4: 259-263.

CROSSING VALLEYS AND REACHING PEAK ON THE FITNESS LANDSCAPES IN MICROBIAL COMMUNITIES UNDER VARIOUS ECOLOGICAL CONDITIONS: A SIMULATION STUDY

Z.S. Mustafin[1]*, D.A. Afonnikov[1, 2], Yu.G. Matushkin[1, 2], S.A. Lashin[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *Corresponding author: lashin@bionet.nsc.ru*

Key words: *Fitness landscape, fitness valley, microbial community, modeling, evolution, mutation rate*

Motivation and Aim: Expression of the majority of an organism's phenotypic traits depends on the state of alleles in several genome loci. To model the evolution of a population with such complex genotype-phenotype relationship, the locus-locus interaction should be taken into account. In these models, the evolution of a population is usually described by its movement over the fitness landscapes [1]. The challenging problem for such models concerns the necessity for a population to overcome intermediate evolutionary stages by random drift. These stages are associated with the fixation of harmful mutations. Nowadays, the interest is also induced by the possibility of models to explain (or at least to try to explain) huge amounts of data recently collected as a result of high-throughput sequencing of genomes [2, 3] as well as by the possibility to check classical and novel theoretical models and to estimate their parameters. These evolutionary models are aimed to answer the following major question: what factors do allow populations to cross valleys with decreased fitness on the fitness landscape during their transition from one peak to another? *Methods and Algorithms:* We analyze mechanisms of crossing fitness valleys for a population of haploid microorganisms, fitness of which depends on allelic states in two loci and determined by complex landscape which shape could be described as "mount in the field surrounded by trench". For this purpose, we used the HEC software [4] to build and simulate necessary models. We considered the impact of various biological factors on evolutionary fate of microbial communities. The factors included both molecular-genetic ones (mutation rate, affinities of subunits), population ones (fitness function landscape), and ecological ones (concentration of accessible substrate in the habitat, flow rate). *Results and discussion:* Our results demonstrate that the success in passing of the fitness valleys in our model is defined by both fitness difference for different allelic combinations and mutation rate. Either gradual or saltation evolutionary modes may be optimal for different fitness landscapes. *Availability:* Software – http://evol-constructor.bionet.nsc.ru, model scripts – upon the requests to the authors.
Acknowledgements: The study is supported by the RSF 14-24-00123 grant.
References:

1. M.C. Whitlock et al. (1995) Multiple Fitness Peaks and Epistasis. Annu Rev Ecol Syst, 26:601–629.
2. T.F.C. Mackay. (2013) Epistasis and quantitative traits: using model organisms to study gene–gene interactions. Nat Rev Genet, 15:22-33.
3. M.V. Meer et al. (2010) Compensatory evolution in mitochondrial tRNAs navigates valleys of low fitness. Nature, 464:279-282.
4. S.A. Lashin et al. (2014) HEC 2.0: improved simulation of the evolution of prokaryotic communities, Mathematical biology & bioinformatics, 9: 585-596.

POLYMORPHISM OF THE *VRN-A1* EXON-4 AND EXON-7 IN POLYPLOID WHEAT

A.F. Muterko*, E.A. Salina
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: muterko@gmail.com*

Key words: *wheat, VRN-A1, exon-4, exon-7, polymorphism*

The manipulation by the quantitative values of such agronomically valuable traits as vernalization requirement, frost tolerance and flowering time allows adapt of plants to a wide range of environmental conditions that is an especially actual in wheat breeding in condition of changing climate.

In recent studies the association of the missense mutations within *VERNALIZATION-A1* exon-4 and exon-7 with modulation of frost tolerance, vernalization requirement duration and flowering time of wheat was shown. However these investigations were carried out exclusively in varieties of *T. aestivum* and has not covered of other species of polyploid wheat and the different *VRN-A1* alleles. Furthermore, often, the *VRN-A1* exon-4 and exon-7 analyzed separately, excluding influence of missense mutations in each of them on the results. In finally, earlier studies not supposed more than one copy of *VRN-A1* per genome and, hence, not estimated ratio of the *VRN-A1* copies with alternative haplotypes of exon-4 and exon-7.

In present study polymorphism of the *VRN-A1* exon-4 and exon-7 was investigated in accessions of 6 tetraploid and 5 hexaploid wheat species carrying the different *VRN-A1* alleles. Furthermore, in contrast to previous studies, investigation was performed given the presence of *VRN-D4*, which only by the several SNPs differs from *VRN-A1*. Specially developed approach, based on anomalous migration of the curved DNA molecules through polyacrylamide gel and the quantitative fluorescence image analysis have been implemented for discrimination and definition of ratio of the *VRN-A1* copy with the exon-4 alternative haplotype. The allele-specific multiplex PCR assay was designed to identification of the *VRN-A1* exon-7 haplotype. Combination of the exon 4 and 7 haplotypes was defined for each copy of *VRN-A1*.

Polymorphism of the *VRN-A1* exon-4 and exon-7 was revealed only in accessions of hexaploid wheat, while tetraploid wheat characterized by the intact sequences (wild type). Furthermore, without two variants, identified only in hexaploid wheat, all dominant *VRN-A1* alleles carry intact exons 4 and 7. Exception for one accession the mutant *VRN-A1* exon-4 was detected only with intact variant simultaneously. This observation allows assume that mutant type and overall polymorphism of exon-4 is associated with no less than two copy of *VRN-A1* per genome. Analysis of the *VRN-A1* exon-4 and exon-7 alternative haplotype combinations in hexaploid wheat, found that wild type of exon-7 and mutant type of exon-4 are associated with analogous haplotype of exon 4 and 7 respectively. In finally, 3 different ratios of alternative *VRN-A1* exon-4 haplotypes per genome were identified. Change of ratio of the alternative *VRN-A1* exon-4 and exon-7 haplotypes can be used to modulate of quantitative values of traits that are important for adaptation of the bread wheat varieties to a large range of environments.

PROTEOMIC OF TCA - EXTRACTED COMPOUNDS, ISOLATED FROM HUMAN BLOOD SERUM REVEALED NEW POTENTIAL BIOMARKERS, ASSOSIATED WITH AUTOIMMUNE AND HEMATOONCOLOGICAL DISEASES

S. Myronovkij[1], M. Starykovych[1], Y. Bobak[1], N. Negrych[2], T. Nehrych[2], M. Shorobura[2], O. Shalay[3], S. Souchelnytskyi[4], R. Stoika[1], Y. Kit[1]*

[1] *Institute of Cell Biology, National Academy of Sciences of Ukraine, Lviv Ukraine*

[2] *Danylo Halytsky Lviv National Medical University, Lviv, Ukraine*

[3] *Institute of Blood Transfusion Medicine, National Academy of Medical Science of Ukraine, Lviv, Ukraine*

[4] *Center for Translational Molecular Medicine, College of Medicine, Qatar University, Doha, Qatar*

** Corresponding author: kit@cellbiol.lviv.ua*

Key words: *proteomics, health, mass-spectrometry*

Blood serum has been extensively explored as a source of markers, as it may contain not only blood proteins per se, but also proteins originating from all tissues of the body. It is estimated that up to 10,000 different proteins (and/or its fragments) may be present in the blood serum, and most of them are in very low concentrations. Selection of a protein preparation, and especially enrichment procedures, may aid in successful search for markers. For concentration of minor protein for MALDI TOF/TOF analysis, and depletion of abundant proteins a 2,2,2-trichloroacetic acid (TCA) precipitation of proteins is frequently applied. However, a significant amount of proteins and peptides may be present in the TCA extracts, and these proteins and peptides are often left non-studied.
Recently, we used two-stops TCA-extraction/acetone precipitation methods in combination with HPLC and MALDI TOF/TOF mass-spectrometry to identify earlier unknown 48 kDa form of human unconventional myosin IC isoform b (48/myo1c) and Ser-Pro-Cys – containing peptides (prosercyne) in blood serum of multiple sclerosis patients (Myronovsky et al,.). Elevating level of 48/myo1c we also detected in blood serum of patients with rheumatoid arthritis, Alzheimer disease, and some hematooncological disease (multiple myeloma, non-Hodgkin's lymphoma, et cetera). Low level of this protein was detected in blood serum of healthy humans but it not detected in blood serum of patients with diabetes 1, cirrhosis, thyroiditis, and recurrent miscarriage. The level of 48/myo1c in blood serum correlates with certain type of human diseases that may have diagnostic value. Cytotoxic effect of 48/myo1c and prosercyne toward some malignant cells and normal lymphocytes in vitro was detected.
Conclusion: The elevating level of 48/myo1c in blood serum correlates with certain type of human diseases that may have diagnostic value. Cytotoxic effect of 48/myo1c and prosercyne toward malignant and normal blood cells in vitro suggests their possible anticancer and potential immunosuppressive activity.

1. S. Myronovskij et al. Isolation and characterization of peptides from blood serum of patients with multiple sclerosis // Studia Biologica. 2015, 9 (2), 5–12.
2. S. Myronovkij et al. Identification of a 48kDa form of unconventional myosin 1c in blood serum of patients with autoimmune diseases // Biochem. Biophys. Reports. 2016, 5. 175-179.
3. S. Myronovskij et al. Identification of Ser-Pro-Cys-Peptide in Blood Serum of Multiple Sclerosis Patients // Protein Pept Lett. 2016 Jun 22. PMID: 27338077

COMPUTER AIDED DRUG DESIGN: DEVELOPMENT MODELS FOR SPECIFICITY, POLYPHARMOCOLOGY AND MEMBRANE PERMEABILITY

G.N. Sastry[1]*, S. Janardhan[1], A.S. Gaur[1], V. Poroikov[2]
[1] *Centre for Molecular Modeling, CSIR-Indian Institute of Chemical Technology, Hyderabad, India*
[2] *Institute of Biomedical Chemistry, Moscow, Russia*
** Corresponding author: gnsastry@gmail.com*

Key words: *Polypharmacology, specificity, membrane permeability, structure based and ligand based approaches, and chemoinformatics*

Motivation and Aim: Developing computational approaches, which provide a direction to the lead identification and optimization is an indispensable area which is of high significance in industry and academia. In addition to QSAR and docking approaches which represent the traditional analog and structure based approaches respectively, optimizing other important factors at times appear to be a major bottleneck. Here, we identify and explore how to effectively and reliable model specificity, membrane permeability and polypharmocology.
Methods and Algorithms: Developed structure and ligand based virtual screening filters. We also describe our drug design platforms being developed in the group. Some of the results are also obtained by employing a range of molecular dynamics (AMBER, CHARMM and DESMOND), quantum chemistry (Gaussian), and other bioinformatics and Chemoinformatics tools.
Results: The talk present how one can effectively use the computational platform Molecular Property Diagnostic Suite (MPDS) that is being developed in the group. The fingerprint, chemotype selectivity and polypharmocology approaches being implemented on selected inhibitors of Mycobacterium tuberculosis and a couple of kianse targets will be discussed.
Conclusion: The talk presents a series of computationally designed lead inhibitors against selected kinase and Mtb targets. It also highlight the ongoing work of developing computational platforms against identified targets.
Availability: Most of the information is available from authors and the software tools are underdevelopment.
Acknowledgements: This work is supported by CSIR and triggered to initiate a collaborative Indo-Russian (DST-RSF) project.
References:
1. P. Badrinarayan and G. N. Sastry, (2013) Curr. Pharm. Design, 19: 4714-4738; ibid (2014) PLoS One, 9: e113773.; ibid (2011) J. Chem. Inf. Model., 34: 89-100.
2. C. Chinmayee, U. Deva Priyakumar and G. N. Sastry, (2015) J. Chem. Inf. Model. 55: 848-860. (2016), (2014) J. Struct. Biol. 187: 38-48. (2016),Mol. Inform., 35: 145-157.
3. S. Janardhan and G. N. Sastry, (2014) Curr. Drug Targets, 15: 600-62.

STUDY ON THE REGULATION OF CELL DIVISION DURING EARLY FRUIT DEVELOPMENT IN TOMATO

H. Nariyama, T. Shibuya, Y. Kanayama*
School of Agricultural Science, Tohoku University, Sendai, Japan
**Corresponding author: yoshinori.kanayama.a7@tohoku.ac.jp*

Key words: *anti-auxin, cell division, 1-naphthylacetic acid, SlHPY2*

Motivation and Aim: Fruit crops exhibit genetic diversity in fruit size, which is one of factors useful for breeding. Generally, fruit size depends on cell division and cell expansion, and fruit development consists of a cell division period and a cell expansion period. Here, we focused on the regulation mechanism of cell division as one of the determining factors for fruit size in tomato as a model plant of fruit crops. In Arabidopsis roots, HIGH PLOIDY 2 (HPY2) induced by auxin reportedly regulates cell division. In this study, auxin and anti-auxin treatments and the expression analysis of SlHPY2, a tomato homolog gene of HPY2, were performed to reveal the effect of auxin on cell division in tomato fruit.
Materials and Methods: For auxin and anti-auxin treatment, tomato 'Ailsa Craig' fruit was dipped into 1-naphthylacetic acid (NAA) every day from 4 days post anthesis, or into anti-auxin every day from anthesis. Fruit diameter was measured at 8 and 15 days post anthesis. Paraffin sections prepared from their pericarp were dyed, and the number of cells per unit length in the mesocarp was measured. Total RNA was extracted from the fruit and SlHPY2 gene expression was measured through quantitative real-time PCR.
Results and Conclusion: The number and size of cells did not differ between NAA-treated fruit and the control at 8 days post anthesis, whereas the number of cells was larger and the cell size was smaller in the treated fruit than in the control at 15 days post anthesis. In the expression analysis, SlHPY2 mRNA levels did not differ between NAA-treated fruit and the control at 8 days post anthesis, whereas the levels rose with NAA concentration at 15 days post anthesis. In the anti-auxin treatment, the number of cells was smaller and the cell size was larger in the treated fruit than in the control at both 8 and 15 days post anthesis; however, SlHPY2 mRNA levels did not differ between the anti-auxin treated fruit and the control. These results suggest that auxin promotes cell division, and that further study is necessary to elucidate the role of SlHPY2. Besides, metabolome and transcriptome analysis was performed throughout fruit development. Metabolites and transcripts specific for early fruit development will be shown and discussed.

WHOLE GENOME OF THE WOOLY MAMMOTH: EVOLUTION THROUGH MILLENIA

A.V. Nedoluzhko[1]*, A.S. Sokolov[2], F.S. Sharko[2], E.S. Boulygina[1], S.V. Tsygankova[1], A.N. Tikhonov[3], K.G. Skryabin[1, 2, 4], E.B. Prokhortchouk[2, 4*]

[1] *National Research Center "Kurchatov Institute", Moscow, Russia*

[2] *Institute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia*

[3] *Zoological Institute, Russian Academy of Sciences, Saint Petersburg, Russia*

[4] *Lomonosov Moscow State University, Faculty of Biology, Moscow, Russia*

** Corresponding author: nedoluzhko@gmail.com, prokhortchouk@gmail.com*

Key words: *genome, woolly mammoth, Mammuthus primigenius, ancient DNA, next-generation sequencing*

Motivation and Aim: Woolly mammoth (*Mammuthus primigenius* Blum.) was an evolutionary dead end of genus Mammuthus which ancestors arose in Africa and migrated to Eurasia almost three million years ago. Paleozoologists described several species in these genera – *M. meridionalis* (Early Pleistocene), *M. trogontherii* (Middle Pleistocene), *M. primigenius* (Late Pleistocene) and other. Woolly mammoth had appeared 300 – 200 thousands years ago in Siberia and colonized Europe and North America.

Methods and Algorithms: In our study we sequenced DNA from different tissues of woolly mammoth calf known as Khroma (AMS age was at background levels, that is >50,000 years). Ancient DNA has been successfully extracted from a skeletal muscle, a piece of skin, a piece of skin covered with hair, and two different bone samples.

Results: Here we report whole genome sequence of the woolly mammoth genome using Illumina GAIIx and HiSeq 1500 platforms we got 16x coverage of the woolly mammoth genome. We used VarScan.v2.3.7 with parameter – $p\text{-}value < 0.05$ for snp-calling. After that we got 17,732,776 SNPs with *p-value* less than 0.05. Of these SNPs only 4,011,148 lie inside gene part. For subsequent analysis we used only SNPs with *p-value* less than 0.001 and with at least 20x coverage. After that procedure only 1,550,394 SNPs (inside gene part) has been retained.

Conclusion: To conduct Gene Ontology analysis we converted particular genes of interest into human orthologues. First we considered genes with missense variants. Next we considered genes with stop gained variants. We founded a lot of stop gains in rapidly evolving olfactory and MHC genes. Using Khroma genome we supposed several SNPs and also founded several interesting non-synonymous substitutions in globin and dynein genes.

Acknowledgements: We would like to thank Dr Ludovic Orlando for allowing the use of the laboratory infrastructure for ancient DNA manipulations and also for his careful reading of the manuscript and his valuable suggestion.

REGULATION OF THIOREDOXIN GENES EXPRESSION IN DESICCATION-TOLERANT INSECT POLYPEDILUM VANDERPLANKI

A.A. Nesmelov[1]*, E.I. Shagimardanova[1], M.D. Logacheva[2], R. Cornette[3], T. Kikawada[3], O.A. Gusev[1, 3, 4, 5]

[1] *Institute of Fundamental Biology and Medicine, Kazan Federal University, Kazan, Russia*
[2] *Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*
[3] *National Institute of Agrobiological Sciences (NIAS), Tsukuba, Japan*
[4] *PMI Riken, Yokohama Campus, Yokohama 230-0045, Japan*
** Corresponding author: nesmelov@gmail.com*

Key words: *P. vanderplanki, thioredoxin, DNA motif discovery*

Motivation and Aim: Polypedilum vanderplanki is the most complex organism able to survive desiccation. *P. vanderplanki* genome contains both common insect thioredoxin (TRX) genes and TRX genes specific for this insect [1]. We test our hypothesis that *P. vanderplanki*-specific TRX genes and common insect TRX genes differ in their regulation. The aim: to compare DNA motifs in regulatory sequences of TRX genes in *P. vanderplanki* genome.
Methods and Algorithms: Isolation of DNA sequences from *P. vanderplanki* genome using bedtools [1]. Analysis of isolated DNA sequences using MEME tool [2].
Results: TRX genes specific for *P. vanderplanki* share similar DNA motifs in their regulatory sequences. These DNA motifs are not associated with common insect TRX genes in *P. vanderplanki* genome.
Conclusion: Revealed difference of regulatory sequences between TRX genes specific for *P. vanderplanki* and common insect TRX genes reflects the difference in their regulation. Specific DNA motifs associated with *P. vanderplanki*–specific TRX genes are responsible for an upregulation of these genes in desiccation.
Acknowledgements: The work was supported by Ministry of Education and Science of the Russian Federation (Research identifier RFMEFI58414X0002).
References:
1. O. Gusev et al., (2014) Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge, *Nature communications, **5:*** 4784.
2. A.R. Quinlan, I.M.Hall. (2010) BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics,* **26:** 841–842.
3. T.L.Bailey et al., (2009) MEME SUITE: tools for motif discovery and searching, *Nucleic Acids Research,* Jul;37(Web Server issue):W202-8.

DEVELOPMENT OF MICROSATELLITE MARKERS ACCORDING TO BAC SEQUENCING DATA AND THEIR PHYSICAL MAPPING TO THE BREAD WHEAT 5B CHROMOSOME

M.A. Nesterov[1]*, D.A. Afonnikov[1], E.M. Sergeeva[1], L.A. Miroshnichenko[2], M.K. Bragina[1], A.O. Bragin[1], G.V. Vasiliev[1], E.A. Salina[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia*
* *Corresponding author: mikkanestor@bionet.nsc.ru*

Key words: *bread wheat, Triticum aestivum, BAC clones, Ion Torrent, MIRA, microsatellites, chromosome 5B, SSR markers, $(AAG)_n$*

The shortage of polymorphic markers for the regions of wheat chromosomes that encode commercially valuable traits determined the need for studying wheat microsatellite loci. In this work, SSR markers for individual regions in the short arm of bread wheat chromosome 5B (5BS) were designed based on sequencing data for BAC clones, and the regions of the corresponding chromosome were saturated with these markers. Totally, 130 randomly selected BAC clones from the 5BS library were sequenced on the Ion Torrent platform and assembled in contigs using MIRA software. The assembly characteristics (N50 = 4136 bp) are comparable to the recently obtained data for wheat and relative species and acceptable for identification of microsatellite loci. An algorithm utilizing the properties of complexity decompositions in the sliding-window mode was used to detect DNA sequences with a repeat unit of 2–4 bp. Analysis of 17770 contigs with the total length of 25879921 bp allowed for designing 113, 79, and 67 microsatellite (SSR) loci with a repeat unit of 2, 3, and 4 bp, respectively. The SSR markers with a motif of 3 bp were tested using nullitetrasomic lines of Chinese Spring wheat homoeologous group 5. Thus, 21 markers specific for chromosome 5B were detected. Seven of these markers were mapped to the distal region of this chromosome (bin 5BS6) using a set of Chinese Spring deletion lines for 5BS. Eight and four markers were mapped to the interstitial region (bins 5BS5 and 5BS4, respectively). Two markers were mapped to a pericentromeric bin. A comparative analysis of the distribution of trinucleotide microsatellites over wheat chromosome 5B and in different cereal species suggests that the (AAG)n repeat has proliferated and has been maintained during the evolution of cereals.

CMSEARCH: TOOL FOR SEARCHING TFBS COMPOSITE MODULES IN DNA SEQUENCES

S.I. Nikitin*, E.S. Cheryomushkin
A.P. Ershov Institute of Informatics Systems SB RAS, Novosibirsk, Russia
Novel Computing Systems in Biology LLC, Novosibirsk, Russia
** Corresponding author: sergey_post@inbox.ru*

Key words: *algorithm, binding sites, transcription factors, composite motif, position weight matrices, sensitivity, positive predictive value, transcription factor binding site, DNA*

Motivation and Aim: The DNA sequences consist of coding areas and regulatory regions, which control gene synthesis. On the start of gene production, transcription factors bind to short sequences called binding sites. As known the transcription factors are often being combined in groups and bind to the respective binding sites coordinately [1]. To predict these transcription complexes binding areas is a scientific challenge nowadays. All existing algorithms show low prediction accuracy and often have specific limitations.
Methods and Algorithms: We introduce new prediction algorithm called CMSearch. It uses position weight matrices (PWM) as transcription factors description and Match algorithm as the accessory instrument for binding sites prediction [2]. Composite module can be defined by setting the weight matrices it contains and rules each PWM interacts with other. We predict single motifs for PWMs contained in selected composite module, and then find all binding sites groups corresponding to the composite module template.
Results: We use the framework developed by the Klepper et al. [3] to compare out algorithm with other developed tools. We consider the sensitivity, positive predictive value and performance coefficient to estimate efficiency of the algorithms. CMSearch showed results comparable with well-known efficient tools like Stubb and CMA [4].
Conclusion: We introduced CMSearch – the novel tool to finding transcription factor binding sites groups defined by setting the position weight matrices and the rules each matrix interacts with other in module. This algorithm showed good results among the other well-known tools.
Availability: The algorithm is distributed as part of the Proteome commercial system, Biobase/Qiagen.
Acknowledgements: This research was supported by Biobase GmbH (now the part of the Qiagen corporation) and made by using Transfac database.
References:
1. Spitz F, Furlong EE. (2012) Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics, 13: 613-626.
2. Kel,A.E. et al. (2003) MATCH: A tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Research, 31: 3576-3579.
3. Klepper, K., Sandve, G., Abul, O., Johansen, J., Drablos, F. (2008) Assessment of composite motif discovery methods. BMC Bioinformatics, 9: 123.
4. Kel A.E. et al. (2006) Composite Module Analyst: a fitness-based tool for identification of transcription factor binding site combinations. Bioinformatics, 22(10): 1190-1197.

A CONGESTION GAME MODEL FOR VIRTUAL DRUG SCREENING IN A DESKTOP GRID

N.N. Nikitina*, E.E. Ivashko
Institute of Applied Mathematical Research, Karelian Research Center RAS, Petrozavodsk, Russia
** Corresponding author: nikitina@krc.karelia.ru*

Key words: *virtual drug screening, protein-ligand docking, desktop grid, BOINC, congestion game*

Motivation and Aim: Virtual drug screening is a significant part of drug development process as it allows reducing the chemical space of the order of 10^{60} potentially synthesizable compounds down to a manageable set for laboratory testing. However, pre- and post-filtering of the chemical space consume a lot of effort and time which makes structure-based virtual drug screening over very large databases unfeasible. Small focused libraries are not always readily available for new drug targets. In this work we address a game-theoretic model for filtering the explored compounds space on the fly when performing structure-based virtual drug screening over very large databases.
Methods and Algorithms: Being a computational technique to process many independent fine-grained tasks, virtual drug screening essentially involves a set of computational nodes that may be seen as independent agents, each of them willing to perform as much useful work as possible. We model virtual drug screening process as a congestion game between computational nodes who compete for a shared pool of resources, namely subsets of computational tasks. The utility of each player depends not only on the value of the chosen resource, but also on the number of other players choosing it (the "congestion level"). According to drug development principles, two primary characteristics of the resulting set of compounds to be laboratory tested are their estimated efficiency of interaction with the disease-relevant target and their structural diversity. Ranking compounds by the former characteristic is performed by molecular docking software. We propose to attain the structural diversity of the results by employing the competition between computational nodes that tend to select the computational task subsets that are in less demand by other nodes.
Results: We propose a congestion game model for virtual drug screening. The game has at least one Nash equilibrium in pure strategies; best- and better-response dynamics are guaranteed to converge to equilibrium in polynomial time. The social utility function expresses efficiency and diversity of the resulting set of compounds. The developed algorithms for taskflow management are being implemented and tested within the BOINC-based Enterprise Desktop Grid for virtual drug screening [1].
Acknowledgements: This work is supported by the Russian Fund for Basic Research (projects 16-07-00622 A and 15-29-07974 ofi_m).
References:
1. N. Nikitina, E. Ivashko, S. Möller (2014) BOINC-based Desktop Grid Infrastructure for Virtual Drug Screening, Abstracts of the International Conference on Mathematical Modeling and High Performance Computing in Bioinformatics, Biomedicine and Biotechnology, MM-HPC-BBB-2014: 62.

MECHANICS OF PLANT CELL UNIDIRECTIONAL GROWTH

S.V. Nikolaev[1, 2]*, S.K. Golushko[2], U.S. Zubairova[1], D.A. Afonnikov[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Design and Technology Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
** Corresponding author: nikolaev@bionet.nsc.ru*

Key words: *plant cell wall, growth mechanics, orthotropic material*

Motivation and Aim: Mathematical modeling is powerful method to study plant cell and tissue growth. Turgor pressure and cell wall strain are considered to be driving force for cell growth. There are vertex dynamics-based (VD) models that are used to simulate plant tissue growth in frame of the growth mechanics. It is worth to note that the mechanics in these models is oversimplified, and problems with adequacy arise especially in 2D-models of unidirectional growth. Some modifications were proposed to account effects of the cell walls in the plane of a 2D simulation (for example, additional structure elements, and restrictions on vertex possible movements). In this work we developed more realistic solid and shell mechanics models to study elastic behavior of plant cells with isotropic and orthotropic materials of cell wall.
Methods and Algorithms: Approach of structural mechanics was applied to model plant cell wall material. The parameters of isotropic and orthotropic materials of cell walls were found in articles. The model of plant cell mechanics were developed in the COMSOL package using shell mechanics and solid mechanics interfaces.
Results: Calculations of stress-strain distribution in cell walls were performed in elastic mode. The calculations were done with different sets of parameters, inside-outside difference of pressure, and additional forces applied to cell walls which imitate mechanical interaction of modeled cell with its neighbors. The results were used to infer growth distribution (plastic deformations) in the cell walls.
Conclusion: The results demonstrate: 1) orthotropy of cell wall (small Young modulus along growth axis, and large one in perpendicular direction) can equalize stress distribution under different additional forces from neighbor cells; 2) wide adopted picture that cell wall growth (plastic deformation) can be local and proportional to stress (strain) has to result not only in stress relaxation, but in increasing plastic deformation, and deviation from ("planned") form, so this picture may demand a correction.
Acknowledgements: The authors expresses gratitude to the supercomputer center of Novosibirsk State University for the opportunity to use COMSOL 4.3b package. The reported study was supported by the Russian Science Foundation according to the research project № 14-14-00734.

IN SILICO SCREENING FOR SULFONATE-BASED INHIBITORS AGAINST PROMISING ANTICANCER TARGETS

D.K. Nilov[1]*, I.V. Gushchina[2], V.K. Svedas[1, 2]
Lomonosov Moscow State University
[1] Belozersky Institute of Physicochemical Biology, Russia
[2] Faculty of Bioengineering and Bioinformatics, Moscow, Russia
** Corresponding author: nilov@belozersky.msu.ru*

Key words: *molecular modeling, docking, enzyme inhibitors, sulfo group, sulfonates, lactate dehydrogenase, tyrosyl-DNA phosphodiesterase 1, cancer*

Motivation and Aim: Human lactate dehydrogenase A (LDH-A) and tyrosyl-DNA phosphodiesterase 1 (TDP-1) constitute attractive therapeutic targets in cancer metabolism [1, 2]. These enzymes convert substrates containing negatively charged carboxyl and phosphate groups. Virtual screening of compounds with a sulfo substituent that mimic abovementioned functional groups was performed to identify novel competitive inhibitors of LDH-A and TDP-1.

Methods and Algorithms: Molecular models of LDH-A and TDP-1 were constructed on the basis of available crystal structures taking into account the ionization states of amino acid residues, and structural criteria for the selection of potential inhibitors were established. A library of commercially available low-molecular-weight sulfo derivatives was subjected to virtual screening. Docking of compounds into the protein models was performed using the Lead Finder software [3]. Inhibitory effects were tested *in vitro* against purified proteins.

Results: The most effective inhibitors selected by virtual screening and experimental validation were able to form hydrogen bonds with Arg168 in the active site of LDH-A, and with Lys265 and Lys495 in the active site of TDP-1. The sulfo group of the inhibitors was shown to occupy the position of the carboxyl or phosphate group of the corresponding substrate. Dependence of inhibitory properties of sulfonates on their structure was analyzed and directions for further structural modification of compounds were proposed.

Conclusion: Molecular models of human LDH-A and TDP-1 were used to screen a library of low-molecular-weight sulfonates what enabled us to identify potential inhibitors of both enzymes. The novel LDH-A and TDP-1 inhibitors will be further investigated as target-driven anticancer agents.

Acknowledgements: This work was supported by the Russian Foundation for Basic Research (grant 14-08-01251) and by the President of the Russian Federation (grant for young scientists MK-7630.2016.4).

References:

1. C. Granchi et al. (2013) Small-molecule inhibitors of human LDH5, *Future Med. Chem.*, 5: 1967-1991.
2. E.Q. Comeaux, R.C. van Waardenburg (2014) Tyrosyl-DNA phosphodiesterase I resolves both naturally and chemically induced DNA adducts and its potential as a therapeutic target, *Drug Metab. Rev.*, 46: 494-507.
3. O.V. Stroganov et al. (2008) Lead finder: an approach to improve accuracy of protein-ligand docking, binding energy estimation, and virtual screening, *J. Chem. Inf. Model.*, 48: 2371-2385.

PROTEOMIC SCREENING FOR AMYLOID-FORMING PROTEINS IN BACTERIA *ESCHERICHIA COLI*

A.A. Nizhnikov[1, 2, 3]*, K.S. Antonets[1, 2], K.V. Volkov[1], A.L. Maltseva[1], A.P. Galkin[1, 2]
[1] *St. Petersburg State University, St. Petersburg, Russia*
[2] *Vavilov Institute of General Genetics (St. Petersburg Branch), St. Petersburg, Russia*
[3] *All-Russian Research Institute for Agricultural Microbiology, St. Petersburg, Russia*
* *Corresponding author: ant.nizhnikov@gmail.com*

Key words: *amyloid, E.coli, HPLC, MALDI, protein identification*

Motivation and aims: Amyloids represent protein fibers exhibiting cross-beta structure. They are found in different organisms, from bacteria to human, and can be pathogenic, useless, and functional. Functions of amyloids in bacteria spread from biofilm formation to parasite-host interaction. Importantly, bacterial amyloids identified to date were found accidentally, and screenings for amyloid-forming proteins in the proteomes of these organisms were never carried out before. The goal of this study was to implement a proteome-wide screen for candidates for amyloid-forming proteins in well-known prokaryote *Escherichia coli*.
Methods: Screening for candidates for novel amyloid-forming proteins was performed with previously developed PSIA (Proteomic Screening and Identification of Amyloids) approach [1] improved by HPLC-separation of the tryptic peptides [2].
Results: We identified 61 detergent-resistant proteins. This protein set was 3-5 fold enriched with potentially amyloidogenic regions predicting by different bioinformatics algorithms (WALTZ, SARP) in comparison with the entire *E. coli* proteome. 56 of 61 proteins contain potentially amyloidogenic regions, and four (BcsC, MukB, YfbK, and YghJ) carry low-complexity N- and Q-rich regions that is the hallmark of a number of known amyloid-forming proteins [2].
Conclusion: There is unexpected diversity of *E.coli* proteins forming detergent-resistant aggregates *in vivo* at the physiological level of expression. These proteins are rich in potentially amyloidogenic low-complexity regions.
Acknowledgements: This work was supported by the grants of the President of the Russian Federation (MK-4854.2015.4), Russian Foundation for Basic Research (16-34-60153, 14-04-01463), and St. Petersburg Government. The authors acknowledge Saint-Petersburg University for research grants 1.50.2543.2013 and 1.37.291.2015, and the opportunity to use facilities of the "Research Resource Center for Molecular and Cell Technologies" and "Chromas."
References:
1. Nizhnikov A.A. *et al.* (2014) Proteomic screening for amyloid proteins, *PLoS One*, 9: e116003.
2. Antonets K.S. *et al.* (2016) Proteomic Analysis of *Escherichia coli* Protein Fractions Resistant to Solubilization by Ionic Detergents, *Biochemistry(Mosk)*, 81:34-46.

IDENTIFICATION OF STURGEON SPECIES WITH MTDNA AND MICROSATELLITE MARKERS IN BELARUS

A.Yu. Nosova
Institute of Genetics and Cytology NAS of Belarus, Minsk, Belarus
** Corresponding author: A.Nosova@igc.by*

Key words: *Genotyping, sturgeon, species identification, hybrids*

Motivation and Aim: Sturgeons are a unique ancient group of fishes and are commercially important. Since the natural populations have dramatically decreased in size, sturgeon domestication and artificial rearing become widespread. Usually, there is no sufficient work on the preservation of the purity of livestock in aquaculture. However, the cost of products is significantly different depending on the species. Identification of the species is complicated by the morphological similarity of sturgeons of commercial size, or impossible for fish products and caviar. This study was conducted to determine the efficiency of using mtDNA and microsatellite markers for the identification of sturgeon species in Belarus.

Methods and materials: We obtained fins fragments of 32 individuals including 7 Siberian sturgeons, 5 Russian sturgeons, 5 starlets, 5 belugas and 10 interspecific hybrids from local fish farm. For species identification variations of the mtDNA control region (D-loop) were studied by PCR analysis [1]. Sturgeons were genotyped by fragment analysis of a set of five microsatellite loci (Afug41, Afug51, An20, AoxD161, AoxD165) [2].

Results: The analysis of mtDNA variations and microsatellite alleles confirmed that 22 individuals of pure species and 10 of interspecific hybrids were provided. Hybrid individuals were represented by 4 Besters (beluga×starlet), 4 Steroses (starlet×Siberian sturgeon). Presumably, two samples differing from the declared were reciprocal hybrids starlet×beluga and Siberian sturgeon×starlet.

Conclusion: The methods of DNA identification can be conventionally divided into two groups addressing nuclear genetic markers and mitochondrial DNA (mtDNA). However, the maternal inheritance of mtDNA limits the use of this methodology for the hybrid individuals. The differences in STR allele frequencies among different species enable to identify individuals of hybrid origin. However, it should be taken into account that STR loci may vary from population to population. In our case, not all species-specific loci according to Barmintseva [2] were presented in studied animals.

Availability: Thus, the use of PCR for identification is appropriate for maintaining the genetic purity of livestock and producers in aquaculture, and for detection of falsified caviar and other sturgeon products. An analysis of the polymorphism of the five microsatellite loci should be used in the genetic pasportisation of aquaculture surgeon stocks and the species verification of the sturgeon products.

References:

1. N.S. Mugue et al. (2008) Polymorphism of the Mitochondrial DNA Control Region in Eight Sturgeon Species and Development of a System for DNA-Based Species Identification, *Russian Journal of Genetics*, 44: 793-798.
2. A.E. Barmintseva, N. S. Mugue (2013) The Use of Microsatellite Loci for Identification of Sturgeon Species (Acipenseridae) and Hybrid Forms, *Russian Journal of Genetics,* 49: 950-961.

MEMBRANE-ASSOCIATED KINASE REGULATORS OF MAKR FAMILY GENES IN *ARABIDOPSIS THALIANA* L.

D.D. Novikova*, N.A. Omelyanchuk, V.V. Mironova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
NSU
** Corresponding author: novikovadd@bionet.nsc.ru*

Key words: *kinase receptor regulators, A. thaliana, kinase receptors, hormone signaling*

Motivation and Aim. Tyrosine kinase receptors and membrane-associated kinase regulators are involved in most signaling pathways, including those for plant hormone signal transduction. Recently, a family of membrane-associated kinase regulators (MAKR) was described in *Arabidopsis thaliana* [1]. MAKR family proteins have distant homology with the basic domain of BRI1 kinase inhibitor 1 (BKI1) which controls activity of tyrosine kinase receptor BRI1 in brassinosteroid signaling pathway. Here, we study phylogeny and systematize the publically available information about protein structure, expression and regulation of this poorly annotated gene family.

Methods and Algorithms. First we analyzed the MAKRs protein sequences to reveal their features and signatures which might imply their potential functions by PhosphoSVM and CBS Prediction servers. Further to investigate phylogenetic background and evolvement of MAKRs we have searched for their homologs throughout the plant kingdom using BLAST and PLAZA on-line services and constructed phylogenetic trees in MEGA 6.06 program. We analyzed data on their mRNA patterns which is presented in publicly available microarray experiments (eFP browser). To study the hormonal regulation of MAKRs in depth we analyzed their upstream [-1000; +1] regions for the presence of potential cis-regulatory sites using online tools AtCOECIS, TRANSFAC programs and CIS-BP database.

Results and conclusions. MAKR family genes probably evolved together with land plants and as a result of duplication series all members of the family appeared. These proteins are widely spread among terrestrial plants. The latter uncovers their potentially important role in plant kingdom in general and in *Arabidopsis* in particular. Also some conservation was discovered among orthologs of all MAKRs, but none among paralogs, except C-terminus. According to protein modification prediction, all of MAKRs have plenty of modification sites. MAKRs are expressed tissue specifically and response to different hormones, they possess hormone responsive motifs corresponding to the sensitivity to the one. These findings suggest that MAKR family genes might play an important role in plant development regulation via signaling pathways. Obtained results may provide new ideas in searching for new regulators of growth and herbicides.

References.

1. Jaillais Y. et al. Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor // Genes & development. – 2011. – Т. 25. – №. 3. – С. 232-237.

MAPPING THE INTERACTION SITE FOR THE MESOBUTHUS SCORPION TOXINS IN THE VOLTAGE-GATED POTASSIUM CHANNEL KV1.2

V.N. Novoseletsky*, A.D. Volyntceva, K.V. Shaitan
Lomonosov Moscow State University, Moscow, Russia
** Corresponding author: valeryns@gmail.com*

Key words: *molecular modeling, homology modeling, molecular dynamics, potassium channel, scorpion toxins*

Motivation and Aim: Voltage-gated potassium (K+) channels are transmembrane pore proteins and contribute to the regulation of membrane potential and, consequently, to cell excitability. Kv1.2 channels play pivotal role in maintaining of resting membrane potential and, consequently, regulation of cellular excitability of neurons. Their blockers have a high importance not only as probes the fundamental channel functioning investigation, but also as a potential drug for treatment epilepsy [1]. Goal of the current study was an interface analysis in complexes of Kv1.2 channel with peptide toxins MeuKTx1, MeuKTx3 и MeuKTx3B, derived from *M. eupeus* venom.
Methods and Algorithms: 3D structure was generated by homology modeling using Kv1.2-2.1 paddle chimera channel in complex with charybdotoxin (pdb-code 4JTA) as a template and equilibrated by molecular dynamic simulation in Gromacs software. Analysis of hydrophobic and stacking interactions, hydrogen and ionic bonds of the toxin and potassium channels was performed for representative frames with optimal toxins orientations using program Platinum [2] and APBS software package [3].
Results: Performed study revealed, that toxin MeuKTx3 demonstrates the highest affinity to Kv1.2 channel. Contacts analysis allowed identifying key residues for binding process and the possible mutation points for enhancing toxin MeuKTx3 activity.
Conclusion: The results of investigation are in good agreement with the experimental values of binding constants, obtained by competitive binding assays [4]. Results of the conducted investigation may find an application in fundamental science and drug design.
Acknowledgements: The research was supported by the Russian Science Foundation grant № 14-14-00239. Simulations were performed using the Supercomputing Center of Lomonosov Moscow State University.
References:
1. X. Wang et al. (2015) Mesomartoxin, a new Kv 1.2-selective scorpion toxin interacting with the channel selectivity filter. Biochemical Pharmacology, 93(2): 232-239.
2. S. Unni et al. (2011) Web servers and services for electrostatics calculations with APBS and PDB-2PQR. Journal of computational chemistry, 32(7): 1488-91.
3. T. Pyrkov et al. (2008) PLATINUM: Protein-Ligand Attractions Investigation Numerically, www.model.nmr.ru/platinum.
4. B. Gao et al. (2010) A potent potassium channel blocker from Mesobuthus eupeus scorpion venom, Biochimie, 92(12): 1847-1853.

DISTRIBUTION OF 2541-2542DELCA *KDPD* FRAMESHIFT MUTATION IN GENOMES OF *MYCOBACTERIUM TUBERCULOSIS* FROM IRKUTSK OBLAST AND YAKUTIA

O. Ogarkov[1, 2, 3]*, V. Sinkov[1], I. Mokrousov[4], S. Zhdanova[1], P. Khromova[1], E. Orlova[1, 3]
[1] *Scientific Centre of the Family Health and Human Reproduction Problems, Irkutsk, Russia*
[2] *Irkutsk State Medical Academy of Continuing Education, Irkutsk, Russia*
[3] *Irkutsk State University, Irkutsk, Russia*
[4] *Laboratory of Molecular Epidemiology and Evolutionary Genetics, St. Petersburg Pasteur Institute, St. Petersburg, Russia*
** Corresponding author obogarkov@yandex.ru*

Key words: *tuberculosis, genomes, CC2-W148*

Motivation and Aim: Whole-genome sequencing of 110 representative isolates of *M.tuberculosis* genotype Beijing [1] found 2541-2542delCA frameshift mutation in gene *kdpD*. This mutation strongly associated with CC2-W148 high-transmissible strain of tuberculosis. More other a partial deletion of the *kdpDE* operon in *M. tuberculosis* has already been associated with greater virulence [2]. The study was carried out for determination of the distribution this mutation among genomes of GMTV database [3] and clinical isolates.
Methods and Algorithms: The publicly available WGS data of from the GMTV database (DB) [3] and DNAs of 256 clinical isolates from Irkutsk Oblast and Yakutia have been analyzed. *In silico* mutation was studied by in-house Perl-written annotation tool snpMiner2 [4]. *In vitro* 2541-2542 delCA was detected by RT-PCR with specially designed TaqMan probes.
Results: In the GMTV DB were found only 40 genomes of Beijing genotypes with investigated mutation. However by PCR results 78 (30.5%) isolates from 256 known Beijing genotypes within two regions had this mutation. In addition two strains were mix W148\nonW genotypes. The most of W148 stains (81.3%) were MDR and only 6 (7.5%) were susceptibility. Among 176 non-W148 strains only 38 (21.6%) were MDR and 102 (58.0%) susceptibility (χ^2=7.9; p<0.01). In addition the analysis of drug resistance profiles of 256 isolates revealed an interesting pattern. W148 strains had significantly less resistance to cycloserine and ethionamide (χ^2 =10.8; p<0.01). However the analysis of mutations in *Rv0486, Rv3199c, Rv3793, Rv3794* of Beijing genomes (GMTV DB) has not confirmed this pattern.
Conclusion: Investigation of 2541-2542delCA is a convenient tool to identify highly transmissible CC2-W148 genotype of tuberculosis *in silico* and *in vitro.*
*Acknowledgments***:** The work was supported by RFBR grant #16-04-00160 A.
Availability: 10.5281/zenodo.51052
References:
1. M. Merker et al. (2015). Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage, Nat Genet., 47(3): 242-249.
2. T. Parish, et al. (2003). Deletion of two-component regulatory systems increases the virulence of Mycobacterium tuberculosis, Infect. Immun., 71: 1134-1140.
3. E.N. Chernyaeva et. al. (2014). Genome-wide Mycobacterium tuberculosis variation (GMTV) database: a new tool for integrating sequence variations and epidemiology, BMC Genomics, 15(1): 308.
4. V.V. Sinkov (2016). snpMiner2: vcf files annotation tool. Zenodo. 10.5281/zenodo.51052.

VASCULAR ENDOTHELIAL GROWTH FACTOR POLYMORPHISMS ARE ASSOCIATED WITH THE EARLIER ONSET OF RHEUMATOID ARTHRITIS

V.O. Omelchenko[1]*, M.A. Korolev[1], E.A. Letyagina[1], A.V. Shevchenko[1], V.F. Prokof'yev[1], T.I. Pospelova[2], V.I. Konenkov[1]

[1] *Scientific Institute of clinical and experimental lymphology SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State Medical University, Novosibirsk, Russia*

** Corresponding author: v.o.omelchenko@gmail.com*

Key words: *rheumatoid arthritis, SNP, VEGF A, rs3025039, rs699947*

Motivation and Aim: Genetic factors are involved in the developing of rheumatoid arthritis (RA). In RA pathogenesis a pannus formation play a crucial role. One of the factors contributing to this process is the vascular endothelial growth factor (VEGF). The existence of an association between the VEGF SNP and RA progress is supposed. The aim of this study was to investigate the involvement of VEGF+C936T (rs3025039) and VEGF-C2578A (rs699947) SNPs in developing of RA.

Methods and Algorithms: 229 Europeoid women with RA were included in our study. Patients had American College of Rheumatology (ACR)-defined RA (1987 classification criteria). The genotyping was performed by restriction fragment length polymorphism analysis of PCR-amplified fragments (PCR-RFLP), furthermore the age of disease onset was determined. Description analysis and Mann—Whitney U-test were employed for statistical processing the results.

Results: No differences were found between groups with VEGF-C2578A (rs699947) SNPs. The age of disease onset was significantly different between group with various genotypes for VEGF+C936T SNPs: 42,5 years for VEGF+936CC compare with 47,8 years for VEGF+936CT (p-value=0,015). 168 (73,36%) and 59 (25,76%) patients had VEGF+936CC and VEGF+936CT genotype, respectively. The last mentioned are consistent with other investigator's results. Two patients (0,88%) had VEGF+936TT genotype and this small group wasn't included in analysis.

Conclusion: Analysis of vascular endothelial growth factor polymorphisms (rs3025039) may be useful in clinical practice for evaluation of predisposition to earlier RA.

THEORETICAL MODEL OF MITOTIC SPINDLE MICROTUBULE GROWTH FOR FRAP CURVE INTERPRETATION

L.V. Omelyanchuk[1, 2]*, A.F. Munzarova[1, 2], T.Y. Mikhailova[2]
[1] *Institute of Molecular and Cellular Biology, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*

Key words: *mitosis, modeling*

Spindle FRAP recovery curve depends upon kinetic parameters of polymerization at microtubule plus ends. Empirical FRAP recovery curve of Salmon et al., 1984 permits to determine only one such dynamic parameter, commonly called as “tubulin turnover”. The aim of our study was to build FRAP recovery curve based upon already known kinetic model of microtubule growth. Analytical expression describing the distribution of polymerizing and degrading microtubule ends as a function of four kinetic parameters was found. Theoretical FRAP recovery curve for spindle was constructed. Fitting the theoretical curve to experimental data gives the values of four parameters, describing spindle microtubule behavior. This gives the opportunity to study how mutation in mitotic proteins affects microtubule growth and shrinking. The alterations in kinetic constants of the transition between growth and shrinking and between shrinking and growth could also be determined for mutant proteins.

SEQUENCING OF CONIFER GENOMES USING NGS

N.V. Oreshkova[1, 2]*, Yu.A. Putintseva[1, 2], D.A. Kuzmin[1], V.V. Sharov[1], V.V. Biryukov[1], S.V. Makolov[1], K.V. Krutovsky[1, 3, 4, 5]

[1] *Siberian Federal University, Krasnoyarsk, Russia*
[2] *V.N. Sukachev Institute of Forest SB RAS, Krasnoyarsk, Russia*
[3] *Georg-August University of Göttingen, Göttingen, Germany*
[4] *N.I. Vavilov Institute of General Genetics, RAS, Moscow, Russia*
[5] *Texas A&M University, College Station, USA*
** Corresponding author: oreshkova@ksc.krasn.ru*

Key words: *genome sequencing, genome assembly, organelle genomes, Pinus sibirica, Larix sibirica*

Motivation and Aim: Pinus sibirica and *Larix sibirica* are the key forest-forming species of great ecological and economical importance. To investigate specific features of these species, interspecific interactions, genetic diversity and for other important studies it is necessary to assemble and annotate reference genomes.

Methods and Algorithms: The Siberian larch (*Larix sibirica* Ledeb.) and Siberian pine (*Pinus sibirica* Du Tour.) nuclear and organelle genomes are being *de novo* sequenced in the Laboratory of Forest Genomics at the Siberian Federal University using Illumina HiSeq 2000 and MiSeq, and their first draft genome assemblies were generated. Estimated genome size was 12.03 Gbp for Siberian larch and 28.90 Gbp for Siberian pine. DNAs isolated from needles, single megagametophytes and a haploid tissue culture of a reference larch tree and from needles and single megagametophytes of a reference pine tree were used to generate multiple PE libraries with 250, 400 and 500 bp long inserts and MPE libraries representing 3 and 5 Kbp long fragments. We tested CLC Assembly Cell, ABySS and MaSuRCA assemblers that were used in the similar *Picea abies, Picea glauca* and *Pinus taeda* conifer genome sequencing projects, respectively.

Results: The best Siberian larch genome assembly was ~5.5 Gbp long (that is 46% of the expected complete genome length) with N50 for contigs equaled 1947 bp. Almost all Siberian pine short reads were successfully mapped to the draft genome assembly v1.0 of closely related sugar pine (*Pinus lambertiana* Dougl.) generated in the PineRefSeq project (http://pinegenome.org/pinerefseq) covering more than 80% of the assembly (~21.26 Gbp). Thus, the reference-based together with *de novo* assembly approaches resulted in a draft genome assembly of Siberian pine with a total length of ~22.9 Gbp (79% of the expected complete genome length) with N50 for contigs equaled 2352 bp. About 80% of Siberian larch and pine nuclear genomes consisted of highly repetitive DNA.

Conclusion: Thus, we obtained 46 and 79% of the length of Siberian larch and Siberian stone pine genomes, respectively. Currently, work is underway to improve the quality of assembly and annotation.

Acknowledgements: This study was supported by Research Grant No. 14.Y26.31.0004 from the Government of the Russian Federation.

CLUSTER ANALYSIS OF STRESS-INDUCED DUPLEX DESTABILIZATION (SIDD) PROFILES FOR *E. COLI* PROMOTERS

M.A. Orlov*, A.A. Ryasik, E.A. Temlyakova, A.A. Sorokin
Institute of Cell Biophysics RAS, Pushchino, Russia
** Corresponding author: orlovmikhailanat@gmail.com*

Key words: *DNA physical properties, machine learning, cluster analysis, bacterial promoters*

Motivation and Aim: The physical properties of promoter DNA play a key role in all steps of transcription initiation. The propensity to undergo stress-induced duplex destabilization (SIDD) [1] could contribute significantly to the formation of open promoter complex. It was shown that SIDD can be used successfully for promoters identification and when used in combination, it is shown to significantly reduce the level of false positive predictions [1].

Methods and Algorithms: Sequences of experimentally found *E. coli* promoters were obtained from RegulonDB (version 8.5). SIDD profiles were calculated for each sequence in the set. Following clusterization of the profiles was performed with Ward method; its results consistency was assessed using consensus clustering technique.

Results: Analysis of the SIDD data shows that approximately half of promoters have no significant maxima corresponding to the opening probability over 50% for .... According to clusterization rest of promoters are grouped into three stable compact clusters with opening probability maxima nearly -175, -45, and +20 positions from transcription start site (TST) accordingly. We suppose that regions with high probability to open located outside of a region, which interact with polymerase directly, serve as a 'safety valves' that keep the superhelicity of promoter DNA stable.

Conclusion: Analysis of SIDD profiles in the promoter area of *E. coli* genome shows that regions of high melting probability do not contribute directly to polymerase-promoter recognition, but help to maintain the promoter in stable superhelical conditions.

Acknowledgements: This work was supported by RFBR grant r_centr_a 14-44-03679.

References:

1. Wang, H.Q. and Benham, C.J. (2006) Promoter prediction and annotation of microbial genomes based on DNA sequence and structural responses to superhelical stress, *BMC Bioinformatics*, 7: 248-262.

COMPUTER ANALYIS OF DISTAL GENE REGULATION USING CHROMOSOME CONTACTS DATA

Y.L. Orlov[1]*, E.V. Kulakova[2], A.G. Bogomolov[1], V.N. Babenko[1], G. Li[3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Huazhong Agricultural University, Wuhan, China*
* *Corresponding author: orlov@bionet.nsc.ru*

Key words: *gene expression regulation, chromosome contacts, CTCF sites, ChIA-PET*

Motivation and Aim: Transcription regulation research is important problem of molecular biology challenging sequencing technologies and bioinformatics data analysis. ChIP-Seq detects interactions between DNA and proteins; ChIA-PET (Chromatin Interaction Analysis with Paired-End-Tag) technology allows detect interactions between pairs of DNA sites affecting gene regulation. Fullwood et al. [1] used ChIA-PET technology to construct chromatin interaction network bound by estrogen receptor alpha from human breast cancer cell line (MCF-7) and found long-range ER binding sites are mostly located at promoter regions. CTCF-mediated interactions found in mouse embryonic pluripotent stem cells and human cell lines [2]. Li et al. [2] detected promoter-centered distant interactions bound by RNA Polymerase II in cancer cells.
Methods and Algorithms: We developed computer programs for statistical data analysis and test it on CTCF binding sites, genes and spatial topological domains. These data have been obtained experimentally by using methods ChIP-seq, Hi-C, ChIA-PET. Five distinct chromatin domains revealed by CTCF ChIA-PET raised a new model of CTCF function for chromosome structure organization and linking enhancers to promoters for gene transcription regulation.
Results: We used data on the spatial domains in the genome of the mouse embryonic stem cells and in the human genome, data on the location of CTCF binding sites clusters obtained by ChIA-PET. Gene annotation was obtained from UCSC Genome Browser (http://genome.ucsc.edu). The result of the analysis is the distribution of CTCF transcription factor binding sites on domains on the human chromosomes and relative gene locations. The distributions of human genes relative CTCF binding sites and a randomly generated list of such sites as the program output were used to estimate statistical significance of the associations found.
Conclusion: Chromatin interaction network is organized into “community”, and genes within community perform related functions and respond to external stimuli in a coordinated manner. In all the promoter-nonpromoter interactions, more than 40% of the non-promoter regulatory elements didn’t interact with their nearest promoters.

References:

1. M.J. Fullwood et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome, *Nature*, 462(7269):58-64.
2. G. Li et al. (2014) Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application, *BMC Genomics*, 15(Suppl 12):S11.

DIOXIN-MEDIATED UPREGULATION OF ONCOSTATIN M IN U937 MACROPHAGES

D.Y. Oshchepkov[1]*, E.V. Kashina[1], V.A. Mordvinov[1], D.P. Furman[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: diman@bionet.nsc.ru*

Key words: *Oncostatin M, macrophage, dioxin, AhR, cytokines*

Motivation and Aim: The environmental pollutant 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD, dioxin) is the most toxic among the dioxin xenobiotics and induces a broad spectrum of biological effects, including immunotoxicity and cancer [1]. Macrophages are key regulators of the innate immune response, as well as one of the first types of cells responding to stress, so the study of dioxin action in these cells is important. It is known that TCDD exposure effects some cytokines expression [2] but our analysis [3] showed that the list of such cytokines is not yet completed. We have investigated an effect of TCDD on Oncostatin M (OSM) expression in U937 macrophages.
Methods and Algorithms: Real-time PCR experiments were performed to investigate OSM mRNA expression dynamics at 6 and 24 hours after TCDD exposure in U937 macrophage-like cells.
Results: The data obtained demonstrate that OSM is upregulated after 6h of TCDD exposure, and maintains its overexpression after 24 hours. Transcription factor AP-1 is known to be the activator of OSM expression in macrophages [4]. We have shown that *FOSB* and *FOSL2* genes, coding AP-1 subunits, are upregulated simultaneously with OSM in U937.
Conclusion: Oncostatin M is supposed to play fundamental roles in mechanisms of inflammation in pathology [5]. Predicted activation of Oncostatin M expression in monocytes/macrophages, which are a primary source of OSM, can explain a spectrum of biological effects of AhR ligands exposure, including immunotoxicity and cancer.
Acknowledgements: This work was supported by state projects nos 0324-2015-0003 and 0324-2015-0004.
References:
1. Mandal P. (2005) J Comp Physiol B. 175: 221-30.
2. Kerkvliet N. (2009) Biochem Pharmacol. 77(4): 746-60
3. Furman D., et al. (2009) Comput Biol Chem. 33(6):465-8.
4. Kastl SP, et al.,.Blood. 2009 Sep 24;114(13):2812-8.
5. Richards C.D. ISRN Inflamm. 2013 Dec 8;2013:512103.

IN SILICO MOUSE CHROMOCENTERS CONTENT

D.I. Ostromyshenskii*, A.S. Komissarov, I.S. Kuznetsova, O.I. Podgornaya
Institute of Cytology RAS, St. Petersburg, Russia
** Corresponding author: necroforus@gmail.com*

Key words: *sequencing, chromocenter*

Motivation and Aim: Chromocenters are interphase nuclei landmark structures of constitutive heterochromatin. Heterochromatin enriched in tandem repeat (TR). There is progress in recent years in revealing chromocenters protein content, though it is not clear what DNA underlay constitutive heterochromatin apart from TR. The aim of the current work was to find out what DNA sequences involved in chromocenters formation.
Methods and Algorithms: Chromocenters was biochemically isolated and chromocenters' DNA library prepared by Nextera kit. The library were sequenced using Illumina MiSeq with paired-end reads of 35 bp.
Results: Bioinformatics comparison of chromocenters MiSeq (chcMiSeq) with whole genome sequencing on Illumina HiSeq (gnHiSeq) revealed NNN content:

Family\Source	chcMiSeq	gnHiSeq
Tandem repeats (TR)	70%	10%
LINE	7%	4%
ERV	1.5%	1%

Among chcMiSeq TR the most abundant is MaSat (61%) and MiSat (4%). The rest of TR (5%) represents the TRs families previously described [1]. The rest 20% of chcMiSeq dataset is mostly unannotated sequences, but some of them have been identified when part of chcMi-Seq dataset have been assembled into contigs by IDBA_UD program. In the contigs assembled there are many fragments of heterochromatic Y chromosome, rRNA and six other pseudo-genes and ncRNA gene. Full scale gene *sfi1* is found in contigs and it is localized to the chromosome 11 pericentromeric region. The ERV based fragments from chcMiSeq assembled contigs went to all the possible locations being mapped to different ERV consensuses from Repbase. This indicate that the whole ERV could be built in TR arrays. In contrast, there is very few full-length LINEs in chcMiSeq or in its' part assembled. Most of the LINE fragments collected in the same 2 kb region at the end of the 2nd ORF and its' flanking region. The same region of LINE is the origin for the L1-based TR [1]. Full-length LINEs enrich facultative heterochromatin, but it is nearly absent in constitutive heterochromatin [2, 3]
Conclusion: The sequencing of chromocenters' DNA (chcMiSeq) reveal full length ERVs and precise LINE' fragment of 2 kb as the substantial mouse constitutive heterochromatin components together with TR of different families.
Acknowledgements: This work was supported by the granting program 'Molecular and cell biology' of the Russian Academy of Sciences.
References:
1. A.S. Komissarov, E.V. Gavrilova, S. Demin et al. (2011). Tandemly repeated DNA families in the mouse genome. *BMC genomics*, *12*(1):531-552.
2. A.T. Chinwalla et al (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915): 520-562.
3. I. Solovei, M. Kreysing, C. Lanctot et al. (2009). Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. Cell 137: 356-368.

THE GENOME WIDE ANALYSIS OF THE LARGE TANDEM REPEATS IN THE CLOSELY RELATED GENOMES

D.I. Ostromyshenskii*, O.I. Podgornaya
Institute of Cytology RAS, St. Petersburg, Russia
** Corresponding author: necroforus@gmail.com*

Key words: *genomes, tandem repeats*

Motivation and Aim: Large tandemly repeated sequences (TR, or satellite DNA) are necessary part of higher eukaryotes genomes and can comprise up to tens percent of the genomes. Much of TRs' functional nature in any genome remains enigmatic because there are only few tools available for dissecting and elucidating the TR functions.
Material and methods: The modified pipeline of the one used previously in our Lab [1] applied to the several databases. Four mammalian genera was used: (1) mice *Mus*: *M.musculus*, *M. caroli* (unassembled genome); (2) guinea pigs *Cavia*: *C. porcellus*, *C. apperea*; (3) bats *Myotis*: *M. brandii*, *M. davidii*, *M. lucifugus*; (4) cows *Bos*: *B. taurus*, *B. mutus*, *B. indicus*.
Results: We tried to find all the 62 M.musculus TR families [1] in raw reads of M. caroli genome (Caroli Genome Project, PRJEB2188). There are only few TR of M. musculus in M. caroli genome. M. musculus major satellite (MaSat or GSAT-MM) occupied nearly 0,7% of M. caroli genome, while in M. musculus genome - ~ 11 %. In M. caroli genome we found 5 other M. musculus's TR's families.
Genus *Cavia*. *C. porcellus* genome possesses 25 TR and *C. apperea* – only 10 TR. 9 out of 10 *C. apperea* TR's family exist also in C. *porcellus* genome except the major TR for this species – Capp-1518. In *C. porcellus* genome there are two major TR – Cpor-783 is absent in the 2nd genome and Cpor-123 exists in C. *apperea* genome as the minor one. Genus *Myotis*. There is no any TR of *Myotis* in Repbase, but 133 TR's families are found in *M. brandtii* genome, 105 - in *M. davidii* genome and 26 - in *M. lucifugus* genome. Only 5 TR families exist in three genome but most of TR families are species-specific. Major TR for *M. davidii* and *M. lucifugus* is common in sequence though differ in monomer length, but the same TR is minor one in *M. brandtii.* The major for *M. brandtii* is not identified in both other genomes at all.
Genus *Bos*. There are three TR known for *Bos* in Repbase and all of them are found in all *Bos* assemblies. Still the major TR in all Bos assemblies differ: in *B. taurus* genome BTSAT4/BTSAT5 is a major TR while BTSAT6 major TR family in *B. indicus* genome. It is visible that most of the top TR families in genus *Bos* exist only in two genomes or even in one, i.e. is species-specific.
Conclusion: The most exhausting analysis of major TR (one for each species) of ~300 animals and plants display no readily apparent conserved characteristics [2]. We compared the TR sets. Our data evidenced that there are species-specific top TR, which are absent in genome of closely related species. In all genera examined major TRs are species-specific and hardly exist in other species of genera even as a minor ones.
Acknowledgements: This work was supported by the granting program 'Molecular and cell biology' of the Russian Academy of Sciences.
References:
1. A.S. Komissarov, E.V. Gavrilova, S. Demin et al. (2011). Tandemly repeated DNA families in the mouse genome. *BMC genomics*, *12*(1):531-552.
2. D.P. Melters, K.R. Bradnam, H.A. Young, et al (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol*, *14*(1), R10.

ELECTROSTATICS: A NEW OLD GENOME SELECTION FACTOR

A.A. Osypov*
Institute of Higher Nervous Activity and Neurophysiology RAS, Moscow, Russia,
Institute of Cell Biophysics of RAS, Pushchino, Russia
** Corresponding author: aosypov@gmail.com*

Key words: *DNA electrostatics, transcription regulation, genome evolution*

Motivation and Aim: Genome DNA electrostatics (E.) properties influence its interactions with proteins, esp. for transcription regulation (TR). DEPPDB was developed to provide all information on genome DNA physical properties, sequence, and annotation of biological properties of genome elements and whole genomes, organized on taxonomical basis.
Methods and Algorithms: DEPPDB and its tools [1, 2] were used to carry out the analysis.
Results: E. potential (EP) is distributed non-uniformly along DNA and correlates with GC content, strongly depending on the sequence arrangement and its context (flanking regions). Observed RNA polymerase binding frequency to DNA correlates to calculated EP. TR areas have EP peculiarities. Binding sites of transcription factors (TF) of different protein families in different taxa are located in long areas of high EP. EP distribution on TF protein surface reflects that of binding sites. Promoters in average have high value and heterogeneity of EP profile. Transcription starting sites of prokaryotic genomes are characterized by hundreds of bp of high EP and some peculiarities directly around TSS. This is aimed to protein binding and physical properties formation for transcription machinery. TSS EP architecture is similar in related taxa. Promoters up-element demonstrates electrostatic nature. E. interact in formation and TR with other physical properties: bending, thermal stability, supercoiling. Curved DNA in promoter regions is preserved and determined by habitat temperature. Mesophiles have different intensity in curvature; (hyper)thermophiles lack it due to life in temperatures above the curvature-relaxing point, rendering it useless in TR. Strongly curved DNA fragments must possess high A+T content (reverse is not true). There is no decrease in size and prominence of electrostatic deep in extremophyles, proving importance of E. and its differential role *vs* curvature. EP properties of intracellular parasite *M. leprae* reflect pseudogenization with reduced TR.
Conclusion: E. plays universal role in prokaryotic TR, affecting proteins binding. It may influence horizontal gene transfer, TR systems evolution and contribute to genome regulatory areas high AT content in such diverse domains as Bacteria and Archea. Physical properties affect such fundamental problems as Chargaff's II rule, genetic code redundancy, nonsynonymy of synonymous substitutions *etc.*, approving biophysical bioinformatics.
Availability: DEPPDB is available at http://deppdb.psn.ru or http://electrodna.psn.ru
Acknowledgements: RFBR grants 14-44-03683 and 16-04-01865.

References:

1. A.A. Osypov et al (2010) DEPPDB – DNA Electrostatic Potential Properties Database. Electrostatic Properties of Genome DNA, *JBCB,* 8(3): 413-25
2. A.A. Osypov et al (2012) DEPPDB – DNA Electrostatic Potential Properties Database. Electrostatic Properties of Genome DNA elements, *JBCB*, 10(2) 1241004.

DIFFERENTIAL EXPRESSION IN HELIX LUCORUM STATOCYSTS UNDER MICROGRAVITY CONDITIONS

A.A. Osypov[1, 2]*, P. Kolosov[1], N. Aceyev[1], E. Chesnokova[1], M. Roshchin[1], N. Bal[1], P. Balaban[1]

[1] *Institute of Higher Nervous Activity and Neurophysiology of RAS, Moscow, Russia,*

[2] *Institute of Cell Biophysics of RAS, Pushchino, Russia*

** Corresponding author: aosypov@gmail.com*

Key words: *Helix lucorum, statocysts, gravity reception, transcriptomics, differential expression*

Motivation and Aim: Helix lucorum snail is a classical model object for studies of the nervous system functions. In order to understand the genome mechanisms of the gravity reception in the snail nervous system we performed a near single-cell transcriptomics analysis of space flight induced differential expression in *Helix* statocysts.

Methods and Algorithms: There were 8 animals in two equal groups of snails - that flied into space (n = 4) and remained on Earth (n = 4), some 13 cells were used per every sample. We performed a full novel transcriptome assembly based on the total mRNA sequenced by means of Ion Proton System.

Results: Near 60% of reads per sample were mapped to the assembly, yielding more than 40 significantly differentially expressed (i.e. downregulated by space flight) genes of high accuracy. Most of them relate to the cell reception and different stages of intracellular signaling pathways, including gene expression regulation. Interestingly they were no significantly differently expressed genes if transcripts were mapped to the whole nervous system transcriptome assembly provided by our colleagues, and the overall portion of the mapped reads was less by nearly 20%.

Conclusion: The data obtained indicates that genes that are differently expressed are specific to the statocysts themselves and are probably related to gravity reception.

Acknowledgements: The work was supported by RSF grant 14-25 00072

TOWARDS A NEUROBIOLOGICALLY REASONABLE *C. ELEGANS* NERVOUS SYSTEM SIMULATION: NEURON, MUSCLE AND SIGNAL PROPAGATION MODELLING

A.Yu. Palyanov[1, 2]*, Kh.V. Samoilova[2]
[1] *A.P. Ershov Institute of Informatics Systems, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *Corresponding author palyanov@iis.nsk.su*

Key words: *C. elegans, nervous system, simulation, ion channels, computational modeling*

Motivation and Aim: C. elegans with its small 302 neurons nervous system and nearly determined connectome is considered as a first multicellular organism to be reverse-engineered and reproduced in the form of a computer simulation. *C. elegans* has a specific mechanism of neural signal transmission based on passive propagation since it lacks voltage-gated Na^+ channels required for typical action potentials. Morphology and electrophysiology of *C. elegans* neurons allows such signals to travel up to 1 mm or longer distances, although with noticeable delay and fading [1]. Body wall and pharyngeal muscle cells are capable of generating specific calcium-dependent action potentials that are driven by the L- and T-type voltage-gated Ca^{2+} and voltage-gated K^+ channels [2, 3]. Reproduction of at least these mechanisms, including ion channels, neurophysiologic parameters and morphology, is required for a basic neurobiologically reasonable simulation which can be further extended with more sophisticated mechanisms.
Methods and Algorithms: In this work we used the NEURON simulation environment, which is particularly well-suited to problems that are closely linked to experimental data, especially those that involve cells with complex anatomical and biophysical properties [4]. It provides an ability for construction of custom models of ion channels and other cellular mechanisms via Neuron Model Description Language (NMODL) expanding the standard repertoire, which we employed as well.
Results: Using the NEURON programming language HOC and NMODL we have developed a model of a typical *C. elegans* neuron with parameters known from experimental study, which works in good accordance with our calculations [1]. The model of *C. elegans* pharyngeal muscle with the models of key ion channels was also created and reproduces its time dependence of membrane potential quite well.
Conclusion: Proposed models of *C. elegans* neuron and muscle provide the basis for further development and construction of neural circuits.
Acknowledgements: The work was supported by Russian Federation President grant MK-5714.2015.9
References:
1. A.Yu. Palyanov, A.S. Ratushnyak (2015) Some Details of Signal Propagation in the Nervous System of C. elegans. Russian Journal of Genetics: Applied Research, 5(6): 642-649.
2. S. Gao and M. Zhen. (2011) Action potentials drive body wall muscle contractions in C. elegans. Proc. Natl. Acad. Sci. USA. 108 (6): 2557–2562.
3. B. Shtonda, L. Avery. (2005) CCA-1, EGL-19 and EXP-2 currents shape action potentials in the C. elegans pharynx. J. Exp. Biol. 208: 2177–2190.
4. N.T. Carnevale and M.L. Hines (2006) *The NEURON Book,* Cambridge, UK: Cambridge University Press.

SEARCH FOR FUNCTIONAL NF-KB BINDING SITES VIA META-ANALYSIS OF NGS EXPERIMENTS IN HUMAN CELL LINES

N. Panyushev[1, 2], E. Lomert[1]*, D. Tentler[1], A. Predeus[2]
[1] *Institute of Cytology RAS, St. Petersburg, Russia*
[2] *Bioinformatics institute, St. Petersburg, Russia*
** Corresponding author: predeus@gmail.com*

Key words: *transcription factors, ChIP-seq*

The NF-kB family of transcription factors plays the critical role in inflammation, immunity, cell proliferation, differentiation and metastasis. NF-kB dimers recognize 9-11 nucleotide sequences called kB sites. There are about 13600 human genes with kB sites in promoter region; however not all of them are transcriptionally active. Thus search for functional kB sites could assist in understanding the basic principles that underlie NF-kB regulation.

The combination of chromatin immunoprecipitation (ChIP) and next-generation sequencing (NGS), namely ChIP-seq, has become a powerful technique to capture potential genomic binding sites of transcription factors, histone modifications and chromatin accessible regions. Using ChIP-seq datasets deposited in the public databases, such as GEO and ENCODE we revealed physical binding sites of p65 (RelA) NF-kB subunit in various human cell lines: MCF-7, HUVEC, HeLa, SGBS and A549. For this purpose, we compared datasets for untreated cells and TNF-alpha-stimulated cells which contain activated p65 (RelA) protein.

The effect of binding of NF-kB with predicted kB-sites on gene transcription in cell lines listed above was confirmed by RNA-seq data. The analysis of accessibility of genomic regions to NF-kB binding was carried out using epigenetic ChIP-seq data. Regulatory elements of the genes were distinguished by the histone modifications (H3K4me1, H3K-4me3, H3K27ac and H3K36me3).

The analysis revealed a consistent pattern of the regulation of NF-kB-dependent transcription, the correlations of histone modifications and localization and consensus sequence of kB-sites. Regulatory elements (promoters and enhancers) were identified genes via the chromatin context information. Obtained data can be used for experimental validation of NF-kb -dependent regulation mechanisms by the binding kB-sites in the regulatory regions.

This work is supported by the Russian Science Fund grant (14-50- 00068).

NEW INSIGHTS INTO THE REGULATION OF REACTIVE OXYGEN SPECIES BY AUXIN THROUGH GENE EXPRESSION ANALYSIS

I.A. Paponov[1, 2]*, V. Budnyk[1], T. Khodus[1], M. Paponov[1], K. Palme[1]
[1] *Institute of Biology II/Molecular Plant Physiology, Faculty of Biology, Albert-Ludwigs-University of Freiburg, Germany*
[2] *Norwegian Institute of Bioeconomy Research, Norway*
** Corresponding author: ivan.paponov@bioforsk.no*

Key works: *auxin, reactive oxygen species, ROS, plant growth, environmental stress*

Motivation and Aim: Plants, unlike animals, cannot move away from environmental stresses. Instead, they capitalise on a wide-ranging potential for growth adjustment in response to stresses. One mechanism believed to underlie this wide-range adjustment is cross-talk between reactive oxygen species (ROS) and the plant hormone auxin. ROS are well known and extensively investigated regulators of auxin activity, acting via oxidation, signalling and distribution. However, the manner by which auxin affects ROS is less well understood.
Methods: We addressed this question by a combination of microscopy staining data analysis and transcriptomic analysis of gene expression data obtained from publically available experiments that used auxin and ROS treatments.
Results: Microscopy staining showed that auxin differentially regulates the level of reactive oxygen species in the roots: it increases hydrogen peroxide levels, as reflected by the fluorescence ratio of ratiometric redox-sensitive GFP (roGFP) and it decreases the cell wall levels of hydroxyl radicals ($OH^{\cdot}$). Gene expression analysis of the auxin regulation of all ROS-related genes showed that peroxidases are the best candidates for differential regulation of ROS. General oxidative response genes, which are induced by several ROS, were highly over-representative among the auxin responsive genes. This finding might reflect a combination of ROS induction by auxin and the presence of general oxidative response cis-elements in the promoters of auxin-responsive genes. Thirteen (13) auxin-repressed genes were up-regulated by superoxide, as indicated their up-regulation in a chloroplast superoxide mutant that displays enhanced chloroplastic superoxide levels. Interestingly, 11 of these 13 genes were expressed in chloroplasts, suggesting a novel regulation of chloroplast genes by auxin through specific reduction of the chloroplast superoxide level.
Conclusions: Differential induction of ROS by auxin and regulation of auxin activity by ROS results in a positive feedback loop that is pivotal in plant growth adjustment to environmental stress conditions.

IDENTIFICATION OF RECOMBINATION SITES IN THE GENOMES OF THE EUROPEAN SUBTYPE OF TICK BORNE ENCEPHALITIS VIRUS

A.I. Paramonov[1]*, Yu.P. Dzhioev[1, 2], I.V. Kozlova[1]
[1] FGBNU Scientific Center of family health and human reproduction problems, Irkutsk, Russia
[2] Irkutsk State Medical University, Institute of Biomedical Technology, Irkutsk, Russia
** Corresponding author: paramonov_a.i@mail.ru*

Key words: *recombination, tick borne encephalitis virus, TBEV*

Motivation and Aim: Tick borne encephalitis virus (TBEV) of the Flaviviridae family is the causative agent in human neuroinfections which often cause disability and death. There are three main subtypes - Far Eastern, Siberian and European (Western). Each subtype has its own habitat; European subtype TBEV has extensive undivided area in Western Europe and mosaic in Asia. Representatives of this serotype circulate in ecosystems significantly different by composition of biocenoses, vectors and hosts. Special attention should be paid to the genetic variation of this subtype, and genetic recombination is one of its leading factors [1, 2, 3]. The aim of this study was to detect potential recombination sites in the genomic sequences of the isolates of European subtype TBEV.
Methods and Algorithms: Genomes of 26 strains of European subtype TBEV available in the GenBank data base, as well as 8 strains we sequenced, were used in this work. The phylogenetic test for the presence of recombination was obtained using Splits Tree v4.1, by Neighbor-net method. Statistical test was carried out using the Phi Test for Recombinations method of Splits Tree software system [4]. Positioning of recombination sites was performed using software methods implemented in the programs package RDP v.4.46 [5].
Results: Phylogenetic network constructed by Neighbor-net has multiple splits indicating the possibility of reticulated evolution and accordingly recombination events at least in some strains. Phi Test for Recombinations showed the presence of recombination in this sequence set with p = 0,008. RDP v4.46 software package found recombination points in strains Joutseno and Absettarov. Interestingly, Joutseno strain contained two independent recombination points.
Acknowledgements: The work was supported by Russian Science Foundation grant №14-15-00615.
References:
1. Gritsun TS, Nuttall PA, Gould EA. Tick-borne flaviviruses //Adv Virus Res. 2003;61:317
2. I.V. Kozlova et al. Genetic and Biological Properties of Original TBEV Strains Group Circulating in Eastern Siberia //Monograph.«Encephalitis» Edited by S.Tkachev, 283 pages, Publisher: InTech, Published: January 09, 2013. P. 95-112.
3. Růžek D, Dobler G, Donoso Mantke O. Tick-borne encephalitis: pathogenesis and clinical implications // Travel Med Infect Dis. 2010 Jul;8(4):223-32.
4. Huson DH, Bryant D: Application of phylogenetic networks in evolutionary studies // Mol BiolEvol. 2006, 23.-P. 254-267.
5. Martin DP, Murrell B, Golden M, Khoosal A, & Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution 1: vev003 doi: 10.1093/ve/vev003

DARWINIAN GENETIC DRIFT

D.V. Parkhomchuk*, A.C. McHardy
Department for Computational Biology of Infection Research, Helmholtz Center for Infection Research, Braunschweig, Germany
** Corresponding author: parkhomc@googlemail.com*

Key words: *genetic drift, neutral evolution, population variability, Shannon's entropy*

Multicellular organisms (including humans) have mutations rates of order of at least few functional mutations per genome per generation. We claim that neither neutral ("non-Darwinian") drift nor Darwinian selection, via chain of fixations of positive variants are directly applicable to describe genetic variability in a population. The aim is to build a general formalism, which can account for drifting of arbitrarily functional variants explicitly, including high mutation rates of functional sites. Neutral drift is then a limiting case of such drift. If we let a "perfect" genome, with the best possible variants in all positions, evolve under the high functional mutation rate, it will "degrade" to some average equilibrium fitness, which is less than the fitness of the perfect genome. Then we can study the general properties of a population drifting around the average fitness in equilibrium. The resulting drift is "Darwinian", because we do not consider neutral variants at all. Instead of neutral evolution, we effectively consider the evolution by compensatory mutations (because the average fitness does not change in time). We can calculate the number of unique genomes, which have the equilibrium fitness. This (very large) fundamental number describes the population variability, defining population boundaries in a sequence space. In statistical mechanics the equivalent is a "number of microstates" and in Shannon's Information Theory this is a "typical set" size. Correspondingly, if we calculate per-site variability from this number, it turns out to be Shannon's entropy. The drift displays unusual characteristics, in comparison with models, which are not accounting for high mutation rates in functional sites: population fitness and evolvability is independent from population size; the fraction of positive mutations among random mutations can be high, in general (a trivial consequence of "compensatory" view on mutations accumulation); variants fixations in an equilibrium population play no role for fitness evolution, and can be averaged out. The differences from traditional views stem from the different and incompatible starting assumptions: we posit that neutrality is an unnecessary oversimplification and it is crucial to consider realistic mutation rates, where the selection force is properly balanced with an equal and opposite stochastic force (for example, but not exclusively, high mutagenesis). When this fundamental balance is not explicitly introduced in a model, such models are critically incomplete and can produce artifactual dependencies. The drift concept is at the core of many bioinformatic models: molecular clock, phylogenetics, coalescent theory and others. Our results suggest that considering the drift as being non-neutral and including compensatory mutations is more realistic and functionally informative in comparison with the neutral drift assumption. It opens a number of bioinformatic challenges, such as to decipher the underlying functionalities of variances in a population, to build consistent models of common polygenic diseases and predispositions, and to improve our understanding of genomic complexity evolution, in general.

References:

1. A.A. Shadrin, D.V. Parkhomchuk. (2014). Drake's rule as a consequence of approaching channel capacity. Naturwissenschaften 101, 939–954.

IMPACT OF 105-DAY ISOLATION CONDITIONS ON PROTEINS EXPRESSED IN ENDOTHELIAL CELLS, IN THE FRAMEWORK OF THE "MARS-500" PROJECT

L.Kh. Pastushkova[1], D.N. Kashirina[1], A.S. Kononikhin[1, 3], A.G. Brzhozovsky[1], I.V. Dobrokhotov[1], E.S. Tiys[2], V.A. Ivanisenko[2], E.N. Nikolaev[3], I.M. Larina[1]

[1] *State scientific center of Russian Federation – Institute for biomedicalproblems RAS, Moscow, Russia*

[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[3] *Emanuel Institute of Biochemical Physics RAS, Moscow, Russia*

Key words: *proteomics, health, space flight*

In order to determine the panel of proteins that may characterize the early stage of vascular endothelial dysfunction, which develops under conditions of physical inactivity in healthy individuals, liquid chromatography-mass spectrometry analysis of urine protein composition of six healthy volunteers was performed. The study was conducted during the experiment with 105-day isolation of healthy volunteers in onground experimental complex. The volunteers were lived inside air chamber in controlled conditions with different levels of salt consumption (from 6 to 12 g/day). Food nutrients intake (electrolytes, water, calories, fat, carbohydrates, protein, vitamins, etc.) on each stage of experiment was normalized per body weight. All obtained biological samples were analyzed using mass ion cyclotron resonance mass spectrometer with Fourier transformation LTQ FT MS (Thermo). After proteomic analysis various bioinformatics possibilities were used. Identification of proteins expressed in endothelial cells was performed by data base Bgee.

The total number of proteins identified in urine was 2037. Among them, it was found 164 proteins expressed in endothelial cells. Of these, 3 proteins, namely osteopontin, prostaglandin D sinthase, protectin CD59 were detected in the urine constantly. The protein angiotensinogen (RAAS member) was present in the urine samples of the background period, but ceased to be detected in the experiment and the recovery period. Then, the correlation between proteins detection frequency and the level of salt consumption was revealed. It has been shown that amyloid beta protein, endosialin, CD90 protein, H subunit of tetramer enzyme lactate dehydrogenase had a high and reliable level of linear correlation with the level of salt consumption. Manual annotation of the most significant proteins in terms of endothelial functions was performed. The data indicate a slight impact of 105-day isolation factors on endothelial function.

We wish to thank "Bruker Daltonics" for administrative and financial support. This work was supported by the Russian Academy of Sciences, Russian Foundation of Basic Research (Grants 15-04-02463 A).

THE ROLE OF KINETOCHORE-DRIVEN MICROTUBULE FORMATION IN DROSOPHILA SPINDLE ASSEMBLY

G. Pavlova[1, 2]*, J. Popova[1, 3]*, A. Munzarova[1, 4]*, J. Galimova[1]*, A. Razuvaeva[1, 4], F. Renda[5], P. Somma[5], A. Pindyurin[1, 4], M. Gatti[5]

[1] *Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russia*

[2] *Kazan Federal University, Kazan, Russia*

[3] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[4] *Novosibirsk State University, Novosibirsk, Russia*

[5] *Department of Biology and Biotechnology, Sapienza, University of Rome, Rome, Italy*

* equal contribution

Key words: *mitosis, modeling*

Mitotic catastrophe (MC) is a form of cell death caused by the disruption of the mitotic process. However, a clear definition of MC is still missing. The current view is that MC is not a specific mode of cell death, but rather a pre-stage that anticipates cell death through necrosis or apoptosis. It has been shown that this "pre-stage" is characterized by strong mitotic defects caused by either ionizing radiation or drugs that affect microtubule (MT) polymerization and/or dynamics. In addition, MC can be triggered by mutations in genes that disrupt fundamental steps of cell division. One this steps is kinetochore-driven MT formation (KDMF). We exploited the *Drosophila* system to investigate KDMF. In fly cells, KDMF is essential for spindle assembly, while MT nucleation from the centrosomes is dispensable, as cells devoid of centrosomes form functional anastral spindles. To genetically dissect KDMF we used *Drosophila* S2 cells and determined their proficiency in the process by analyzing spindle MT regrowth after cold- or colcemid-induced MT depolymerization. Specifically we examined KDMF in prometaphases/metaphases of cells depleted of specific spindle proteins by RNA interference (RNAi). We identified several factors that positively affect the process (Eb1, Mast/Orbit, Mars, Mei-38 and Dgt6), as well as factors that appear to delay KDMF (Asp and Patronin). These results provide novel insight into the molecular mechanisms of KDMF and suggest a model for the regulation of the process.

MICROBIAL COMMUNITY OF THE OIL SITE OF THE UZON CALDERA (KAMCHATKA)

S.E. Peltek[1], A.V. Bryanskaya[1]*, Y.E. Uvarova[1], A.S. Rozanov[1], T.V. Ivanisenko[1], T.K. Malup[1], V.A. Ivanisenko[1], E.V. Lazareva[2], O.V. Saik[1], S.M. Zhmodik[2], O.P. Taran[3], N.M. Slynko[1], S.V. Shekhovtsov[1], V.N. Parmon[3], N.L. Dobretsov[2], N.A. Kolchanov[1]
[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[2] V.S. Sobolev Institute of Geology and Mineralogy SB RAS, Novosibirsk, Russia
[3] Boreskov Institute of Catalysis SB RAS, Novosibirsk, Russia
** Corresponding author: alla@bionet.nsc.ru*

Key words: *microbial community, Uzon caldera, oil site, hydrocarbon metabolism*

Uzon Caldera is a unique place of the Earth where natural oil and bitumen yields (naphthides) were discovered within the active hydrothermal fields. Some studies showed that hydrothermal naphthides (hydrothermal petroleum) are modern formations. Oil of the Uzon caldera is considered as the youngest oil in the world. The preliminary results have indicated that their age is about 1,000 years, whereas the later studies have shown that they are only 50 years old.

Our field studies in the Uzon caldera during 2010-2012 in spring and autumn describe the features of the Oil site in the caldera. The Oil site is a small part of the thermal field with two natural thermal vents, with significant gradients of temperature, Eh-pH, and geochemical parameters. Oil film or oil droplets were formed on the surface of the solution when digging test pits in almost every location site. 10 test pits and hydrothermal vents have been examined for Oil field and adjacent areas.

More than 300 thousand sequences with the length of at least 250 base pairs were analyzed in the test samples by the high-throughput sequencing. We have determined more than 1 thousand of the individual taxons.

Detailed analysis of the microbial community composition showed that it contains microorganisms which are able to metabolize hydrocarbons. By correlation analysis in accordance with environmental factors and occupancy of metabolic pathways, the environmental factors were grouped into 4 major clusters and pathways into 7 largest clusters, respectively. It was shown that the cluster of metabolic pathways including Lipid metabolism correlates positively with the temperature and negatively with the Ba concentration. For the cluster including the Lysine degradation pathway, Other types of O-glycan biosynthesis, Carotenoid biosynthesis, strong positive correlation between biodiversity and a strong negative correlation with the elements Re, Co, U were found. Maximum negative correlation with biodiversity was observed for the filling the pathways of Fluorobenzoate degradation and Naphthalene degradation. Maximum filling of pathways was observed for metabolic pathways involved in the metabolism of petroleum products.

Thus, the Uzon caldera is a natural laboratory of modern petroleum formation from organic matter of sediments. The unique microbial communities were formed at high (up to 97 °C) and average temperatures, significant fluctuations in Eh-pH, and high content in solution and solid substance of sulphides, arsenic, antimony, and mercury. These communities are adapted to living in hydrocarbon surroundings, so they could be viewed as the source of unique enzymes and metabolites.

PQ: A NEW PROGRAM FOR PHYLOGENETIC RECONSTRUCTION

D. Penzar[1]*, M.S. Krivozubov[2], S.A. Spirin[3]
[1] *Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia*
[2] *Gamaleya Center of Epidemiology and Microbiology, Moscow, Russia*
[3] *Belozersky Institute of Physico-Chemical Biology of Moscow State University, and Institute of System Studies, Moscow, Russia*
** Corresponding author: dmitry_penzar_1996@mail.ru*

Key words: *Molecular phylogeny, algorithm, software, web interface*

Motivation and Aim. Programs for protein phylogenetic reconstruction are widely used by evolutionists and molecular biologists. A recent study [1] show that distance methods of phylogenetic reconstruction (such as minimum evolution) outperforms symbol-oriented ones (such as maximum likelihood). The aim of the work is to elaborate and implement a new symbol-oriented method that could compete with distance methods.
Methods and Algorithms. With a given protein multiple sequence alignment and a given tree on sequences of the alignment, the PQ (position-quartets) score is calculated by the following formulas: $W = \sum_{pq} \sum_{pq}$ where p runs over all positions of the alignment and q over all quartets (forths) of sequences of the alignment; $W_{pq} = \max(S(a_{ip}, a_{jp}) - M, 0) + \max(S(a_{lp}, a_{kp}) - M, 0)$, where $M = \max(S(a_{ip}, a_{kp}), S(a_{ip}, a_{lp}), S(a_{ajp}, a_{kp}), S(a_{jp}, a_{lp}))$, a_{ip} is the letter of i-th sequence in p-th position of the alignment, $S(a, b)$ is a scoring matrix (e.g., BLOSUM62), and i, j, k, l are sequences of the quartet q such that $\{i, j\}$ is separated from $\{k, l\}$ by at least one branch of the tree. The search for the tree with maximal W is performed with standard protocols: tree growing, nearest-neighbor interchange, and subtree pruning and regrafting.
Results. The program PQ is implemented and tested on a number of sets of alignments of protein evolutionary domains. PQ outperforms maximum likelihood and maximum parsimony programs for all tested sets. For some sets PQ outperforms distance-oriented programs, too. We also investigate effectiveness of different search strategies.
Conclusion. The new program PQ can be a good alternative to phylogenetic programs based on maximum likelihood and distance methods, especially for small (<30 sequences) protein alignments.
Availability. The program code in plain C is available on request from the authors. The web interface is available without registration at http://mouse.belozersky.msu.ru/tools/pq.
Acknowledgements: the work is supported by the Russian Scientific Foundation, grant No. 16-14-10319.
References:
1. G.Gonnet (2012) Surprising results on phylogenetic tree building methods based on molecular sequences, *BMC Computational Biology*, 13: 148

THE BIOINFORNATIONAL COMPARISON OF CRISPR/CAS SYSTEM STRUCTURE OF YERSINIA PSEUDOTUBERCULOSIS STRAINS ISOLATED FROM DIFFERENT REGIONS

N.P. Peretolchina[1]*, Y.P. Dzhioev[1, 2], A.Y. Borisenko[1], E.A. Voskresenskaya[3], A.I. Paramonov[2], L.A. Stepanenko[1], V.I. Zlobin[1]

[1] *Irkutsk State Medical University, Irkutsk, Russia*

[2] *Scientific Center of family health problems and human reproduction, Irkutsk, Russia*

[3] *Institut Pasteur, Saint Petersburg, Russia*

* *Corresponding author: nadine1lenz@gmail.com*

Key words: *CRISPR, Yersinia pseudotuberculosis, bioinformatics*

Motivation and Aim. CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated proteins) system is a specific and adaptive protection of bacteria against alien nucleic acids. CRISPR locus usually consists of cas-genes, leader sequence and spacers interspaced by short palindromic repeats. New spacer is incorporated at the beginning of CRISPR locus from bacteriophages and plasmids when bacterium meets them. Furthermore, bacterium with such spacers could be tolerant to phage or plasmid invasion. The aim of the research is a comparison of CRISPR/Cas systems of Yersinia pseudotuberculosis IP32953 and IP31758.

Materials and Methods. Objects of research were complete genome sequences of Y. pseudotuberculosis IP32953 and Y. pseudotuberculosis IP31758 presented in GenBank: CP000720 and CP009712. Structural and functional characteristics of cas-genes were studied by MacSyFinder (ver. 1.0.2) and accessory programs (makeblastdb ver. 2.2.28 and HMMER ver. 3.0). Three algorithms of CRISPR locus identification were used for more precise and accurate CRISPR-locus interpretation. Phages and plasmids screening was carried out by online-tool «CRISPR Target: a tool to explore targets of CRISPRRNAs», phage and plasmid database ACLAME, and BLAST. «PHAST: A Fast Phage Search Tool» was used for prophage region search.

Results. Y. pseudotuberculosis IP32953 and IP31758, both have CRISPR/Cas system referred to the type IF. Systems consist of two loci with different sizes. YP1 locus of Y. pseudotuberculosis IP31758 is specific for infectious agent circulation area. At the same time spacers of YP1 locus of Y. pseudotuberculosis IP32953 and spacers of YP3 loci of both strains are found in different strains of Y. pseudotuberculosis circulated in different continents. Y. pseudotuberculosis IP32953 interacted with Escherichia phage, Salmonella phage and Enterobacteria phage. Also this strain has protospacer in its own plasmid. Y. pseudotuberculosis IP31758 interacted with Vibrio phage, Enterobacteria phage and Yersinia phage during strain evolution. Also the strain has protospacer in transfer plasmid of Y. frederikssenii Y225.

Conclusion. Both strains have active CRISPR/Cas system. CRISPR/Cas systems of Y. pseudotuberculosis IP32953 and IP31758 have different origin. They do not have similar spacers. Strains could be resistant to phages action or to new plasmid acquiring. The results of the research could be used in developing of new typing methods of Yersinia based on the CRISPR system, and in selection of specific bacteriophages for target strain treatment.

SEARCH OF GENETIC SEQUENCES OF POTATOES IN DATABASES

A.I. Perfileva
Siberian Institute of Plant Physiology and Biochemistry, SB RAS, Irkutsk, Russia
** Corresponding author: alla.light@mail.ru*

Key words: *databases, potatoes, NCBI, Spud DB*

The potato genome is interesting object of research because potatoes is impotent for agriculture. However, now there is an insufficient numgunber of the works devoted to research of a genome of potatoes. The problem consists in complexity of definition of the necessary gene of potatoes because of a large number of databases, and also ambiguity of information provided in them, sequence considerably differ. The most widespread databases, containing gene sequences of potatoes are: GenBank (http://www.ncbi.nlm.nih.gov/genbank/), UniProt (http://www.uniprot.org/), Spud DB (http://solanaceae.plantbiology.msu.edu/). Most of scientists-biologists actively uses the BLAST NCBI base, however, it contains information on a wide range of biological objects of various kingdoms therefore not all genes of potatoes are present at her. Some works refer to the UniProt base, however, of information in her less, than in BLAST NCBI.

Most of researchers in the works refer to the Spud DB base (http://solanaceae.plantbiology.msu.edu/). The real base is characterized by what contains data not only on potatoes genes, and also on tomato genes. All base consists of these two research groups which have carried out a genome sequencing. One group of researchers has carried out a sequencing of a genome of a tomato and potatoes (designation of a gene begins "S. tuberosum group Phureja DM1-3-ITAG"), and other group potatoes sequencing (designation of a gene begins "S. tuberosum group Phureja DM1-3-PGSC"). The base contains information on a gene, his transcript, proteinaceous sequence, and also allows to carry out search of homologous genes.

Potatoes belong to family Solanacea therefore for search of genes of potatoes it is possible to use homologous genes of a tomato, and also other organisms. More truly to pick up the necessary sequence of a gene it is possible to bring sequence of the interesting gene in a search box of the Spud DB base and to find a potato homolog (with the indication of a protsentn of identity). Further the found sequence can be checked in Blast NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank/), at her existence in this base.

Distinctions in the found genetic povtornost can be connected with high-quality features, various techniques of research groups. However the main part of a gene at various sequencings is identical that allows to carry out work with these sequences.

The work has been conducted with the support of Russian Foundation for Basic Research, Grant mol_a No. 16-34-00806.

HEAT SHOCK PROTEINS OF POTATO *IN VITRO* UNDER HEAT AND BIOTIC STRESS

A.I. Perfileva*, E.G. Rikhvanov
Siberian Institute of Plant Physiology and Biochemistry, SB RAS, Irkutsk, Russia
** Corresponding author: alla.light@mail.ru*

Key words: *heat shock proteins, Clavibacter michiganensis spp. sepedonicus, stress*

Motivation and Aim. Under soft heat shock synthesized are heat shock proteins (HSP), which protect the cell from cell death under stronger heat shock. It is known that in a number of cases HSP is individuated in case of introduction of pathogens in plants. At the same time, there is no information in the literature on both the role of HSP in the response reaction of potato to its infecting with the ring rot pathogen *Clavibacter michiganensis* spp. *sepedonicus (Cms)* and the HSP content potato tissues under hyperthermia. In our work, investigated was the effect of infecting potato Cms and their processing with iodide acetic sodium salt (IASS) on the content of HSP in them.
Methods and Algorithms. To the end of defining the influence of pathogenesis on the content of HSP, the plants of resistant (Lugovskoy) and susceptible (Lukyanovsky) potato varieties were infected with Cms, underwent heat shock, and the quantity of HSP 101, HSP60 and HSP17,6 was measured. For the purpose of understanding of the role mitochondria in the process of HSP synthesis regulation, in a series of the following experiments we investigated the influence of processing IASS on the content of HSP in potato plants *in vitro*. IASS specifically inhibits of glyceraldehyde 3-phosphate dehydrogenase (the key enzyme of glycolysis), therefore, sodium salt irreversibly suppresses cellular respiration. The plants were infected with *Cms*; after 48 h of cultivating, these were processed with IASS; nest, the plants underwent heat shock. After all these processes, the content of HSP was determined by the Western-Blotting.
Results. It has been experimentally ascertained that the maximum content of HSP is observed in case of heating the plants at 39 °C. So, the temperature of 39 °C is seen as tempering for potato plants. No difference has been revealed in the level of synthesis of this protein at 26 °C for these varieties. Consequently, stability of the Lugovskoy variety to infecting with *Cms* is not bound up with synthesis of HSP. The content of HSP60 did not change in all the samples. This may be considered to be natural, no wonder that this protein has been attributed to the group of "house keeping proteins". As it has been verified experimentally, heat stress leads to the increase in the content of HSP17.6 and to substantial increase in content of HSP101 in tissues of both varieties of potato. Infecting in our experiments induced some increase in the content of HSP101 and HSP17.6 for the susceptible Lukyanovsky variety and some decrease in the content of these proteins in tissues of resistant Lugovskoy variety. Processing with IASS provoked a decrease in the content of HSP in potato tissues of both the varieties indicated under all the variants of processing the plants with IASS. Probably, this is a result of suppressing the functioning of mitochondria with IASS.
Conclusion. Probably, there are differently directed protective programs realized in plants under biotic and abiotic stresses. It may be supposed that HSP gene expression activation under heat stress is normally accompanied by suppression of expression of the genes protecting the cell under pathogenesis. So, decrease in the synthesis of HSP in infected plants under heat stress may be explained via the hypothesis stating that increase in the temperature suppresses expression of protective potato genes activated in response to biological stress. The work has been conducted with the support of Russian Foundation for Basic Research, Grant mol_a No. 16-34-00806.

MOSAIC GENE NETWORK MODELLING IDENTIFIED NEW REGULATORY MECHANISMS IN HCV INFECTION

O.V. Petroskaya[1]*, E.D. Petrovskiy[1, 2], E.L. Mishchenko[1], I.N. Lavrik[1, 4], V.A. Ivanisenko[1, 2]

[1] *The Federal Research Center Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *PB-soft, LLC, Novosibirsk, Novosibirsk, Russia*

[3] *International Tomography Center SB RAS, Novosibirsk, Russia*

[4] *Otto von Guericke University Magdeburg, Medical Faculty, Department Translational Inflammation Research, Magdeburg, Germany*

** The corresponding author: popik-olga@bionet.nsc.ru*

Key words: *Mosaic gene network, Mathematical modeling, Apoptosis, Tumor necrosis factor, Hepatitis C virus*

Motivation and Aim: Modeling of gene networks is widely used in systems biology to study the functioning of molecular genetic systems, including pathogen-host interactions. Most of the existing mathematical modeling techniques are useful for the analysis of the well-studied biological processes, for which there is complete information about rates of reactions. However, the complex biological processes corresponding to, for example, phenotypic traits of an organism, or pathological disease processes, including pathogen-host interaction, usually involves a set of interacting networks that can be represented by a mosaic gene network. In such networks, for each mosaic fragment there is a mathematical model, however, information about the molecular and genetic relationships between these fragments may be missing.

Methods and Algorithms: The approach of mosaic gene networks modeling in case of absence of precise information about the molecular and genetic links between the mosaic pieces is based on the genetic screening data on the effect of the elements perturbation of one piece and response of elements in remaining piece. The method uses approaches of control theory and applies mathematical models, written in the form of a system of ordinary differential equations (ODE).

Results: Modeling of random mosaic gene regulatory networks, consisting of two pieces, has shown high efficiency of developed method (mean deviation of the dynamics of mosaic networks elements from behavior of the original gene networks has been less than 10%). Using this approach, we have been integrated the model of hepatitis C virus subgenomic replicon replication and the model of apoptosis in cells. Such mosaic network allowed us to describe the regulation of apoptosis process by HCV. Analysis of the mosaic model revealed that the regulation of TNF-induced signaling by the HCV network is crucially dependent on the RIP1, TRADD, TRAF2, FADD, IKK, IκBα, c-FLIP, and BAR genes.

Conclusion: Overall, the developed mosaic gene network modelling approach demonstrated good predictive power and allowed the prediction of new regulatory nodes in HCV action on apoptosis and the NF-κB pathway. Those theoretical predictions could be a basis for further experimental verification.

GUT MICROBIOTA IN CASE OF PARKINSON'S DISEASE AND OTHER NEUROLOGICAL PATHOLOGIES: COMPARATIVE STUDY

V.A. Petrov[1]*, V.M. Alifirova[1], I.V. Saltykova[1], Y.B. Dorofeyeva[1], A.V. Tyakht[2], E.S. Kostryukova[2], A.E. Sazonov[1, 3]

[1] *Siberian State Medical University, Tomsk, Russia*

[2] *Scientific Research Institute for Physical-Chemical Medicine, Moscow, Russia*

[3] *Lomonosov Moscow State University, Moskow, Russia*

* *Corresponding author: vyacheslav.a.petrov@mail.ru*

Key words: *gut microbiota, metagenome, Parkinson's disease, 16S sequencing*

Motivation and Aim: Recently it is shown that nervous system of host organism interact with gut microbiota. In case of neurological disorders, especially for Parkinson's disease gut microbiota altered1. But little known about differences in microbiota composition among different neurological pathologies.

Methods and Algorithms: The study was conducted in a three groups of patients (60 control subjects, 60 subjects with Parkinson's disease and 32 subjects with other neurological pathologies, including multiple sclerosis, idiopathic dystonia and essential tremor). After the total DNA isolation and library preparation sequencing of the variable V3–V4 16S rRNA gene regions was performed by using MiSeq Reagent Kit v2 and MiSDefault device according to the manufacturer's recommendations. Taxonomic classification, alpha- and beta-diversity performed using QIIME Software 1.9.0. Determination of operation taxonomic units (OTU) performed with the usage of Greengenes v. 13.5 database (OTU's representative set picking) and HITdb (taxonomy assigning using RDP Classifier). Statistical comparison of the groups of samples was performed using Galaxy-based LefSe algorithm.

Results: The overall composition of fecal microbiota was affected by disease status both in terms of α- (chao1, Shannon and observed OTUs indices) and β- (weighted UniFrac) diversity. Gut microbiota of patients with Parkinson's disease and other neurological pathologies is characterized by lower taxonomic diversity in comparison with healthy control without significant difference between Parkinson's disease and other neurological pathologies. In addition, there were significant differences in compositional dissimilarity between groups based upon ANOSIM and MRPP statistical algorithms. According to LefSe algorithm, there were 44 species and 22 genera of bacteria and archaea with difference in abundance within groups. The top bacterial markers with highest LDA scores among the groups were Christensenella minuta on the species level and Bifidobacterium on the genus level for Parkinson's disease group, Bacteroides massiliensis on the species level and Bacteroides on the genus level for control group, Anoxystipes fissicatena and Blautia for other neurological pathologies group.

Conclusion: Gut microbiota composition is specifically altered in case of different neurological disorders.

Acknowledgements Ministry of Education and Science of the Russian Federation supported the work (RFMEFI60414X0075, № 14.604.21.0075).

1. F. Scheperjans et al. (2015). Gut microbiota are related to Parkinson's disease and clinical phenotype// Mov Disord, 30(3):350-8.

WORKFLOW FOR EXOME SEQUENCING IN IDENTIFICATION OF DE NOVO MUTATION IN THE *NCL6* GENE

D.A. Petukhova*, N.R. Maksimova, P.I. Guryeva, V.S. Kaymonov, M.T. Savvina
Laboratory of Genome Medicine, Clinics of Medical Institute, North-Eastern Federal University
** Corresponding author: todanilovadiana@gmail.com*

Key words: *Next-generation sequencing, whole-exome sequencing, de novo mutation, neuronal ceroid lipofuscinosis type 6 (NCL6), trio family analysis*

Whole-exome sequencing using next-generation sequencing (NGS) technologies is gaining popularity in clinical practice. Associations between undescribed mutation with the clinical picture of disease is the true difficulty in analyzing of NGS data. Trio family analysis (father, mother and affected child) is a very powerful approach to identify potentially pathogenic 'de novo' mutations in the proband. We sequenced the exome (~552 genes) of affected child with suspected of leukodystrophy and both unaffected parents by using the TruSight Inherited Disease sequencing panel (Illumina inc., San Diego, CA, USA) on a Miseq sequencing system (Illumina inc., San Diego, CA, USA). In total, we obtained 7 Gb of sequence data with 2091 variations from the human reference genome sequence that were subjected to several filtering steps. The MiSeq system provides fully integrated on-instrument data analysis software. Basespace software performs secondary analyses on the base calls during the sequencing run. SNP's and short INDEL's are identified using the Genome Analysis Toolkit (GATK) by default. The number of candidate variants is reduced using a three-step filtration strategy to generate a short candidate mutation list. For the variant filtering and annotation we used Variant Studio version 2.2 (Illumina inc., San Diego, CA, USA) data analysis software. Initial quality filter removed less reliable variant calls and resulted in the identification of 1499 genetic variants. In the second step given the rare incidence of autosomal – recessive disease, we excluded known dbSNP variants from our variants database, reducing the number of candidate by more than 98% to a total of 37 variants. For further analysis, we applied an autosomal-recessive disease model and assumed that the mutation was inherited from both parents [1,2,3]. Therefore, we have customized trio family data filtering and have found a common homozygous mutation c.396dupT (p.Val133fs) in exon 4 of *CLN6*. Validation by Sanger sequencing also confirmed that the c.396dupT (p.Val133fs) mutation was indeed present in a homozygous state in the affected child and in a heterozygous state in the both parents. We identified a pathogenic de novo mutation c.396dupT (p.Val133fs) in the *CLN6* gene, and made a diagnosis of neuronal ceroid lipofuscinosis type 6, suggesting that relatively high number of patients with neuronal ceroid lipofuscinosis type 6 may be hidden under the guise of leukodystrophy in Yakut population.

References:

1. Yohe S, Hauge A, Bunjer K, Kemmer T, Bower M, Schomaker M, et al. Clinical validation of targeted next-generation sequencing for inherited disorders. Arch Pathol Lab Med. 2015;139:204–10. doi: 10.5858/arpa.2013-0625-OA.
2. Becker J, Semler O, Gilissen C, Li Y, Bolz HJ, Giunta C, et al. Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. American journal of human genetics. 2011;88(3):362–71. doi: 10.1016/j.ajhg.2011.01.015 pmid:21353196; PubMed Central PMCID: PMC3059418.
3. Kmoch S, Stranecky V, Emes RD, Mitchison HM. Bioinformatic perspectives in the neuronal ceroid lipofuscinoses. Biochim Biophys Acta. 2012 S0925-4439(12)00301-8 [pii] 10.1016/j.bbadis.2012.12.010.

MIRNA BINDING SITES IN THE MRNA OF HUMAN TITIN GENE

I.V. Pinsky[1]*, A.T. Ivashchenko[1], S.B. Labeit[2]
[1] *Al-Farabi Kazakh National University, Almaty, Kazakhstan*
[2] *Institute of Integrative Pathophysiplogy, Mannheim, Germany*
**Corresponding author: ilya.pinskyi@mail.ru*

Keywords: *titin, sarcomeres, isoform, miRNAs, mRNAs, CDS*

Motivation and Aim: Titin, the human muscle protein, is the largest in the nature (the longest isoform IC contains 35991 amino acid residues) and plays an important role in providing the elasticity and structural integrity of sarcomeres. Interruption of its synthesis leads to the development of a number of serious cardiovascular diseases such as heart failure, cardiomyopathy, ischemic heart disease, and myocardial infarction. Titin gene expression is controlled by miRNAs (microRNAs) that bind with the mRNAs of the gene and block their translation. Therefore, it is important to determine which miRNAs most strongly regulate the synthesis of human titin and what exons of the gene contain the binding sites because different exons of the gene are expressed in different types of muscle tissue at different stages of the human body development.
Methods and Algorithms: The binding of 2563 human microRNAs with mRNA of human titin IC isoform, including all 363 exons of the human titin gene was determined using program miRTarget. The human miRNA sequences were taken from miRBase site (*www.mirbase.org/*), and the mRNA sequence of the titin gene was taken from Genbank (*www.ncbi.nlm.nih.gov/genbank*). The degree of binding ($\Delta G/\Delta G_m$, %) was estimated according to the value of the $\Delta G/\Delta G_m$ ratio, where ΔG was equal to the free energy of miRNA-mRNA binding and ΔG_m was equal to the energy of miRNA binding with its perfect complementary nucleotide sequence.
Results: As a result of this research, 15 miRNA binding sites with scores not less than 90% were found and marked in exons of titin mRNA. miR-6861-5p has the largest number of binding sites. This miRNA bound with the mRNA of titin at positions 37324, 38077, and 38830 nt at the boundaries of exons 178-179, 187-188, and 196-197, respectively. Other miRNAs had only one binding site each. miR-494-5p bound with titin mRNA at position 1301 nt in the seventh exon. miR-578 bound with titin mRNA in the eleventh exon at position 1960 nt. The 58th exon contained overlapping binding sites of two miRNAs (miR-374b-3p and miR-374c-3p) in positions 17239 nt and 17241 nt, respectively. Exon 59 was a target for miR-3714, which interacted with the mRNA at position 17450 nt. Exon 75 was the target of miR-34a-3p, which interacted with the mRNA at position 22116 nt. The 85th exon of the titin gene was the target for miR-1278, which interacted with titin mRNA at the position of 24928 nt. The 89th exon had a binding site for miR-544b at position 26044 nt. The 326th exon contained binding sites for miR-4738-3p and miR-136-3p. miR-4738-3p interacted with titin mRNA at position 74955 nt and miR-136-3p bound with mRNA of titin at position 71469 nt. Exon 339 also contained binding sites for miR-4693-5p and miR-4495, which bound with the mRNA for titin at positions 92464 nt and 93909 nt, respectively.
Conclusion: The results of the computer analysis provide a theoretical basis for further experiments to validate the miRNA binding sites found in the mRNA of titin and to determine the miRNA concentrations in the blood and other cells and tissues of humans and mice. These results could then be used for diagnosis and treatment of human cardiovascular diseases.

BIOSTORE: A CLOUD-COMPATIBLE HUB FOR BIOINFORMATICS RELATED TOOLS AND PLATFORMS

S. Pintus[1]*, T. Valeev[1, 2], I. Yevshin[1, 2], F. Kolpakov[1, 2]
[1] *Institute of Systems Biology Ltd., Novosibirsk, Russia*
[2] *Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
** Corresponding author: sspintus@biouml.org*

Key words: *containerization, Docker, cloud, workflow, web-server*

Motivation and Aim: The growing demand of flexible access and control of computational resources and big data in bioinformatics have recently resulted in development of a wide range of cloud-based approaches targeted to bioinformatics, from cloud-compatible pipelines like GATK and workflow managers like Galaxy to large-scale bioinformatics oriented cloud services like Google Genomics.
Today there is high demand for web based scalable platforms which would allow biomedical and computations researchers and specialists to collaborate in developing and making use of bioinformatics solutions.
Results: Here we introduce BioStore – a cloud-compatible hub for bioinformatics related tools and platforms.
The BioStore web server uses VNC client and command prompt SSH client to access GUI-based and command-line applications, respectively, which, in turn are organized in containerized manner. We use Docker to apply containerized approach which allows to create an isolated working environment for each project, and efficiently allocate computational resources. The union filesystem (AUFS) used in Docker allows to inherit from already developed images and thus develop new containers.
At the current state BioStore provides a number of tools for systems biology – Cell Designer, Tellurium, COPASI, Cytoscape and , chemoinformatics, gene expression regulation and genome annotation
The concept of BioStore suggests three kinds of users – developers of conainerized bioinformatics workplaces, advanced computational biologists who fluently use the functional of the tools and platforms and “layman” pure biologists or medical scientists who would prefer to run production-ready workflows
The workflows are available in BioUML platform which also could be used as an IDE for various bioinformatics environments like R or Python
Availability: http://wiki.biouml.org

COMPUTATIONAL MODEL FOR MAMMALIAN CIRCADIAN OSCILLATOR INTERACTING WITH NAD+ / SIRT1 PATHWAY

O.A. Podkolodnaya[1], N.N. Tverdokhleb[1, 3], N.L. Podkolodnyy[1, 2, 3]*
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk*
[3] *Novosibirsk State University, Novosibirsk*
* *Corresponding author: pnl@bionet.nsc.ru*

Key words: *computer modeling, mammalian circadian oscillator, SIRT1 pathway*

Motivation and Aim: The mammalian circadian timing system is a finely tuned hierarchical system which regulates a wide range of processes in the body (molecular genetic, physiological and behavioral) with a period close to 24 hours, allowing the body to optimally adapt to the cyclical changes of environment. Almost every cell in an organism contains autonomous molecular genetic circadian oscillator (CO). The structure of this oscillator can be described by a complex gene network with the feedback process mediated transcription, post-translational modification of proteins, protein-protein interactions, chromatin modification, and others. The basis of the circadian oscillator structure constitute by the two interlocked negative feedback loops generating the circadian rhythm. Additional feedbacks of this gene network provide stability of the oscillator functioning and its relationship with the other molecular-genetic systems and pathways of the organism. One of them is formed with the participation of NAD-dependent deacetylase SIRT1, which couples the deacetylation of a number of transcription factors and co-factors to the cleavage of NAD+. Therefore, SIRT1 play a vital role in metabolism, inflammation, apoptosis, stress resistance, energy responses to restriction and high calorie intake, development, and reproduction, which will ultimately affect the processes of aging and disease.
Methods and Algorithms: The circadian rhythm gene network was reconstructed using the GeneNet system (Ananko et al., 2005). Computer model for mammalian circadian oscillator was implemented in MATLAB (Mathworks) as a system of 187 ordinary differential equations.
Results and conclusion: We have modified and extended the most detailed circadian clock mathematical model, developed by Kim and Forger in 2012. In particular, the subsystem comprising genes / proteins NAMPT and SIRT1, as well as NAD + and NAM was added into the model. The clock gene expression data, kinetic constants and characteristics of the dynamics of the mammalian circadian processes for the wild type genotype and different mutations in the clock genes were collected and used to verify the extended mathematical model. A numerical study have demonstrated that the dynamic characteristics of the model, including the period, the amplitude and phase changes of concentrations agrees well with the experimentally observed values.
Acknowledgements: The work was supported by the RSF (the project № 14-24-00123).
References
1. Ananko EA, Podkolodnyy NL, Stepanenko IL, Podkolodnaya OA, et al. GeneNet in 2005 // Nucleic Acids Res.,2005, vol. 33, (Database issue), pp. D425-D427.

COMPUTER ANALYSIS OF BIOLOGICAL NETWORKS OF MAMMALIAN CIRCADIAN OSCILLATOR

N.L. Podkolodnyy[1, 2, 3]*, N.N. Tverdokhleb[1, 3], E.O. Sambilova[3], S.A. Lobynya[3], Z.D. Yakubova[3], O.A. Podkolodnaya[1]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk*

[3] *Novosibirsk State University, Novosibirsk*

* *Corresponding author: opodkol@bionet.nsc.ru, nata@bionet.nsc.ru, pnl@bionet.nsc.ru*

Key words: *computer network analysis, gene network, interatomic network, mammalian circadian oscillator*

The methods of network analysis and searching patterns of structural organization of biological networks including gene networks, interactomics networks, gene co-expression networks, the diseases networks, etc. currently used for solving practical problems of bioinformatics and systems computational biology. The paper presents the methods for analysis of biological networks, including methods for analysis the local and global topological properties of networks, methods for identifying the subsystems on the network, the methods for comparing of network structures, etc., and the results of analysis of gene network applied to research mammalian circadian oscillator. A method for constructing structural models of biological networks (null model) as random graphs with structural patterns similar to the patterns found in the analyzed biological network are developed. Null structural models of biological networks are important for building statistical hypotheses for solving various types of applications is proposed. The original method for description of the integrated structural characteristic of the biological network was proposed. The integrated structural characteristic of the network corresponds to a principal component, built on the basis of the structural characteristics of local units - graphlets frequencies (small related isomorphic induced subgraphs), which include the vertex. On this basis, we developed a method of comparing the network and a statistical criterion for testing the hypothesis under consideration the network of its structural model in the form of random graphs. The circadian rhythm gene network was reconstructed using the GeneNet system (Ananko, 2005). A network of protein-protein interactions (PPI) in the liver at different times of day was reconstructed using experimental data on protein-protein interactions, gene/protein expression data in liver tissue. A PPI network in the liver at different times of day and an expanded version of the gene network of the mammalian circadian oscillator have been reconstructed. A computer analysis of the gene regulatory network of the circadian oscillator and biological interpretation of the identified structural features, including central peaks gene network (hubs), structural patterns of regulation and non-random structural motifs, strongly connected components, regulatory circuits and structural-functional units (clusters) were done. As a result, we identify the central component of the circadian oscillator, which includes basic regulatory circuits passing through the key element of the circadian clock---the protein Clock/Bmal1. The reconstructed structural model, which includes both the central component and functional subsystems interacting with it, became the basis for building an extended mathematical model of the dynamics of the gene network regulating the circadian oscillator.

The work was supported by the RSF (the project № 14-24-00123).

QUANTITATIVE CONTRIBUTION OF IL2RΓ TO THE DYNAMIC FORMATION OF IL2-IL2R COMPLEXES

L.F. Ponce*, K.García-Martínez, K. León
Center of Molecular Immunology, System Biology Department, Habana, Cuba
** Corresponding author:luisf@cim.sld.cu*

Key words: *membrane, modeling, immune cells*

Interleukin-2 (IL2) is a growth factor for several immune cells and its function depends on its binding to IL2Rs in the cell membrane. The most accepted model for the assembling of IL2-IL2R complexes in the cell membrane is the *Affinity Conversion Model* (ACM). This model postulates that IL2R receptor association is sequential and dependent on ligand binding. Most likely free IL2 binds first to IL2Rα, and then this complex binds to IL2Rβ, and finally to IL2Rγ (γc). However, in previous mathematical models representing this process, the binding of γc has not been taken into account. In this work, the quantitative contribution of the number of IL2Rγ chain to the IL2-IL2R apparent binding affinity and signaling is studied. A mathematical model of the affinity conversion process including the γ chain in the dynamic, has been formulated. The model was calibrated by fitting it to experimental data, specifically, Scatchard plots obtained using human cell lines. This paper demonstrates how the model correctly explains available experimental observations. It was estimated, for the first time, the value of the kinetic coefficients of IL2-IL2R complexes interaction in the cell membrane. Moreover, the number of IL2R components in different cell lines was also estimated. It was obtained a variable distribution in the number of IL2R components depending on the cell type and the activation state. Of most significance, the study predicts that not only the number of IL2Rα and IL2Rβ, but also the number of γc determine the capacity of the cell to capture and retain IL2 in signalling complexes. Moreover, it is also showed that different cells might use different pathways to bind IL2 as consequence of its IL2R components distribution in the membrane.

EPIGENOMIC CHANGES IN POSTMORTEM BRAINS OF HUMAN ALCOHOLICS

I. Ponomarev
Waggoner Center for Alcohol and Addiction Research, University of Texas, Austin, Texas, USA
** Corresponding author: ponomarev@utexas.edu*

Key words: *Alcoholism, Neuroscience, Genomics, Epigenetics, Chromatin*

Motivation: Chronic alcohol abuse is associated with epigenetic changes including DNA methylation and histone modifications that ultimately control long-term changes in gene expression and behavior. Histone H3 lysine 4 trimethylation (H3K4me3), a promoter-enriched chromatin (epigenetic) mark of actively transcribed genes, has been implicated in psychiatric disorders including drug addiction. The effects of chronic alcohol on genome-wide distribution of the H3K4me3 mark and its relationship with alcohol-induced changes in gene expression are not well understood.
Methods: We used chromatin immunoprecipitation followed by next generation sequencing (ChIP-Seq) to obtain genome-wide distributions of this mark in superior frontal cortex from postmortem brains of 24 human alcoholics and 24 matched control cases. An antibody against H3K4me3 was used for the ChIP step to isolate specifically bound DNA and non-immunoprecipitated DNA was used as input control. Sequencing was carried out using Illumina HiSeq paired-end sequencing (2x100 base pairs), yielding 20-30 million reads per sample.
Results: We identified multiple H3K4me3 peaks differentially regulated in gene promoters between the alcoholic and control groups. Our gene network approach highlighted genes involved in synaptic transmission and myelination, as two functional groups potentially regulated by alcohol-induced changes in H3K4me3. The network approach identified subsets of functionally related transcripts that are regulated in agreement with H3K4me3 changes, suggesting cause and effect relationships between this epigenetic mark and gene expression. In addition, we identified H3K4me3 peaks differentially affected by alcohol in males and females.
Conclusions: These data provide support for our previous findings showing global epigenetic changes caused by alcohol in the human cortex. Taken together, our results point to an important role of the H3K4me3 modification in the regulation of alcohol-induced changes in gene expression and downstream neuroadaptations and pathologies associated with alcohol use disorders.
Acknowledgements: Funding: NIH, NIAAA grants R21 AA021462 and U01 AA0209260.

PUNCTUATED EVOLUTION: THE RELATIONSHIP BETWEEN RARE MUTATIONS AND CLADOGENESIS OF VERTEBRATES

K. Popadin[1], K. Gunbin[2, 3]*
[1] *Center for Integrative Genomics, University of Lausanne, Switzerland*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3] *Novosibirsk State University Novosibirsk, Russia*
* *Corresponding author: genkvg@gmail.com*

Key words: *molecular evolution*, *cladogenesis, rare amino-acid substitutions*

Motivation and Aim: Is there a link between the rate of molecular evolution and the rate of cladogenesis? At one side, there is evidence that critically important mutations are acquired during short speciation period [1, 2]. On the other side, there is evidence that the rate of cladogenesis is decoupled from the rate of molecular evolution [3, 4]. Analyzing statistically rare amino acid changes we demonstrated that fast rate of their fixation is associated with periods of diversification in mammalians.
Methods and Algorithms: The initial orthology relations between proteins were taken from OrthoDB 6. To select highly related sets of proteins we define Strict Orthologous Protein Groups (SOPGs) as subgroups of Large Orthologous Protein Groups (LOPGs), containing proteins from at least 5 species with monophyletic origin. We followed several computational steps in order to reveal statistically rare amino-acid substitutions on each branch of the phylogenetic trees. Firstly, the multiple alignments of the SOPG and LOPG have been constructed. Secondly, a symmetric matrix of relative rates of amino acid substitution has been derived on the basis of the multiple alignment of the LOPG. Thirdly, the phylogram describing the evolutionary rates of SOPG proteins has been reconstructed using species tree. Fourthly, the ancestral sequences has been reconstructed (separately for amino acids and indels) using the substitution matrix, phylogram and purified multiple alignment. Finally, using the ancestral sequences we calculated the number of observed amino acid substitutions. A comparison of the observed and expected (coming from computer simulations of molecular evolution) numbers of each substitution type was performed using permutation test.
Results: Is there a connection between the frequency of rare replacements and genus birth in the fossil record? To check this, we compared rates of taxon formation in the fossil record and the fraction of SOPG clusters that fix rare substitutions on the internal nodes of Metazoa tree. We selected Vertebrata taxa for the reasons of high quality of paleontological description. Positive associations were observed between the frequencies of rare substitutions and the genus birth rate in the paleontological history. To interpret the obtained results we discuss two different scenarios: periods of positive selection or relaxed negative selection associated with fast speciation.
Availability: Data available upon the requests to the authors.
Acknowledgements: The study is supported by grant 14.B25.31.0033 (Resolution No. 220) from the Russian Federation Government.
References:
1. M. Pagel et al. (2006) Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science 314:119-121.
2. J. Flegr (2010) Elastic, not plastic species: frozen plasticity theory and the origin of adaptive evolution in sexually reproducing organisms. Biol Direct 5:2.
3. X. Goldie et al. (2011) Diversification and the rate of molecular evolution: no evidence of a link in mammals. BMC Evol Biol 11:286.
4. R. Lanfear et al. (2010) Mutation rate is linked to diversification in birds. Proc Natl Acad Sci USA 107:20423-20428.

TASSE: A NEW APPROACH TO SOLVENT TREATMENT IN MOLECULAR DYNAMICS

A.V. Popov*, Yu.N. Vorobjev, D.O. Zharkov
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
** Corresponding author: apopov@niboch.nsc.ru*

Key words: *computer simulation, molecular dynamics, implicit solvent*

Motivation and Aim: Molecular dynamics approach is one of the most important ways to learn about atomic structure of proteins, nucleic acids, and their complexes, and is the cornerstone of modern methods of rational drug design. Both classic molecular dynamics and docking methods based on it keep evolving towards increased reliability. One of the central problems that have to be addressed for computations to be meaningful is finding out the optimal way of solvent treatment, which is important for maintaining adequate structures and interactions of proteins and nucleic acids. Nowadays there are two major solvent models: explicit and implicit, i.e. approximate mathematical model-based. Although the explicit solvent easily accounts for molecular effects by design, it suffers from decreased mobility due to viscosity, resulting in larger conformational transition time scales. Besides, higher atom count has a severe impact on performance
Methods and Algorithms: A hybrid model can be defined as follows. Tightly bound solvent molecules forming close contacts with biopolymer are modeled explicitly, and the remote solvent is represented with an infinite continuum. Thus computational efficiency depends on a careful minimization of the solvent molecule count.
Results: TASSE (Tightly Associated Solvent Shell Extractor) is new software developed to solve the problem of selecting explicit solvent molecules. It gathers the information about importance of particular solvent molecules from a preliminary, relatively short modeling in an explicit solvent. The program analyses hydrogen bonding and electrostatics along the trajectory, and collects information about the solvent bridges between atoms of the biopolymer. Then the bridges are sorted by relevance, and the final structure containing solvent molecules at statistically important sites is built. There are also fine tuning options, including explicit cutoff by solvent molecule count. The resulting structure can be further modeled within the limits of continual models. TASSE supports the popular AMBER software suite, BISON package, and any other package that is able to export trajectories in PDB format.
Conclusion: The new software, TASSE, performs effective minimization of explicit solvent molecules, which is important in development of a new hybrid molecular-continual approach to solvent treatment.
Availability: TASSE program will be freely available online at http://bison.niboch.nsc.ru. The source code will be released under GPL terms.
This work was supported by Russian Foundation of Basic Research (15-04-00387), RAS MCB Program (6.11), Federal Agency of Science Organization (SB RAS VI.57.1.4) and Russian Science Foundation (16-14-10038).

COMPUTER-AIDED DRUG REPURPOSING: NEW USES FOR OLD DRUGS OR FILLING GAPS IN BIOMEDICAL KNOWLEDGE?

V.V. Poroikov*, D.A. Filimonov, A.A. Lagunin, T.A. Gloriozova
Institute of Biomedical Chemistry, Moscow, Russia
** Corresponding author: vladimir.poroikov@ibmc.msk.ru*

Key words: *drug-like substances, biological activity, incompleteness of information, bioinformatics, chemoinformatics, computer-aided drug repurposing, PASS, PharmaExpert, "drug – target – effect – disease" associations*

Motivation and Aim: Biological activity is one of the most important properties of organic substances, which provides the possibility of their use as medicines. The current dimensionality of chemical-biological space for available chemical compounds and known pharmacological targets is about 10^{10}, while for virtual compounds and potential targets it is about 10^{16}. Thus, experimental study of the interaction of all drug-like structures with each known pharmacological target could not be accomplished from both economical and practical point of view, and all existing information about biologically active compounds is incomplete. Bio- and Chemoinformatics shed light on a hidden pharmacological potential of launched drugs that may provide the reasons for their repurposing. Due to the current knowledge about pharmacodynamics and pharmacokinetics of the launched drugs, their repurposing may significantly reduce the time & financial expenses and risks of the development.
Methods and Algorithms: We will present an overview of the currently existing target-based and ligand-based methods of computer-aided drug repurposing with particular highlighting of our software PASS and PharmaExpert [1, 2].
Results: Published in 2001 PASS predictions of novel pharmacotherapeutic actions for eight from the list of Top200 drugs have been further confirmed either by the experimental or by clinical studies for Sertraline (Cocaine dependency treatment), Amlodipine (Antineoplastic enhancer), Oxaprozin (Interleukin 1 antagonist), Ramipril (Antiarthritic). Later we anticipated the nootropic action of some antihypertensive drugs (Perindopril, Ramipril, Quinapril, etc.) that has been confirmed by the experiment [3] and in clinical trials [4].
Conclusions: The considered examples undoubtedly demonstrate the potential of computer-aided methods in drug repurposing. Moreover, computer-aided approaches leads to the filling the gaps in the existing biomedical knowledge due to the extraction of novel "drug – target – effect – disease" associations.
References:

1. Filimonov D.A. et al. (2014) Prediction of the biological activity spectra of organic compounds using the PASS online web resource. *Chem. Heterocycl. Compnds.*, 50: 444-457.
2. URL [http://www.way2drug.com].
3. Kryzhanovskii S.A. et al. (2012) Nootropic action of some antihypertensive drugs: computer predicting and experimental testing. *Pharmaceut. Chem. J.*, 45: 605-611.
4. Gao Y. et al. (2013) Effects of centrally acting ACE inhibitors on the rate of cognitive decline in dementia. *BMJ Open*, 3: e002881.

INTRON EVOLUTION: SLIDING AND VARIABILITY OF LENGTH

I.V. Poverennaya[1]*, D.D. Gorev[2], T.V. Astakhova[3], M.A. Roytberg[2, 3]
[1] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*
[2] *Moscow Institute of Physics and Technology, Moscow, Russia*
[3] *Institute of Mathematical Problems of Biology RAS, Pushchino, Russia*
* *Corresponding author: ipoverennaya@gmail.com*

Key words: *introns, intron evolution, intron phase, sliding, exon-intron structure*

Motivation and Aim: Intron sequences evolve so fast that even their lengths seem to be highly unconserved. Thus, intron evolution is usually considered in terms of evolution of exon-intron structures (EIS). One of the rarest and intriguing evolutionary event is intron sliding that is a shifting of exon-intron boundaries over short distances. Such relocation could lead to the change of intron phase, i.e. the position of intron relative to the open reading frame that might affect the "golden ratio" of intron phase distribution. Here we analyze the exon-intron structure of four different eukaryotic genes in order to find out the preferable choice of intron phase during intron sliding and to study the correlation between lengths of orthologous introns.
Methods and Algorithms: Identification of orthologous introns requires exon-intron structure alignment of orthologous genes. To construct them we have applied our own protocol of manual EIS aligning based on multiple sequence alignment of genes products and exon lengths. The obtained alignments have been used as training and testing sets for multiple EIS alignment program developed by us.
Results: For each analyzed gene about 100 orthologues from different vertebrates (mammals, amphibians, fishes, birds, etc.) was obtained through the Annotation Pipeline NCBI database (http://www.ncbi.nlm.nih.gov/gene/) to make a dataset. Analysis of EIS alignments has revealed several cases of sliding for each dataset. In every case, the sliding caused an intron phase change; however, there seem to be no preference of novel or initial phase. Analysis of the intron lengths showed that despite high variability of intron length, some correlations could be observed especially in separate taxa. Moreover, if instead of intron length L we will consider a normalized length N = (L-A)/A, where A is an average length within a group of orthologous introns. E.g. for *ptprd* genes of birds (28 species were considered) the normalized value is in the interval (-0.15, 0.15) for 85.2% of introns what is significantly higher than the values for random set of lengths in accordance with the distribution of the lengths of the introns. Also, for the interval (-0.5; 0.5) the according proportion of introns is 96.8%.
Conclusion: Obtained results did not confirm our initial hypothesis that in the process of sliding introns prefer to change its phase to 0 more frequently. However, it is necessary to expand the analysis on a larger dataset of genes for making a final conclusion. Despite the wide range of orthologous intron lengths, some intron length conservation could still be observed and leads us to the question what intron length the ancient introns had.

DNA DAMAGE AND GENERATION OF REACTIVE OXYGEN SPECIES BY PLATINUM DRUGS: EXPERIMENTS ON BACTERIA

E.V. Prazdnova*, V.A. Chistyakov, M.S. Mazanko, M.N. Churilov, V.K. Chmyhalo
Academy of Biology and Biotechnologies of Southern Federal University, Rostov-on-Don, Russia
** Corresponding author: prazdnova@sfedu.ru*

Key words: *cisplatin, DNA damage, oxidative stress, mutagenesis, antibiotic resistance*

Motivation and Aim: Cisplatin is anticancer drug that provides the cytotoxic effect by inducing apoptosis through the formation of intrachain DNA–adducts [1]. The use of cisplatin is limited by a wide range of side effects, strong mutagenic and genotoxic effects [2]. We suppose that the drug may cause mutations in the patient's microflora metagenome. That leads to secondary infections, caused by microorganisms resistant to antimicrobial agents. The effects of cisplatin may also be mediated by the generation of reactive oxygen species [3]. Therefore, antioxidants may reduce the mutagenic potential of cisplatin. We performed a study to investigate whether the mutagenic effect of cisplatin to bacteria is also due to oxidative stress.

Methods and Algorithms: We used a range of bacterial biosensors reacting to oxidative stress and DNA damage, based on *E.coli* strains MG1655 pKatG-lux (registers formation of hydrogen peroxide in the cell), pSoxS-lux (reacts to the increased superoxide-anion-radical level), and pColD-lux (registers DNA damage). To estimate mutation rate, we applied the standard serial dilution method.

Results: The biosensor assay demonstrated high genotoxic activity of cisplatin, and a slight induction of superoxide anion radical, with no generation of hydrogen peroxide. It was shown that ascorbate reduces the genotoxic effect of cisplatin by 41% in this model system. Non-lethal doses of cisplatin induced 3-7-fold increase in the frequency of mutant resistant to rifampicin and ciprofloxacin in *E. coli* MG 1655. It was found that ascorbate reduces mutagenesis induced by cisplatin by 65%. However, it decreased the toxicity of drug too.

Conclusion: The mechanism of drug action on bacteria appears to be associated with the generation of superoxide anion-radical. To reduce the risk of secondary infections, complicated by antibiotic resistance, it seems reasonable to use antioxidants, but it should be taken into account that they can reduce the general cytotoxicity of drug.

Acknowledgements: This work was supported by the Ministry of Education and Science of the Russian Federation (project No. 6.1202.2014/K)

References:

1. M.J.Silva et al. (2005) Comparative analysis of the mutagenic activity of oxaliplatin and cisplatin in the Hprt gene of CHO cells, *Environmental and molecular mutagenesis***,** 46.2: 104-115.
2. G.Brozovic et al. (2008) Evaluation of DNA damage *in vivo* induced by combined application of cisplatin and sevoflurane*, European journal of anesthesiology*, 25.08: 642-647.
3. A. Laurent et al. (2005) Controlling tumor growth by modulating endogenous production of reactive oxygen species, *Cancer research* 65.3: 948-956

GENEQUERY: GLOBALLY CONNECTED NETWORKS OF GEO TRANSCRIPTIONAL PROFILES SHOW HYPOTHESIS GENERATION POTENTIAL AND REVEAL THAT TOCOPHEROLS RESCUE TREM2-ASSOCIATED MICROGLIAL DYSFUNCTION

A.V. Predeus[1, 2]*, T. Ulland[1], Y. Wang[1], V. Lampropoulou[1], W. Song[1], I. Arbuzov[3], F. Towfic[4], S. Gilfilan[1], E. Loginicheva[1], B.T. Edelson[1], B. Zeskind[4], M. Colonna[1], M.N. Artyomov[1]

[1] *Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, MO, USA*
[2] *Bioinformatics institute, Saint Petersburg, Russia*
[3] *ITMO University, Saint Petersburg, Russia*
[4] *Immuneering Corporation, Cambridge, MA, USA*
* *Corresponding author: predeus@gmail.com*

Key words: *GeneQuery, GEO, transcriptional networks, phenotype search, hypothesis generation, TREM2 deficiencies, Nasu-Hakola disease*

Modern collections of transcriptional profiling experiments contain enormous wealth of information, which is severely underutilized due to inconsistent annotation, cross-platform differences and wide spectrum of conditions and tissues profiled. On the other hand, most of the modern pathway analysis tools rely on curated gene sets that quickly become outdated and often fail to capture true diversity of transcriptional responses in real biological systems. To reveal hypothesis generation potential of transcriptional profiling databases, we developed *GeneQuery*, new geneset-based global phenotype searching tool that makes use of gene expression data in GEO database. *GeneQuery* circumvents aforementioned difficulties by introducing digital definition of phenotypes through gene modules co-expressed in a given dataset. Since there is an established connection between co-expression and co-regulation of groups of genes, we used co-expressed modules as a representative of particular phenotype in the transcriptional universe. Careful application of WGCNA approach allowed us to automatically and unbiasedly obtain co-expressed modules of genes that are subsequently compared to the geneset in question. Using regular Fisher's exact test with Bonferroni correction, we were able to establish a phenotype search engine that finds biologically similar experiments based on the transcriptional signatures. Furthermore, using network methods we have analyzed the cross-connectivity of the overall "transcriptional universe" graphs of humans, mice, and rat, and have found that both conserved and species-specific clusters are present for each species. Overall, nearly half of all available transcriptional experiments (spanning over 400,000 samples) are included in the database, which is dynamic and easily expandable.We first used *GeneQuery* to unbiasedly characterize the "microarray expression universe" and then explored its hypothesis generation potential in various biological settings. GeneQuery revealed an unexpected connection between transcriptional signatures of patients with Nasu-Hakola disease, a rare neurodegenerative disease caused by TREM2 and DAP12 mutations, and a portion of the aging-signature in mouse brain consisting of genes responsive to α/γ-tocopherol treatment. Utilizing a mouse model of TREM2-associated microglial deficiency, we demonstrated that α/γ-tocopherol treatment rescued microglial function in *Trem2*-deficient mice but did not affect WT microglia. These results validate a powerful new computational approach, highlight the critical role of TREM2 in microglial function, and suggest new therapeutic approaches for treating TREM2-associated neurodegeneration. *GeneQuery* is available free of charge at https://artyomovlab.wustl.edu/genequery-alpha/searcher/.

THE ROLE OF HUNTINGTIN PROTEIN-PROTEIN INTERACTIONS IN THE PROCESSES OF CHANGING AND MAINTENANCE OF NEUROTRANSMISSION IN HIPPOCAMPUS

A.L. Proskura*, S.O. Vechkapova, T.A. Zapara, A.S. Ratushniak
Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia
** Corresponding author: annleop@mail.ru*

Key words: *huntingtin, protein-protein interactions, vesicular transport, synaptic plasticity, long-term potentiation*

Motivation and Aim: Huntington's disease is caused by a pathological expansion of a CAG repeat in the first exon of the gene coding the huntingtin (HTT), resulting in an abnormally long polyglutamine stretch. Huntingtin is normally present in the cytoplasm where it may interact with structural and synaptic elements. Studies indicate a toxic gain-of-function possibly through the aberrant protein interactions. Presymptomatically Huntington's disease patients often exhibit the cognitive deficits before the onset of classical symptoms. At early stages there are structural alterations in the synapses of the hippocampus, abnormal synaptic plasticity [1, 2], progressive imbalance in the interaction between spatial and procedural memory systems [3]. The aim was to study the huntingtin contribution to the changing the synaptic plasticity in the hippocampus.
Results: Using GeneNet technology and Pubmed database it was shown that huntingtin is not involved in exocytosis of glutamate receptors while inducting the long-term potentiation, but it involved in the processes of maintaining the efficiency of synaptic transmission. Normally HTT acts as a scaffold molecule providing the order of events when endocytosis of glutamate receptors takes place, and also is involved in the process of moving the vesicles with receptors along the tubulin dendritic cytoskeleton. The pathogenic form of protein leads to disruption of the intermolecular interactions and also may enter into competitive interaction with proteins of postsynaptic density, and regulators of cytoskeleton remodeling.
Conclusion: The appearance of abnormally long polyglutamine stretch in the huntingtin molecule results in enhanced the affinity of its interaction with domains of synapse proteins that at early stages of the disease may be associated with certain disorders of functional synaptic plasticity, and in later stages results in a toxic gain-of-function and the death of neurons in the higher regions of the brain.
Availability: http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/viewer/AMPA.html
Acknowledgements: The work was supported by VI 35.1.5 basic project of fundamental researches of RAS and RFBR grant 15-29-04875.
References:
1. DiProspero N.A. et al. (2004) Early changes in Huntington's disease patient brains involve alterations in cytoskeletal and synaptic elements. J Neurocytol. 33:517-533.
2. Murphy K.P. et al. (2000) Abnormal synaptic plasticity and impaired spatial cognition in mice transgenic for exon 1 of the human Huntington's disease mutation. J Neurosci. 20: 5115-5123.
3. Ciamei A., Morton A.J. (2009) Progressive imbalance in the interaction between spatial and procedural memory systems in the R6/2 mouse model of Huntington's disease. Neurobiol Learn Mem. 92: 417-428.

STUDY OF *ARMILLARIA BOREALIS* PATHOGENICITY BY THE COMPARATIVE WHOLE GENOME SEQUENCING

Yu.A. Putintseva[1, 2]*, I.N. Pavlov[1, 2], N.V. Oreshkova[1, 2], V.V. Sharov[1], D.A. Kuzmin[1], S.V. Makolov[1], K.V. Krutovsky[1, 3, 4, 5]

[1] *Siberian Federal University, Krasnoyarsk, Russia*
[2] *V.N. Sukachev Institute of Forest SB RAS, Krasnoyarsk, Russia*
[3] *Georg-August University of Göttingen, Göttingen, Germany*
[4] *N.I. Vavilov Institute of General Genetics, RAS, Moscow, Russia*
[5] *Texas A&M University, College Station, USA*
* *Corresponding author: yaputintseva@mail.ru*

Key words: *genome sequencing, Armillaria borealis, genome assembly, transcriptome, gene expression*

Motivation and Aim: Armillaria mellea s.l. is one of the main fungal pathogens of woody plants of the boreal forests. Similar species, *Armillaria borealis* (Marxm. & Korhonen) is widely distributed in Siberia and the Far East and is also causing the root rot disease leading to a weakening or often death of woody plants. Yet, no large-scale "-omics" data are available for these species that can help to uncover their pathogenicity. Therefore, we sequenced the genome and five transcriptomes of *A. borealis* to further promote the study of basidiomycete pathogenicity.

Methods and Algorithms: The *A. borealis* genome and five transcriptomes have been *de novo* sequenced, assembled and characterized in this study. Fungal material was collected from an active mycelia of *A. borealis* taken from *Abies sibirica* trees died in 2015 and differed by the factors that caused their death. DNA and cDNA were sequenced using 250-bp insert paired-end libraries on the Illumina MiSeq platform at the Laboratory of Forest Genomics of the Siberian Federal University. A *de novo* genome assembly and gene expression analyses of five transcriptomes were performed using the CLC Genomics Workbench. Gene model predictions were conducted using the MAKER2 pipeline. A preliminary functional annotation of predicted gene models and gene ontology assignment were performed using Blast2GO.

Results: A. borealis genome assembly was ~99.6 Mbp long (twice as larger compared to 58.35 Mb for the *A. mellea* genome, but close to 84 Mbp for the *A. gallica* genome) with N50 for contigs equaled 8600 bp. The length of the mitochondrial genome was 103.5 Kbp. Gene expression analysis of five transcriptomes from samples with different degrees of pathogenicity showed significant difference in expression of genes involved in in plant cell wall degradation.

Conclusion: Currently, work is underway to improve the quality of assembly and annotation of *A. borealis* and to construct metabolic pathways of *Armillaria* pathogenicity. These genome and transcriptomes contribute to the studies of woody plants fungal pathogen, and our results provide useful insights towards identifying specific genes potentially associated with pathogenesis and other metabolic functions.

Acknowledgements: This study was supported by Research Grant No. 14.Y26.31.0004 from the Government of the Russian Federation.

RNA SEQ ANALYSIS OF MARINE AND FRESHWATER FORMS OF THREE-SPINED STICKLEBACK *(GASTEROSTEUS ACULEATUS)*. EVOLUTIONARY AND PHYSIOLOGICAL MECHANISMS OF ADAPTATION

S.M. Rastorguev[1]*, A.V. Nedoluzhko[1], A.M. Mazur[2], E.B. Prockhorchouk[2]
[1] *National Research Center "Kurchatov Institute", Moscow, Russia*
[2] *Institute of Bioengineering, Research Center of Biotechnology of RAS, Moscow, Russia*
** Corresponding author: rastorgueff@gmail.com*

Key words: *Gasterosteus aculeatus, evolution, adaptation, RNAseq, netx-generation sequencing*

The three-spined stickleback (*Gasterosteus aculeatus*) is known as a model organism for studying the population genetics and speciation during habitat changes. Besides the marine form, it exists in multiple divergent freshwater populations along the coast of the Northern hemisphere. The native marine population of the three-spined stickleback uses freshwater streams and lakes for spawning. However, the isolation of the freshwater habitat can result in a new resident freshwater population, which eventually changes the morphotype and acquires other adaptive features. A large number of such freshwater populations of the three-spined stickleback originate alongshore of the Northern hemisphere providing a convenient model for the study of adaptive evolution in similar habitats. RNA seq is common way to study organisms on function level. Defining a gene list, whose expression was significantly different in the control and experimental group, allows functionally characterize the genetic changes under the designed influence. The RNA-seq analysis was done on samples of marine and freshwater forms of three-spined stickleback (Gasterosteus aculeatus). We used control marine and freshwater samples, which were been held in their native environment, and case samples, holding in nonnative environment. Accordingly, we had four groups of samples - marine, freshwater, marine samples in freshwater environment and freshwater samples in marine environment, five samples in each group. After the experiment, gills of the samples were isolated for subsequent RNA-seq analysis. Gills were chosen as target tissues as they play an important role in osmotic balance, and the expression of genes, involved in osmotic regulation, is likely to be observed. RNA-seq analysis reveal 2983 genes with significantly different expression level (95% confidence level) between marine and freshwater controls. Totally 22,457 genes annotated in the Ensembl database for the three-spined stickleback. The expression of samples, placed in the nonnative environment showed a lower difference level - 380 genes differ when freshwater samples placed in sea water, and 413 genes differ when marine samples placed into fresh water. Approximately half of the genes from differential expression gene lists coincide with the list of genes that differ between the controls, indicating that the rapid physiological adaptation response to changing conditions partially overlaps with evolutionary adaptation that occurs in different ecological forms of three-spined stickleback. A series of tests with the obtained gene lists, when they were correlated to each other, Gene Ontology enrichment and other analysis confirmed the point of view.
Acknowledgements: This work is supported by the Russian Science Foundation (Grant 04-01237 data sampling and analyzing; and 14-24-00175 sequencing).

GENETIC AND MOLECULAR MECHANISMS CRUCIAL FOR HYPERTENSION DEVELOPMENT IN THE ISIAH RATS

O.E. Redina[1]*, L.O. Klimov[1], M.A. Ryazanova[1], L.A. Fedoseeva[1], T.O. Abramova[1], Yu.V. Alexandrovich[1], S.E. Smolenskaya[1], Ye.V. Antonov[1], N.I. Ershov[1], V.M. Efimov[1, 2], A.L. Markel[1, 2]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

** Corresponding author: oredina@ngs.ru*

Key words: *stress-sensitive hypertension, ISIAH rats, RNA-Seq, soluble epoxide hydrolase*

Motivation and Aim: Arterial hypertension (AH) is a multifactorial disease with a significant genetic component. Search for new molecular targets and novel approaches to prevention and treatment of the hypertensive disease still remains an actual problem. The aim of the current study was to analyze the age-dependant changes in molecular mechanisms related to blood pressure control in the ISIAH rats with stress-sensitive AH and to reveal the genes which might be considered as potential genetic targets for AH treatment. It was suggested that the target genes should be associated with AH and contribute significantly to its manifestation not only in established disease, but also in the early stages of it pathogenesis.

Methods: The work was carried out on 1-month old prehypertensive, and 3-month old hypertensive ISIAH/Icgn, and normotensive control WAG/GSto-Icgn males of the same ages. Each experimental group consisted of 3 rats maintained under standard conditions. RNA-Seq approach was used for comparative transcriptome profiling in the brain stem, hypothalamus, adrenals and kidney (renal cortex and renal medulla), which are known as the main target organs in AH development. Experiments were approved by the Institute Animal Care and Use Committee. The Cufflinks/Cuffdiff programs were used to detect the differentially expressed genes (DEGs). The PLS-DA method and Pearson correlation were used to find a set of variables (genes) that contribute the most to inter-strain differences.

Results: The development from the prehypertensive state to hypertensive led to the decrease in the number of DEGs in all organs analyzed. Less than a half of DEGs was common between the groups of different ages in each organ. The PLS-DA approach helped to define DEGs making the most contribution to inter-strain differences. One of these, *Ephx2,* associated with AH was common for all organs in rats of both ages. *Ephx2* was significantly overexpressed in organs of ISIAH rats. *Ephx2* encodes the soluble epoxide hydrolase (sEH) that metabolizes the epoxyeicosatrienoic acids having vasodilation properties. sEH overexpression was linked to hypertension in different rat models of AH and was considered as a promising target for treatment of AH [1].

Conclusion: The AH development in the ISIAH rats is associated with multiple age-dependant changes in molecular mechanisms. *Ephx2* may be indicated as deserving high priority in future molecular investigations of the stress-sensitive hypertension development in ISIAH rats.

Acknowledgements: This work has been supported by the Russian Science Foundation, grant No. 14-15-00118.

References: 1. Chiamvimonvat N., et al. (2007). The soluble epoxide hydrolase as a pharmaceutical target for hypertension. J Cardiovasc Pharmacol., 50: 225-237.

ROLE OF MEMBRANE POTENTIAL IN NITRITE UTILIZATION BY ESCHERICHIA COLI CELLS UNDER LOW SUBSTRATE CONCENTRATIONS: THE MATHEMATICAL MODEL

N.A. Ree*, V.A. Likhoshvai, T.M. Khlebodarova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: seahourse@mail.ru*

Key words: *nitrite respiration, membrane potential, mathematical model*

Motivation and Aim: E.coli is a facultative anaerobe that can use nitrite as electron acceptor for ATP synthesis. Under low nitrite concentrations in the environment, the main nitrite-reducing enzyme is periplasmic Nrf reductase. It was shown [1], that molecular-genetic mechanisms of Nrf activity regulation are not sufficient to describe the nitrite accumulation dynamics in the chemostat in the micromolar range. The hypothesis was suggested [1] that the local change of Nrf concentration while transition from the cytoplasm to the periplasm under membrane potential may be a mechanism that cause higher activity of Nrf reductase in the periplasm than it could be expected according to *nrf* operon expression level [2]. The aim of this work is theoretical verification of this assumption.
Methods and Algorithms: Generalized Hill functions [3] were used to describe gene expression mechanisms, involved in nitrite electron transport chain and nitrite metabolism (*fdn*, *nrf* and *nir*) in *E.coli* cells. Rate of enzyme reactions were described by Michaelis-Menten equations. Parameters of the model were evaluated from the published data or were estimated during model's adaptation to experimental data. STEP+ was used for numerical calculations of the model.
Results: The mathematical model was created that describes *E.coli* cells cultivation in the chemostat at anaerobic conditions on nitrite and represents molecular-genetic mechanisms of respiratory chain formation, nitrite metabolism regulation and kinetic of substrate utilization during continuous culture growth. The model analysis revealed that taking into account the influence of membrane potential on Nrf subunits transport from the cytoplasm to the periplasm is sufficient to describe nitrite utilization kinetic in the chemostat under micromollar substrate concentration range from the experiments.
Conclusion: Significant contribution of membrane potential on periplasmic Nrf activity and nitrite utilization dynamic under substrate concentration <1 mM was theoretically proved.
Acknowledgements: The work was partially supported by the RFBR (No 16-01-00237a) and Project Fundamental Research of SB RAS (No 0324-2015-0003).
References:
1. N.A. Ree et al. (2015) About mechanisms of nitrite utilization by Escherichia coli cells under stationary growth cultivation, Math. Biol. Bioinf., 10: 193-205.
2. H. Wang, R.P. Gunsalus (2000) The nrfA and nirB nitrite reductase operons in Escherichia coli are expressed differently in response to nitrate than to nitrite, J. Bacteriol. 182: 5813–5822.
3. V. Likhoshvai, A. Ratushny (2007) Generalized Hill function method for modeling molecular processes, J. Bioinform. Comput. Biol. 5: 521–531.

MODELING RESTRICTION-MODIFICATION SYSTEMS: EXPRESSING TOXIC MOLECULES WITHIN A CELL

A. Rodic*, M. Djordjevic
Faculty of Biology, University of Belgrade, Serbia
** Corresponding author: andjela.rodic@bio.bg.ac.rs*

Key words: *restriction-modification systems, modeling transcription regulation, single cell experiments, expression dynamics*

Motivation and Aim: Bacterial restriction-modification (RM) systems encode a restriction enzyme (R), which cuts specific DNA sequences, and a methyltransferase (M), which methylates and protects the same sequences. Expression of these enzymes during the system establishment in a naïve bacterial host has to be tightly regulated to prevent the host DNA being damaged by the toxic molecule (R), with this regulation often being accomplished by a control (C) protein. We aim understanding which features of RM systems allow this tight regulation. To that end, we provided modeling for recent state-of-the-art single cell measurements of RM dynamics, and also systematically *in silico* abolished the main RM features.
Methods: We developed a model for *in vivo* dynamics of R and M expression in a cell, which we compared with the first single-cell measurements done for Esp1396I RM system [1]. We used statistical thermodynamics to model the system transcription regulation, which was then used as an input for the dynamical modeling, that we subsequently implemented in a numerical procedure allowing direct comparison with the measured data. We used a similar model to perturb (*in silico* mutate) the main system features in another RM system with convergent architecture (AhdI). These features include an extremely high binding cooperativity, the differential translation efficiency, and the high dimerization constant. We then systematically analyzed the effect of these perturbations on the system dynamics [2].
Results: Our model successfully reproduces the main experimentally measured qualitative features of the expression dynamics – the significant delay of R with respect to M expression, including a high pic in M expression for the early times [1]. Regarding the *in silico* mutations, the perturbations generally abolish the three main dynamical features of the system: a delay in R expression, a fast transition from "OFF" to "ON" state, and the steady state stability [2].
Conclusion: The developed theoretical model can satisfactory explain the first direct measurements of the enzyme expression in RM systems. Our results suggest that the prominent RM features are likely optimized to satisfy few dynamical constraints. The inferred principles provide guidelines for constructing synthetic gene circuits capable of efficiently expressing toxic molecules [3].
Acknowledgement: This work is funded by the Swiss National Science Foundation (IZ73Z0_152297), and Serbian Ministry of Science and Technology (ON173052).
References:
1. N. Morozova, A. Sabantsev, E. Bogdanova, Y. Fedorova, A. Maikova, A. Vedyaykin, A. Rodic, M. Djordjevic, M. Khodorkovskii, K. Severinov (2015), *Nucleic Acids Researc*h, 44: 790-800.
2. A. Rodic, et al. (2016) submitted.
3. A. Rodic, et al. (2016) in preparation.

GENETIC FITNESS OF DEAF PEOPLE IN THE SAKHA REPUBLIC

G.P. Romanov[1]*, N.A. Barashkov[1, 2], F.M. Teryutin[1, 3], L.A. Klarov[3], A.V. Solovyev[1], N.N. Gotovtsev[1], V.G. Pshennikova[1, 2], N.N. Sazonov[1], I.V. Morozov[4, 5], A.A. Bondar[4], L.U. Dzhemileva[6], E.K. Khusnutdinova[6, 7], O.L. Posukh[5, 8], S.A. Fedorova[1, 2]

[1] *Institute of Natural Sciences, M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia*
[2] *Yakut Scientific Center of Complex Medical Problems, Yakutsk, Russia*
[3] *Republican Hospital # 2 – Center of Emergency Medicine, Yakutsk, Russia*
[4] *Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia*
[5] *Novosibirsk State University, Novosibirsk, Russia*
[6] *Institute of Biochemistry and Genetics, Ufa Scientific Centre, Ufa, Russia*
[7] *Bashkir State University, Ufa, Russia*
[8] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
* *Corresponding author: gpromanov@gmail.com*

Key words: *GJB2, deafness, genetic fitness, Eastern Siberia, Russia*

Motivation and Aim: Introduction of a sign language in schools for deaf people led to growth of their genetic fitness, which has doubled the *GJB2* gene, associated deafness in the USA over the past 200 years. High prevalence of the *GJB2*-deafness and relatively recent (~ 60 years ago) introduction of sign language among deaf people were recorded among indigenous Yakut population (Eastern Siberia, Russia).
Methods and Algorithms: We have performed study of fertility of deaf people compared to their hearing siblings in Eastern Siberia. Fertility was determined as mean number of children born to specific group. Genetic fitness of deaf people was calculated as the ratio of their fertility to fertility of their hearing siblings [1].
Results: Data on 83 deaf people and 185 hearing siblings, aged 35-69 years was collected. 143 children accounted for 83 deaf individuals, whereas 185 hearing siblings had 422 children. Fertility of deaf people was estimated as 1.72 vs 2.28 of their hearing siblings. Overall the genetic fitness for deaf individuals is 0.75. There was no difference between genders. Our results are comparable with fitness of deaf women in Sweden – 0.76 [2], and lower than in USA – 0.88 [1] and higher than in Mongolia – 0.62 [3].
Conclusion: Thus, genetic fitness of deaf in Eastern Siberia is slightly reduced, which is still could possibly increase frequency of *GJB2*-deafness.
Acknowledgements: The study was supported by the RFBR grants #15-04-04860_a, #16-34-00564_mol_a, the State project #6.656.2014/K.
References:
1. Blanton SH. et al. Fitness among individuals with early childhood deafness: Studies in alumni families from Gallaudet University Ann Hum Genet. 2010;74(1):27-33.
2. Carlsson PI, Danermark B, Borg E. Marital status and birthrate of deaf people in two Swedish counties: the impact of social environment in terms of deaf community. Am Ann Deaf. 2004-2005 Winter;149(5):415-20.
3. Tekin M. et al. GJB2 mutations in Mongolia: complex alleles, low frequency, and reduced fitness of the deaf Ann Hum Genet. 2010 Mar;74(2):155-64.

EVOLUTION OF MITOCHONDRIAL GENOMES IN BAIKALIAN AMPHIPODS

E.V. Romanova[1]*, V.V. Aleoshin[2], K.V. Mikhailov[2], R.M. Kamaltynov[1], M.D. Logacheva[2], E.A. Sirotinina[1], D.Yu. Sherbakov[1, 3]

[1] *Limnological Institute SB RAS, Irkutsk, Russia*

[2] *A.N. Belozersky Institute of Physicochemical Biology MSU, Moscow, Russia*

[3] *Irkutsk State University, Irkutsk, Russia*

* *Corresponding author: ERA_85@inbox.ru*

Key words: *Baikal, amphipods, mitochondrial genome*

Motivation and Aim: Amphipods of Lake Baikal present the perfect example of adaptive radiation. Along with cichlid fishes of African Rift Lakes and Darwin's finches of the Galapagos archipelago these invertebrates make a good model for studying of speciation process, however in comparison to the former groups, amphipods of Lake Baikal are incommensurably less investigated. Despite previous efforts [Sherbakov et al 1999] the evolutionary history and taxonomy of Baikalian amphipods remains not sufficiently understood. The aim of this study is to infer phylogeny of Baikalian amphipods based on all mitochondrial protein-coding gene sequences and also to look for possible peculiar features of mitochondrial genomes structure in this group of invertebrates.

Methods and Algorithms: Sequencing DNA samples was performed using Illumina systems (Illumina, San Diego, CA). Mitochondrial DNA sequences were assembled both from reads of large mitochondrial DNA amplicons and from reads of total DNA using reference sequence.

Results: We obtained complete and nearly complete mitochondrial genome sequences of nine Baikalian amphipod species. A phylogenetic inference based on both nucleotide sequences of mitochondrial protein coding genes and their amino acid sequences revealed Baikalian species to be a monophyletic group within other amphipods. The amphi-Atlantic-distributed amphipod species *Gammarus duebeni* become the nearest outgroup to Baikalian clade. A structural analysis of mitochondrial genomes showed that Baikalian species under study possess different gene order patterns in comparison to each other and to available non-Baikalian amphipods. Also in mitochondrial genomes of four Baikalian amphipods the duplicated tRNA genes were found. It was discovered that additional tRNA copies in genomes of three species become underwent a remolding (changing of tRNA identity via point mutation in anticodon). Another feature of mitochondrial genomes studied is different numbers and lengths of intergenic spacers that occupy from 1.8 to 21.3 % of their entire length.

Conclusion: Our study supported the hypothesis about monophyly of Baikalian amphipods [Kamaltynov, 1999]. It was also shown the origination of some shallow-water species of *Eulimnogammarus* genus from a deep-water ancestor. The mitochondrial genomes of studied species possess sufficient amount of differences in patterns of genes and non-coding parts that suggest the hypothesis of an intense rearrangements processes in the evolution of mitochondrial genomes of Baikalian amphipods.

References:

1. D.Yu. Sherbakov et al. (1999). On the Phylogeny of Lake Baikal Amphipods in the Light of Mitochondrial and Nuclear Dna Sequence Data, *Crustaceana, 72(8): 911-919.*
2. R.M. Kamaltynov (1999). On the evolution of Lake Baikal amphipods. *Crustaceana*, 72, 921–931.

GENETIC DIVERSITY AND METABOLISM OF THE GARGA HOT SPRING MICROBIAL MAT

A.S. Rozanov*, A.V. Bryanskaya, T.K. Malup, T.V. Ivanisenko, Yu.E. Uvarova, S.E. Peltek
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: sibiryak.n@gmail.com*

Key words: *microbial mat, hot spring, microbial community, metagenomic, metabolism*

Motivation and Aim Geothermal springs located at Baikal rift zone are interesting objects for studying the thermophilic bacteria. In this study we present the data about genetic diversity of phototrophic microbial mat of the geothermal Garga spring (Baikal, Russia).
Methods and Algorithms: We have applied the 16S rRNA metagenomic approach. To this aim, we have used the primers U341F (5'-CCTACGGGRSGCAGCAG-3', where R is A or G, S is G or C) and U806R (5'- GGACTACNVGGGTWTCTAAT-3', V - C or R, W – A or T). NGS sequencing of the variable V3-V4 regions of the 16S rRNA gene was performed on MiSeq (Illumina) using the MiSeq reagent kit v.2 (Illumina). For data analysis we have used the software: QIIME pipeline, USEARCH (Ultra-fast sequence analysis) tool v5.2.236.
Results Metagenomic sequencing of the 16S rRNA gene was performed to analyze the genetic diversity of the microorganisms of the Garga hot-spring. We have studied 4 points with the temperature varying from 74 to 45 °C. Bioinformatics analysis gained more than 13 thousands sequences of post filtering quality for each of the eight studied samples. Total number of sequences for all the points equals to 222201.
Phylogenetic analysis has demonstrated that Archaea (mainly Crenarchaeota) inhabit only a single point with the highest temperature (20%). The majority of the microbial mat was represented by cyanobacteria genus Leptolyngbya. The heterotrophic microorganisms were mainly represented by Actinobacteria and Proteobacteria in all samples of the Garga hot spring. Planctomicetes, and anoxic phototrophs, mainly Chloroflexi, were found in significant amounts in the middle layer of the microbial mat. In the lower part of the microbial mat, the heterotrophic microorganisms were dominated. Representatives of phylum Firmicutes (Clostridia, strict anaerobes) were most frequent. Based on the data about distribution of different Bacterial in microbial mat layers we have made assumption of forming of closed cycles for the basic for life chemical elements: carbon, nitrogen, and sulfur.

NEUROTHROPHIN SIGNALING PATHWAY IN DEVELOPMENT OF ALZHEIMER'S DISEASE-LIKE PATHOLOGY

E.A. Rudnitskaya*, N.A. Muraleva, N.A. Stefanova, N.G. Kolosova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: ekaterina.rudnitskaia@gmail.com*

Key words: *Alzheimer's disease, neurotrophin, senescence-accelerated OXYS rats*

Motivation and Aim: Alzheimer's disease (AD) is the most common type of age-related dementia worldwide, with dramatically increasing incidence because of ageing of the population. However, the precise mechanisms of AD progression are still not fully understood. One of the processes that may contribute to neurodegeneration is age-related alteration of neurotrophic signaling since neurotrophins manage synaptic plasticity, growth of neurite and neuron survival and death. To investigate a link between alterations of neurotrophic signaling pathway (NSP) and progression of AD pathology we used OXYS rats that are considered as a suitable model of AD.
Methods and Algorithms: 20-day-old, 3, 5 and 18-mo-old OXYS and Wistar (control) rats (n = 16) were used. The RNA-seq data obtained for frontal cortex were used to analyze changes in NSP. ELISA was used to quantify level of Brain-Derived Neurotrophic Factor (BDNF) in the hippocampus. Western-blot analysis was used to quantify levels of TrkB and phosphorylated TrkB (phTrkB) receptors in the hippocampus and cortex. Immunohistochemistry was used to localize proBDNF and mature BDNF (mBDNF), TrkB and phTrkB receptors in the hippocampus and cortex.
Results: According to KEGG pathway, 4 genes related to NSP were upregulated in the cortex of OXYS rats at the age of 20 days and 5 mo compared to Wistar rats. However, 18 mo-old OXYS rats had 20 genes related to NSP with differential expression in the cortex. From the age of 20 days to 5 mo expression of 45 genes related to NSP changed unidirectionally in the cortex of OXYS and Wistar rats. With age only 5 genes of NSP changed its expression in the cortex of Wistar rats, while the expression of 54 genes was changed from the age of 5 to 18 mo in the cortex of OXYS rats. Analysis of protein content in the hippocampus showed that BDNF levels were increased in 3-mo-old OXYS rats and decreased with age. In addition, proapoptotic proBDNF became prevailing form of BDNF in the hippocampus of 18-mo-old OXYS rats. As for BDNF receptor, TrkB, its activation (that's mean phTrkB/TrkB ratio) was decreased in the hippocampus of 18-mo-old OXYS rats.
Conclusion: Activation of NSP in the hippocampus and frontal cortex of 3-5-mo-old OXYS rats may be considered as compensatory process addressed to slow down the development of AD-like pathology. However, compensation was ineffective and considerable alterations of NSP occurred at the age of 18 mo that coincided in time with active progression of AD-like pathology in OXYS rats. *This work was supported by grant from the Russian Foundation for Basic Research (project #15-04-06066).*

DEVELOPMENT OF CATARACT AS THE BASIC SELECTION TRAIT IN THE ONTOGENY OF SENESCENCE-ACCELERATED *OXYS* RATS

Yu.V. Rumyantseva*, A.Z. Fursova, E.E. Korbolina, N.G. Kolosova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: rumyantseva@bionet.nsc.ru*

Key words: *accelerated aging, cataract, lens, crystallins, molecular chaperone, small heat-shock protein, Alzheimer's disease, OXYS rats*

A non-transgenic model of accelerated senescence and associated age-related disorders in OXYS rats was established by selection and inbreeding of Wistar rats, which were sensitive to the cataractogenic effects of galactose-enriched diet. The development of cataracts was induced by galactose overconsumption at the start, but, after five generations, the early spontaneous cataract became the selection trait against the background of a normal diet. To date, as a result of carefully controlled selection, OXYS rats stably and spontaneously develop the characteristic accelerated-senescence phenotype, including early cataract (similar to human senile cataract), AMD-like retinopathy, osteoporosis, arterial hypertension, and – according to the recent findings - brain neurodegenerative pathology with the features specific for Alzheimer's disease. These led to the use of OXYS strain in the fundamental research and in the study of drug therapeutic effectiveness. It is widely known, that age-related cataracts are associated with degenerative changes in the ocular lens including the aggregation of proteins, mainly molecular chaperones - crystallins, but also amyloids. This biochemical aspect might be a fascinating hypothesis for a cataract as a "biomarker" of systemic changes, including neurological processes, in the selection of OXYS rats. We recently reported the downregulation of a-crystallin gene expression during retinopathy progression in OXYS rats. So, *the aim* of the present study was to analyze the dynamics of morphological changes in the OXYS lens and to compare it with lens mRNA levels for αA- and αB-crystallins in order to search for potential systemic commonalities between cataract and retinopathy. We examined OXYS rats' lens by means of light microscopy at 20 days (no clinical signs of cataract), 3 months (cataract prevalence is 100 %), and at 12 months (at the pronounced stages of disease) in comparison with age-matched Wistar rats (control group). In the lens of 20-day-old OXYS rats the minor aberrations in the packing of cortical fibers, and the signs of alterations in the transport activity and/or cell-to-cell contacts were detected. The likely-compensatory increase in the density of the lens epithelium was accompanied by upregulation of the αA- and αB-crystallin genes. At the age of 3 months, there were noticeable aberrations (and at 12 months, significantly enhanced aberrations) in the structure of the lens capsule and in organization of the cortical fibers in OXYS rats, whereas a-crystallin expression dipped below than in the Wistar rats. Summarizing, we showed that systemic changes in the expression and function of crystallins may underlie cataract and retinopathy progression in OXYS rats. Thus the selection for "cataract" trait might led to inheritable systemic changes including AMD-like retinopathy and brain Alzheimer's disease-like pathology in OXYS rats.

This study was supported by the Russian Foundation for Basic Research (Grant 14-04-00376 A).

COMPARATIVE ANALYSIS OF EXPRESSION OF ANHYDROBIOSIS-RELATED GENES IN RESPONSE TO DIFFERENT TYPES OF IONIZING RADIATION IN THE SLEEPING CHIRONOMID *(POLYPEDILUM VANDERPLANKI)*

A.V. Ryabova[1], A.V. Cherkasov[1]*, K. Mukae[2], T. Kikawada[3], T. Okuda[3], T. Sakashita[4], O. Gusev[1, 5, 6]

[1] *Institute of Fundamental Biology and Medicine, KFU, Kazan, Russia*
[2] *Department of Regulatory Biology, Saitama University, Saitama, Japan*
[3] *Anhydrobiosis Research Group, NIAS, Tsukuba, Japan*
[4] *Takasaki Advanced Radiation Research Institute, Takasaki, Japan*
[5] *Division of Genomic Technologies, CLST, RIKEN, Yokohama, Japan*
[6] *Preventive Medicine & Diagnosis Innovation Program, CLST, RIKEN, Yokohama, Japan*
* *Corresponding author: urban-nomad@yandex.ru*

Key words: *anhydrobiosis, ionizing radiation, resistance, microarrays*

Motivation and Aim: The larval stage of African chironomid Polypedilum vanderplanki has an extraordinary ability to withstand different types of external stress, such as complete water loss and high-dose radiation exposure. Both hazards cause a similar trend in molecular damage, eventually leading to severe DNA lesions. The main purpose of this study was to clarify cross-tolerance mechanism to desiccation and irradiation by analyzing alterations in expression patterns of the most renowned stress-resistant groups of enzymes, such as antioxidants, late embryogenesis abundant (LEA) proteins and heat-shock proteins.

Methods and Algorithms: For RNA expression analysis larvae were sampled according to the time (in hours) passed from the irradiation (0.5h; 3h; 12h), beginning of desiccation (24h; 48h) and of rehydration (24h). Custom microarrays for P. vanderplanki (4 × 44k format) were prepared by Agilent Technologies, Japan. Probe design was performed using 16652 genes selected from Pv-EST database. Data analysis was accomplished in Subio Platform (v.1.19). Blast2go software was applied for ESTs annotation and Gene Ontology (GO) mapping via BLASTx results running against TrEMBL, Flybase and Wormbase. GO enrichment analysis was carried out with BinGO plugin for Cytoscape platform.

Results: The expression patterns of desiccation-responsive genes demonstrated mainly similar trends in all types of stress exposure. Functional analysis showed a presence of several closely related clusters of enriched biological processes and molecular functions, associated with different metabolic response to unfavorable conditions. Most of enriched GO terms were connected with protein modification and repair or neutralization of reactive oxygen species (ROS).

Conclusion: Evolutionary-based adaptation of *P. vanderplank*i larvae to anhydrobiosis produces an activation of a multiple gene associations for elimination negative effects obtained from other types of abiotic stress impact. Tolerance to ionizing radiation is a side adaptation to desiccation resistance.

Acknowledgements: The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University and Russian Foundation for Basic Research (grant №14-04-01657).

GRAPH DATABASE FOR HUMAN MICROBIOME

A.A. Ryasik, E.A. Temlyakova, M.A. Orlov, A.A. Sorokin*
Institute of Cell Biophysics RAS, Pushchino, Russia
** Corresponding author: arc7an@gmail.com*

Key words: *graph database, omics data, big data*

Motivation and Aim: Human microbiome project (HMP) [1] is an important and perspective work with huge significance in biology and medicine. This project data grow year by year and nowadays it already contains 3055 sequenced and annotated organisms. So there is an obvious need in a tool that can provide convenient way of storing, updating and analyzing such amount of information. We believe that graph representation of biological data is the most natural and convenient way, because graphs have already been successfully used in system biology, bioinformatics etc.
Methods and Algorithms: We develop a graph-oriented storage for omics data using Neo4j [2] database engine. The algorithms for collecting, uploading and updating data was implemented by Python and Scala. We collect data from different sources and databases such as GenBank, UniProt, CheBI, Gene Ontology etc. It is also possible to store profiles of physical characteristics of the genome sequences in our database. All this data is processed and represented as graph, and accessible via native graph query language Cypher from web portal.
Results: Recently we have reimplemented the algorithms of the uploading data from external sources and achieved critical acceleration in upload. Now full upload of the most heavy part - genomics and proteomics data - can be done within hours instead of weeks as it used to be.
Conclusion: As sequencing data is updated every month, consequently new version of graph databases can be updated much faster according to new releases of HMP data. So the releases of our graph database with the most relevant information will be available almost the same time as HMP releases.
Availability: Source code is available on GitHub: http://github.com/promodel/
Acknowledgements: This work was supported by RFBR, grant r_centr_a 14-44-03679.
References:
1. NIH Human Microbiome Project: http://www.hmpdacc.org/catalog/
2. Neo4j web-site (Neo Technology): http://www.neo4j.org/

NEURONAL TRANSCRIPTIONAL REGULATION OF *DROSOPHILA* LIFE SPAN

O.Y. Rybina[1, 2], A.V. Symonenko[1], N.V. Roshina[1], A.V. Krementsova[1, 3], E.R. Veselkina[1], M.I. Schelkunov[4], S.V. Sarantseva[5], E.G. Pasyukova[1]*

[1] *Institute of Molecular Genetics of RAS, Moscow, Russia*

[2] *Moscow State Pedagogical University, Institute of Biology and Chemistry, Russia*

[3] *N. M. Emmanuel Institute of Biochemical Physics of RAS, Moscow, Russia*

[4] *M. V. Lomonosov Moscow State University, Russia*

[5] *B. P. Konstantinov Petersburg Nuclear Physics Institute, Russia*

* *Corresponding author: egpas@rambler.ru*

Key words: *Life span, transcription factors, the nervous system, Drosophila*

Motivation and Aim: The nervous system is responsible for processing information from internal and external sources and propagation of vital clues throughout the organism, which ensures its key role in maintaining structural and functional homeostasis and, hence, in the control of longevity. Main aging pathways known today are not specifically neuronal and function in other tissues. At the same time, the identity of neuronal cells is established and maintained due to the expression of specific neuronal genes. The role of these genes in life span control remains largely obscure. We used genetic approaches combined with q-RT-PCR and RNA-seq analyses to assess effects of genes controlling neuronal transcription factors on life span and to find mechanisms providing their impact on aging.

Results: Earlier, we demonstrated that several genes (*stc*, *Lim3*, and others) that encode transcription factors and are involved in the development of the nervous system affect life span in *Drosophila melanogaster*. The role of these genes in life span control was directly demonstrated by assessing effects of their mutations/reversions and tissue-specific RNA-i knockdown and overexpression on life span and locomotion (a conventional marker of aging). Decreased life span was associated with reduced synaptic function, whereas elevated locomotion – with changes in microtubule network. A naturally occurring polymorphism in the regulatory regions of *stc* and *Lim3* was associated with variation in life span. SNPs associated with life span variation were located in target sites of regulatory proteins, including global regulators of transcription; the functionality of SNPs was directly confirmed in experiments with cell culture and flies. Tissue-specific changes in *stc* and *Lim3* transcription were associated with sex-specific changes in life span. Of interest, changing *Lim3* or *stc* transcription exclusively in embryos affected lifespan of adult flies, indicating either epigenetic inheritance of some functional modifications or setting of some structural properties of the nervous system, which might lead to alterations of the adult life span. Changes in the expression level of transcription factors should lead to systemic changes in transcription of many primary and secondary target genes. Several primary Lim3 target genes specific for the nervous system were revealed. Secondary targets were associated with mitochondrial function and energy homeostasis, which indicates probable molecular mechanisms providing *Lim3* impact on longevity.

Conclusion: Systemic regulation of neuronal transcriptional networks is proposed as one of the mechanisms regulating life span.

SEARCH FOR GENE MUTATIONS THAT CAN POTENTIALLY AFFECT THE SUSCEPTIBILITY TO TUBERCULOSIS

O.V. Saik[1]*, P.S. Demenkov[1], E.U. Bragina[2], M. Freidin[2], A. El-Seedy[3], R. Hofestaedt[4], V.A. Ivanisenko[1]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Research Institute of Medical Genetics SB RAMS, Tomsk, Russia*
[3] *Alexandria University, Alexandria, Egypt*
[4] *Bielefeld University, Bielefeld, Germany*
** Corresponding author: saik@bionet.nsc.ru*

Key words: *susceptibility to tuberculosis*, *SNP, ANDSystem, gene prioritization*

Motivation and Aim: Tuberculosis is one of the most wild spread infectious diseases (World Health Organization et al., 2015). It is known that genetic mutations are important in susceptibility to infectious diseases including tuberculosis (Burgner et al., 2006). This work was aimed on search of gene mutations that can potentially affect the susceptibility to tuberculosis based on an automatic analysis of databases and literature.
Methods and Algorithms: information from databases ClinVar, SNPedia, miRdSNP, lincRNA, GWAS Catalog, dbSNP, dbNSFP, KEGG was used in this work. "dhyper" function from R package "stats" was used to calculate overrepresentation of common genes in tuberculosis and other infectious diseases. ToppGene was used for candidate gene prioritization. ANDSystem (Ivanisenko et al., 2015) was used for automatic extraction of knowledge from PubMed publications about genes associated with tuberculosis.
Results: Gene mutations associated with susceptibility to 64 infectious diseases were analyzed. It was shown for 30 diseases (bacterial, viral, protozoal, fungi and flatworms) that there is a statistically significant overrepresentation of genes with mutations common with tuberculosis. 368 candidate genes were proposed as potentially involved in the resistance to tuberculosis with the use of prioritization tool ToppGene. The Gene Ontology enrichment analysis showed that these candidate genes are involved in the immune response, cell proliferation, apoptosis, and others. It was shown that genes associated with tuberculosis in PubMed publications (according to ANDSystem) are statistically significant overrepresented ($p<0.001$) among predicted candidate genes. 36 candidate genes (for example TNFRSF1A) are presented in the KEGG Tuberculosis pathway and can be the most promising candidates for the genotyping.
Conclusion: 368 genes with mutations that can potentially affect susceptibility/resistance to tuberculosis were proposed based on the automatic analysis of databases and literature.
Acknowledgements. This work was supported by the Germany-Ukraine-Russia VW grant.
References:
1. D. Burgner, S.E. Jamieson, J.M. Blackwell. (2006) Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better?, The Lancet infectious diseases. 6.10: 653-663.
2. V.A. Ivanisenko, O.V. Saik, N.V. Ivanisenko, E.S. Tiys, T.V. Ivanisenko, P.S. Demenkov, N.A. Kolchanov. (2015) ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. BMC systems biology, 9(Suppl 2), S2.

ASSOCIATIVE NETWORKS OF GLAUCOMA AND APOPTOSIS

O.V. Saik[1]*, P.S. Demenkov[1], O.S. Konovalova[2], M.N. Ponomareva[2], N.A. Konovalova[2], N.A. Kolchanov[1], I.N. Lavrik[3], V.A. Ivanisenko[1]
[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[2] Tyumen State Medical Academy, Ministry of Health of the Russian Federation, Tyumen, Russia
[3] Otto von Guericke University Magdeburg, Magdeburg, Germany
** Corresponding author: saik@bionet.nsc.ru*

Key words: *primary open-angle glaucoma (POAG)*, apoptosis*, ANDSystem, gene networks, comorbidity*

Motivation and Aim: Primary open angle glaucoma (POAG) is one of the leading causes of irreversible blindness affecting over 44 million people worldwide (Tham, Cheng, 2016). In POAG the degradation of retinal ganglion cells occurs mostly via apoptosis (Lindner et al., 2015). It is known that some chronic systemic conditions (hypertension, diabetes, obesity) are comorbid to POAG and can serve as risk factors for POAG (Tham, Cheng, 2016). The aim of this study was to propose potential comorbid disease for POAG and to investigate the role of apoptotic genes in the POAG gene network.
Methods and Algorithms: Genes associated with POAG were extracted from OMIM, ClinVar, GWAS catalog, SNPedia databases and ANDSystem (Ivanisenko et al., 2015). Gene networks of POAG and apoptosis were reconstructed by ANDSystem. "dhyper" function from R package "stats" was used to calculate overrepresentation of common genes in POAG and other diseases to asses potential comorbidity. The betweenness centrality of apoptotic genes in the POAG gene network was assessed by R package "igraph".
Results: A list of 96 genes associated with POAG was automatically extracted from databases and scientific publications. The gene ontology enrichment analysis revealed that POAG genes are involved in extracellular matrix organization, angiogenesis, apoptosis and other. For 348 diseases overrepresentation of genes common with POAG was statistically significant. The central role in the POAG gene network of genes involved in apoptosis was shown.
Conclusion: It was shown that genes involved in apoptosis have the central role in the POAG gene network. A list of 348 potential comorbid disease for POAG was proposed.
References:
1. Y.C. Tham, C.Y. Cheng. (2016) Associations between chronic systemic diseases and primary open angle glaucoma: an epidemiological perspective, Clinical & Experimental Ophthalmology.
2. E. Lindner et al. (2015) Analysis of functional polymorphisms in apoptosis-related genes in primary open angle glaucoma, Molecular vision, 21: 1340.
3. V.A. Ivanisenko, O.V. Saik, N.V. Ivanisenko, E.S. Tiys, T.V. Ivanisenko, P.S. Demenkov, N.A. Kolchanov. (2015) ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. BMC systems biology, 9(Suppl 2), S2.

SLEEP OF REASON IN THE ANALYSIS OF THE RESULTS OF RESEARCH ON MATERIALS «PROTEOMIC INFORMATION OFSPRING WHEAT VARIETIES DIFFERING IN RESISTANCE TO INFECTION AFTER *PUCCINIA RECONDITA* INOCULATION»

K.N. Sarsenbayev*, A. Sarsenbayeva
L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
** Corresponding author: kanat-50@mail.ru*

Key words: *proteomics, wheat*

Motivation and Aim: Systemic acquired resistance (SAR) to Puccinia recondite (pathotype TKT/Y) in wheat cultivars were studied by using a proteomic approach.
Methods: The effect of leaf rust on the regionalized in the Akmola region of Kazakhstan spring wheat cultivars was studied in pot experiments. The peptides were analyzed by using nano-HPLC (Agilent Technologies 1200), which is directly related to the ion-trap mass spectrometer (Bruker 6300 series), equipped with nano-electrospray source. The gradient of acetonitrile from 5% to 90%, duration 25 minutes. Voltage of fragmentation 1.3 V.
Results: A total of 104 proteins were identified using a combination of peptide mass fingerprinting (PMF) and MSMS fragmentation. Only detected in the control plants: Fructan 1-exohydrolase w1;w2; w3; Histone H1; NAD(P)H-quinoneoxidoreductase subunit 1,chloroplastic; NAD(P)H-quinoneoxidoreductase subunit H, chloroplastic; NADP-dependent glyceraldehyde-3-phosphate dehydrogenase; Mitochondrial outer membrane porin; Cytochrome c oxidase subunit 2; Arf-GAP with Rho-GAP domain; Elongation factor 1-beta; 30S ribosomal protein S8, chloroplastic; 30S ribosomal protein S7, chloroplastic; Eukaryotic initiation factor iso-4F subunit p82-34; Retinoblastoma-related protein 1; Ubiquitin-activating enzyme E1 1;Cysteine synthase. Only detected in the infected plants: peroxidase; ATP synthase subunit 9, mitochondrial; 50S ribosomal protein L23, chloroplastic; cytochrome b6-f complex subunit 4; Ubiquitin; 50S ribosomal protein L16, chloroplastic; 30S ribosomal protein S3, chloroplastic; DNA-directed RNA polymerase subunit beta; DNA-directed RNApolymerase subunit alpha; glutathione S-transferase 1; Small heat shock protein, chloroplastic; probable non-specific lipid-transfer protein 3. Multiple changes in the activity of enzymes, although the intracellular pool of substrates has been found to not allow such a significant change
Conclusion and Availability: The results suggest the impossibility of plant life without the above proteins and errors in the identification of compounds chromatograph. Interpretation for the identification of proteins and determine the direction of changes activity of enzymes should be based on common sense

MODULATION OF COGNITIVE FUNCTION BY OXIDATIVE DNA BASE LESION REPAIR

K. Scheffler[1], V. Rolseth[2], M.D. Bjorge[2], G. Hildrestrand[2], W. Wang[1], R. Suganthan[1], A. Kusnierczyk[1], C. Neurauter[2], H. Korvald[2], C. Vågbø[1], L. Luna[2], G. Slupphaug[1], L. Eide[2], M. Bjoras[1, 2]*

[1] *Department of Microbiology, University of Oslo and Oslo University Hospital, Oslo, Norway*

[2] *Department Cancer Research and Molecular Medicine, Norwegian University of Technology and Natural Sciences, Trondheim, Norway*

Keywords: *aging, DNA repairs, neurobiology*

Accumulation of oxidative DNA damage has been proposed as a potential cause of age-related cognitive decline. Adult neurogenesis is crucial for maintenance of hippocampus-dependent functions involved in behavior. We have generated DNA glycosylase deficient mice that display cognitive and behavior abnormalities, or altered recovery after ischemic stroke. We showed a decline in post hypoxic-ischemia neurogenesis of neil3-/- mice when compared to wild type mice (Sejersted et al 2011). The numbers of neuronal progenitors and microglia in striatum were reduced and reconstitution of neuronal tissue was decreased. Further, we demonstrated that neil3-/- mice displayed learning and memory deficits and reduced anxiety-like behavior (Regnell et al 2012). It appears that Neil3-dependent repair of oxidative DNA damage in neural stem/progenitor cells is required for maintenance of adult neurogenesis to counteract the age associated deterioration of cognitive performance. Unexpectedly, Neil2 deficient mouse show improved learning and reduced anxiety but no change in global accumulation of oxidative lesions in brain relevant for cognitive function (i.e. hippocampus and amygdala). Furthermore, Neil2 deficient mouse was hyper-resistant to post hypoxic-ischemia neuronal cell death. In another set of experiments Myh and Ogg1 deficient mouse (double knock-out) showed increased activity and less anxiety, although there is no differences in global accumulation of 8-oxoG in any brain region as compared to wild type (Dahl-Bjørge et al 2015). Molecular mechanisms underlying these phenotypes will be discussed.

DISEASE MODELS FOR CANCER TO SELECT CANDIDATE BIOMARKERS AND DRUG TARGET

E. Schwartz[1]*, A. Yuryev[2], C. Ross[3], I. Riz[4] and A. McPherron[1]
[1] *Ami-Go-Science, 5917 Barbados Place, Rockville, MD, USA*
[2] *Elsevier, Rockville, MD, USA*
[3] *Johns Hopkins University, Baltimore, MD, USA*
[4] *George Washington University, Washington DC, USA*
** Corresponding author: elenasch@gmail.com*

Key words: *companion diagnostic tests, cancer drug resistance, Sub-Network Enrichment Analysis, pathway analysis*

Motivation and Aim: The majority of molecularly targeted anti-cancer agents currently on the market or in clinical development are only efficacious in a limited subset of patients diagnosed with a specific type of cancer (typically 5-40%). Another serious problem of current anti-cancer agents is drug resistance. Stratification of patients based on their responses to current therapy identified molecular features with prognostic significance in relation to therapy of choice. However, further co-development of novel drugs together with companion diagnostic tests (CDx) is very important. A better educated choice of therapy that is ensured by successful CDx is expected to enhance the therapeutic response, minimize toxicity or decrease cost of treatment. While pharmaceutical companies have traditionally focused on the selection and development of novel therapeutic agents, it has become clear that CDx and monitoring tests are important components of a drug development plan, and regulatory agencies (e.g. the FDA and EMEA) are beginning to require that these tests be co-developed with the drug. The aim of this work is to establish in vitro models for cancer drug resistance and, together with data-mining of clinical data, develop a therapeutic strategy to avoid drug resistance. It will include CDx - pre-validated set of sensitive and specific biomarkers predicting the therapeutic response to basic therapy – and corresponding specific agent to enhance sensitivity, if indicated.
Methods and Algorithms: We analyzed proprietary and publicly available data sets including GSE66297 and others by using Sub-Network Enrichment Analysis (SNEA) implemented in Pathway Studio 9.0 from Elsevier. We also used Kaplan–Meier survival analysis by exploring the R2 database ('R2: Genomics Analysis and Visualization Platform – http://r2.amc.nl').
Results: In silica we built a library of disease models (the underlying pathways and their multiple inter-connections) that demonstrate a mechanism of platinum-based therapy resistance in patients with different cancers. These pathways were categorized as those that activate multidrug transporters, detox and anti-ROS enzymes and components of autophagy as well as regulators of apoptosis. Key regulators of each of the multiple sub-pathways along with potential biomarkers that can indicate the status of these sub-pathways have been identified.
Conclusions: The proposed system will help to establish a library of pre-validated biomarkers and related drug targets for rapid selection and incorporation into CDx for molecularly targeted cancer therapies. Furthermore, we plan to use a series of cell-based assay systems to verify proposed biomarkers and related drug targets. Our approach can be applied to potentially any chemotherapeutic drug.

HUMAN BLOOD BISPECIFIC ANTIBODIES – NEW BIOCHEMICAL MARKERS OF AUTOIMMUNE DISEASES

S.E. Sedykh*, V.V. Printz, V.N. Buneva, G.A. Nevinsky
[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: sedyh@niboch.nsc.ru*

Key words: *Human blood, antibodies, bispecific antibodies, autoimmune diseases, systemic lupus erythematosus, multiple sclerosis*

Motivation: Antibodies are the products of clonal B cell populations recognizing a single antigen. There is a common belief that IgG molecules presented in biological fluids are monovalent molecules with stable structures containing two identical antigen-binding sites.

Previously we have shown that human milk and placenta as the result of Fab-arms exchange contains bispecific antibodies. This phenomenon was first described for IgG4, but have shown that bispecific antibodies of human milk and placenta contains all IgG subclasses (IgG1-4). Interestingly, bispecific human milk κλ-IgG are presented mostly by IgG1 (74%), IgG2-IgG4 (5-16%) and placenta κλ-IgGs consisted of 43.5% IgG1, 41.0% IgG2, 5.6% IgG3 and 7.9% IgG4. Moreover, human milk contains up to 54% of chimeric κλ-IgG, 17% κλ-sIgA and placenta in average contains up to 15.0% κλ-IgG.

Methods: Using affinity chromatography, SDS PAGE, Western blotting and ELISA we obtained the bispecific IgG fractions from blood of healthy donors, systemic lupus erythematosus and multiple sclerosis patients and determined the content of IgG1-4 subclasses.

Results: Here we show that the serum of autoimmune patients contains significantly higher concentrations of bispecific IgG molecules than in healthy donors. Since the formation of bispecific antibodies may be due to unknown processes occurring in immune system during autoimmune pathology, the presence of such antibodies in the serum of patients with systemic lupus erythematosus and multiple sclerosis may be a new biochemical marker of autoimmune disorders. Factors of human blood, stimulating the antibody Fab-arm exchange and formation of bispecific molecules are not yet established.

Acknowledgements: The reported study was funded by RFBR, according to the research projects 16-34-60066 mol_a_dk, 16-04-00603 a, 16-04-00604 a.

PROTEOMIC ANALYSIS OF HORSE MILK EXOSOMES

S.E. Sedykh*, L.W. Purvinsch, V.N. Buneva, G.A. Nevinsky
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
** Corresponding author: sedyh@niboch.nsc.ru*

Key words: *horse milk, exosomes, proteins, proteomic analysis*

Motivation and Aim. It has been shown that mammalian milk contains exosomes - endocytic membrane vesicles that are secreted by cells of different types and detected in most biological fluids. There are described multiple protocols of exosome isolation from biological fluids. Here we to compare different protocols of exosome purification from horse milk and exosome protein content.
Methods. To purify exosomes from preparations of horse milk we used centrifugation and ultracentrifugation, ultrafiltration, gel-filtration. To elucidate the protein components we used wide range of chromatography, electrophoresis and mass-spectrometry analysis.
Results. The comparison of several exosome isolation methods from horse milk has shown that the most pure preparations of vesicles according to transmission electron microscopy were obtained with a combination of ultrafiltration, ultracentrifugation and chromatography. Using transmission electron microscopy we confirmed the presence of vesicular structures in horse milk preparations and analyzed their morphology. With the immunohistochemical analysis we confirmed the presence of CD63, CD81 receptors in exosomes.
Conclusion. We have shown that the protein components of horse milk exosomes depends on the purification method used.

PHYLOGENETIC ANALYS OF DAHPS II TYPE AMINOACID SEQUENCES

A.I. Semashko*, E.G. Veremeenko, N.P. Maksimova
Belarusian State University, Minsk, Belarus
** Corresponding author: tassigo@gmail.com*

Key words: *3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, aromatic amino acids, phenazine antibiotics, phylogeny*

Motivation and Aim: DAHP (3-deoxy-D-arabino-heptulosonate 7-phosphate) synthase catalyses the condensation reaction between phosphoenolpyruvate and D-erythrose 4-phosphate as the first committed step in the biosynthesis of aromatic compounds in plants, fungi and bacteria. It is known about two types of DAHP synthases[1]. The first type of this enzyme contributes formation of aromatic amino acids, siderophores and quinones. The second type promotes formation of phenazine antibiotics and was discovered in plants, fungi and bacteria in contrast to the first type which was find only in bacteria[2]. That is why the main aim of this research was detection of DAHP synthase II type evolution pathway.
Methods and Algorithms: Sequencing of genes which encodes DAHP synthase II type in Pseudomonas bacteria was carried out. Sequences for plants, fungi, agrobacteria, cyanobacteria and mammals were found in GeneBank database. Alignment and plotting of a phylogenetic tree were made with help of program MEGA (version 6.0) using sequences of unique binding sites of DAHP synthase II type with substrate[3].
Results: The UPGMA-tree shows unexpected results. The first major clade comprises all members of plants: both with cloroplastic and nuclear DAHP-genes. The second clade unites Basidiomycota fungi. The third clade is complex of two following groups: cyanobacteria Mastigocladus and Ascomycota fungi such as Penicillium and Aspergillus. The forth clade unites bacteria such as Agrobacterium and Rhizobium which can interract with plants. The last clade combines cyanobacteria Scytonema, Pseudomonas bacteria and Pantholops. Thus such unusual allocation of different organisms can be explained by the old origin of DAHP-genes and absence of reverse mutation detection. For these reasons it can be observed the join of organisms from various kingdoms in one group.
Conclusion: According to received cladogram it can be assumed that gene which encodes DAHP synthase II type primary appeared in ancestral cyanobacteria. After this part of them gave rise for modern bacteria, cyanobacteria, fungi and plants' chloroplasts. Plants' nuclear genes of DAHP synthase descended from cloroplastic genes by acquisition of introns. Appearance of such genes in grass-feeding mammals genome can be connected with endosymbiotic bacteria which decompose cellulose.
References:

1. Diversity and evolution of the phenazine biosynthesis pathway / D.V. Mavrodi et al. // Appl. Environ. Microbiol. – 2010. – Vol. 76, № 3. – P. 866-879.
2. Light, S.H. The diversity of allosteric controls at the gateway to aromatic amino acid biosynthesis / S.H. Light, W.F. Anderson // The Protein Society. – 2013. – P. 395-404.
3. In Silico sequence analysis and molecular modeling of the three-dimensional structure of DAHP synthase from Pseudomonas fragi / S. Tapas et al. // J. Mol. Model. – 2011. – Vol. 17. – P. 621-623.

TARGETED HIGH-THROUGHPUT SEQUENCING FOR MODY GENES IN WEST SIBERIA

E.V. Shakhtshneider*, E.N. Voropaeva, D.E. Ivanoshchuk, A.K. Ovsyannikova, O.D. Rymar, Y.I. Ragino, M.I. Voevoda
Institute of Internal and Preventive Medicine, Novosibirsk, Russia
** Corresponding author: 2117409@mail.ru*

Key words: *maturity onset diabetes of the young, high-throughput sequencing*

Motivation and Aim: MODY (Maturity-Onset Diabetes of the Young) is an autosomal dominant hereditary disease caused in most cases by mutations in a gene encoding the carbohydrates metabolism [1]. MODY is characterized by the early beginning (up to 35 years). MODY increases the risk of wrong diagnosis and not effective treatment of hyperglycemia [2, 3]. Molecular-genetic diagnostics of this condition is an effective way of reducing of complication of diabetes in young population.
Methods: 16 patients were selected based on the full clinical examination and biochemical analysis (including C-peptide, glycosylated hemoglobin, antibodies to b-cell). We selected targets for high-throughput sequencing HNF4A, GCK, HNF1A, PDX1, HNF1B, NEUROD1, KLF11, CEL, PAX4, INS, BLK, ABCC8 and KCNJ11. These genes were sequenced using the Ion Torrent PJM (Life Technologies) and verified using Sanger sequencing (ABI PRISM 3500, Applied Biosystems). All exons, the intron-exon boundaries and the promoter region of these genes were examined. Human reference version is hg38. Free online protein structure prediction program PolyPhen-2 was used to predict mutational consequence of GCK, HNF1A, CEL and ABCC8 genes.
Results: we found g.44189633T>C, g.44189037G>C, g.44198738G>A, p.Arg37Trp, p.Leu147Val, p.Gly258Cys, p.Trp257Term in the GCK gene (MODY2). We found p.Asn62Ser and p.Met412Val in the HNF1A (MODY3). In one patient we found p.Leu247Pro in the CEL gene (MODY8). In two patients we found p.Ala1457Thr in ABCC8 gene (MODY12).
Conclusion: MODY2 diabetes was confirmed in 11 patients; MODY3, in 2 patients; MODY8, in 1 patient; MODY12, in 2 patients. The identification of a pathogenic MODY genes mutation is important for the correct and definite diagnosis of MODY and helps the clinician to predict the disease course and to initiate the appropriate therapy.
Acknowledgements: the reported study was supported by RSCF, research project No. 14-15-00496.
References:
1. Murphy R., Ellard S., Hattersley A.T. Clinical implication of a molecular genetic classification of monogenic β–cell diabetes. Nature Clinical Practice. 2008; 4 (4): 200-213.
2. Kuraeva T. L., Zilberman L. I., Titovich E. V., Peterkova V. A. Genetics of monegenic forms of diabetes mellitus. Sakharnyi diabet. 2011; 1: 20-28.
3. Dedov I.I. Novel technologies for the treatment and prevention of diabetes mellitus and its complications. Sakharnyi diabet. 2013; (3): 2-10.

THE ROLE OF THE MECHANISMS OF RESISTANCE TO IONIZING RADIATION IN *DROSOPHILA MELANOGASTER* AGING AND LONGEVITY

M.V. Shaposhnikov[1, 2]*, E.N. Proshkina[1, 2], L.A. Shilova[1], D.O. Peregudova[1], S.O. Zhikrivetskaya[3], A.A. Moskalev[1-4]*

[1] *Institute of Biology of Komi Science Center of Ural Branch of RAS, Syktyvkar, Russia*
[2] *Syktyvkar State University, Syktyvkar, Russia*
[3] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*
[4] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*
** Corresponding author: mshaposhnikov@mail.ru, amoskalev@list.ru*

Key words: *radioresistance, ionizing radiation, Drosophila melanogaster, aging, longevity*

Motivation and Aim: The ionizing radiation damages DNA and cellular macromolecules directly or via production of reactive oxygen species. A number of mechanisms protect cell from impact of ionizing radiation. The purpose of this work was to investigate the role of cellular stress-resistance mechanisms including DNA repair, DNA damage response, detoxification of free radicals, and heat shock response in *Drosophila melanogaster* resistance to irradiation and longevity.
Methods and Algorithms: We analyzed the effect of ionizing radiation on the level of expression of stress response genes, and the lifespan of wild-type *Drosophila melanogaster* laboratory line. To elucidate the role of stress response genes in longevity we used mutant lines, that were defected in stress response genes, lines with overexpression of these genes, or flies in which stress signaling pathways were pharmacologically inhibited.
Results: We found that radiation hormesis, radioadaptive response, and hyperradiosensitivity can be observed not only in cell cultures but also at the organism level of *Drosophila melanogaster* by integral indicators such as lifespan. We also showed that the reaction of an organism to irradiation is determined by cellular mechanisms of stress resistance (DNA repair, DNA damage response, detoxification of free radicals, and heat shock response).
Conclusion: Thus, we investigated the roles of some components of cell stress signaling pathways in the lifespan alteration after the irradiation. We found that despite low-dose irradiation affects the level of genes expression predominantly stochastically, the mutations in stress response genes, pharmacological inhibition of their products or genes overexpression play crucial role in radioresistance and lifespan.
Acknowledgements: The study was supported by grant of the Presidium of Russian Academy of Sciences N15-4-4-23.

PATTERNS AND MECHANISMS OF CHROMOSOMAL EVOLUTION INFERRED FROM PHYSICALLY MAPPED GENOME ASSEMBLIES

I.V. Sharakhov[1, 3, 4]*, G.N. Artemov[4], A. Peery[1], X. Jiang[3], A.B. Hall[3], Z. Tu[2, 3], A.N. Naumenko[1], V.N. Stegniy[4], M.V. Sharakhova[3]

[1] *Department of Entomology, Virginia Polytechnic Institute and State University, Blacksburg, USA*

[2] *Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, USA*

[3] *The PhD Program in Genomics Bioinformatics and Computational Biology, Virginia Polytechnic Institute and State University, Blacksburg, USA*

[4] *Laboratory of Evolutionary Cytogenetics, Tomsk State University, Tomsk, Russia*

** Corresponding author: igor@vt.edu*

Key words: *inversions, chromosomes, rearrangements, genomes, evolution*

Motivation and Aim: Polymorphic inversions are highly non-uniformly distributed among chromosomal arms of malaria mosquitoes concentrating in just two autosomal arms. They are associated with ecological, behavioral and physiological adaptations of mosquitoes related to pathogen transmission. The recently sequenced and assembled genomes of 16 anophelines from Africa, Asia, Europe, and Latin America allowed us to investigate the mechanisms and biological significance of chromosomal rearrangements.
Methods and Algorithms: Using a combination of cytogenetic and bioinformatics approaches we mapped the genomic scaffolds to chromosomes of *Anopheles arabiensis, An. stephensi*, *An. funestus*, *An. atroparvus*, and *An. albimanus*. We identified the genomic coordinates for evolutionary breakpoints and conserved synteny blocks and estimated the number of chromosomal inversions between *An. gambiae* and each of the species.
Results: Our study found that contrary to polymorphic inversions, fixed rearrangement accumulated ~3 times faster on the X chromosome than on autosomes. The highest densities of TEs and satellites of different sizes have also been found on the X chromosome suggesting a mechanism for the rapid inversion generation. The high rate of X chromosome rearrangements is in sharp contrast with the paucity of polymorphic inversions on the X in anophelines.
Conclusion: The high density of fixed inversions and the relatively rare occurrence of polymorphic inversions on the X chromosome characterize chromosome evolution in malaria mosquitoes. This finding could be indicative of a greater role of the X chromosome rearrangements in speciation of malaria mosquitoes. The development of high-resolution physical maps in combination with computational approaches identified contrasting biological roles of inversions on the sex chromosome and autosomes.
Acknowledgements: The work was funded by the grants from the National Institutes of Health AI099528 and the Russian Science Foundation № 15-14-20011.

GENOME AND CHROMOSOME EVOLUTION OF MOSQUITOES – VECTORS OF HUMAN DISEASES

M.V. Sharakhova[1, 2]*, V.A. Timoshevskiy, A.N. Naumenko[1], A. Peery[1], G.N. Artemov[2], V.N. Stegniy[2], I.V. Sharakhov[1, 2]

[1] *Virginia Polytechnic and State University and Fralin Life Science Institute, 360 West Campus Dr, Blacksburg, USA*

[2] *Tomsk State University, Tomsk, Russia*

* *Corresponding author: msharakh@vt.edu*

Key words: *genome evolution, physical mapping, mosquitoes, vectors of human diseases*

Motivation and Aims: The knowledge about the genome structure and evolution in mosquitoes may help to better understand their ability to transmit different pathogens. Initial genome studies demonstrated that two mosquito subfamilies *Culicinae* and *Anophelinae* have striking differences in their genome sizes related to the abundance of the transposable elements. However, chromosomal evolution and distribution of different repetitive elements in mosquito genomes remained largely unexplored because of the lack of detailed physical genome maps for most of the mosquito species. This study aims to better understand the genome and chromosome evolution in mosquitoes.

Methods: In addition to *An. gambiae,* physical genome maps for 4 malaria mosquitoes *An. atroparvus, An. albimanus, An. funestus* and *An. stephensi* and vector of viruses *Ae. aegypti* were developed using fluorescent *in situ* hybridization. Genome analysis was performed using VectorBase [1], Repeatmasker [2], TEfam [3] and Tandem Repeat finder [4].

Results: The analysis revealed highly nonuniform rates of gene order reshuffling among the chromosomes. The X chromosome in malaria mosquitoes and sex-determining autosome 1 in *Ae. aegypti*, which is homologous to the X and 2R arm in *An. gambiae*, demonstrated significantly more breaks per megabase than the rest of the autosomes. We found a high abundance of simple tandem repeats in sex-determining chromosomes of mosquitoes suggesting their role in genomic plasticity. Our study also demonstrated that transposable elements (TEs) in *Ae. aegypti* do not follow the pattern in malaria mosquitoes. The majority of the TEs in the malaria mosquito genome are distributed in heterochromatic areas around the centromeres. In contrast, TEs in *Ae. aegypti* chromosomes are mostly spread in euchromatic regions because of their tendency to localize in gene introns. A comparative genomic analysis of *Ae. aegypti* with *An. gambiae* determined that the previously proposed whole-arm synteny is not fully preserved; a number of pericentric inversions have occurred between the two species.

Conclusion: The research tools and information generated by this study contribute to a more complete understanding of the genome organization and evolution in mosquitoes.

Availability:

http://www.vectorbase.org

http://www.repeatmasker.org/

http://tefam.biochem.vt.edu

https://www.tandem.bu.edu/trf/trf.html

DISSECTING VARIANCE HETEROGENEITY IN HUMAN SERUM METABOLOME

S.Zh. Sharapov[1, 2]*, Y.A. Tsepilov[1, 2], J.S. Ried[3], K. Strauch[3, 4], C. Gieger[3], Y.S. Aulchenko[1, 2]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

[3] *Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany*

[4] *Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany*

** Corresponding author: sharapovsodbo@gmail.com*

Key words: *Genome-Wide Association Study, Gene interactions, Variance heterogeneity*

Motivation and Aim: Significant difference in the trait variance between genotypes of particular locus indicates that this locus (usually called as vQTL) is potentially involved in interactions with unknown factors. Based on this assumption variance heterogeneity (VH) test was proposed to identify statistical genetic interactions in the framework of genome-wide association analysis. Identification of genetic interaction can expand our knowledge about genetic control of complex traits.

Methods and Algorithms: Here we applied VH approach to detect genetic interactions in the control of human serum metabolome as a two-step strategy. We analyzed concentrations of 151 human blood serum metabolites from KORA dataset (N=2,901). In the first step we conducted genome-wide VH test. In the second step we identified genetic interactions to find vQTL. All findings were replicated in TwinsUK dataset (N=843).

Results: In the first step we found six significant vQTL effects and replicated three out of them on TwinsUK data. In the second step we found two intra-locus and two inter-locus genetic interactions. One intra-locus interaction was replicated. Further analysis showed that the biggest amount of variance heterogeneity of these vQTL was due to scale effect, which is typical for metabolome. Replicated intra-locus interaction was induced by another variant in same locus with recessive model of effect.

Conclusion: We presented an application of variance-heterogeneity approach to detect epistatic interactions. Our findings indicate the presence of variance-heterogeneity in human serum metabolome. We conclude that variance-heterogeneity approach has moderate potential to find genetic interactions.

Availability: In our study we used “GenABEL” and “VariABEL” packages, which are available from {{http://www.genabel.org}}

THE MITOCHONDRIAL GENE ORDER AND CYTB EVOLUTION IN HYMENOPTERA AND OTHER INSECTS

F.S. Sharko[1]*, A.V. Nedoluzhko[2], S.M. Rastorguev[2], A.A. Polilov[3], K.G. Skryabin[1, 2, 3], E.B. Prokhortchouk[1, 3]

[1] *Institute of Bioengineering, Research Center of Biotechnology RAS, Moscow, Russia*

[2] *National Research Center "Kurchatov Institute", Moscow, Russia*

[3] *Lomonosov Moscow State University, Faculty of Biology 1-12 Leninskie Gory, Moscow, Russia*

** Corresponding author: fedosic@gmail.com*

Key words: *genome, mitochondrial, gene order, Megaphragma amalphitanum, Scydosella musawasensis, phylogenetic analysis*

Motivation and Aim: The modern insect genomes have accumulated a number of nucleotide substitutions during hundred millions years of evolution, which become a major challenge in phylogenetic analysis. Moreover, molecular evolution isn't only single mutations, but it also depends on much bigger structural changes such as gene order rearrangements, insertions, deletions and gene and ever genome duplications. Mitochondrial (Mt) DNA are considered absent of recombination and much faster rate of evolution than nuclear DNA, this molecule is becoming increasingly important for comprehensive evolutionary and population genomics. Development of modern deep sequencing technologies leads to considerable progress in the study of evolutionary processes of plants, animals, fungus and etc. A large number of complete mitochondrial genomes are sequenced, annotated and uploaded to different databases, but nevertheless phylogenomic analysis of these data has its own difficulties and pitfalls especially for arthropods that have a very long evolutionary history.

Results: Here we report two mitochondrial genome: The mitochondrial genome of the smallest known free-living insect *Scydosella musawasensis* and the mitochondrial genome of the parasitic wasp *Megaphragma amalphitanum.* Also we describe two types of phylogenetic analysis in several insect orders, using the nucleotide sequence of *CYTB* gene and the gene order in the mitochondrial genomes of *Dipteria, Coleoptera Orthoptera, Heteroptera, Hymenoptera and Lepidoptera*. For this we have developed software, *mitoSpider*, that searches for specific gene sequences in mitochondrial genomes of the entire NCBI base. Phylogenetic trees which were constructed according to the mitochondrial gene order of within the Hymenoptera and Heteroptera were largely congruent with those were constructed using the nucleotide sequence of *CYTB* gene, in contrast to *Dipteria, Coleoptera, Orthoptera, and Lepidoptera*.

Conclusion: Our study of the mitochondrial genome *Scydosella musawasensis* revealed the absence of trnI gene in the mitochondrial genome, bioinformatics analysis carried out on complete mitochondrial genomes of insects which were deposited in the NCBI database for last years allowed to identify a significant number of insect species with uncommon mitochondrial DNA genes set.

Acknowledgements: This work is supported by the Russian Science Foundation (Grant No. 14-24-00175).

THE INFLUENCE OF SNP RS201381696 OF A TATA BOX IN THE HUMAN *LEP* GENE ON EXPRESSION OF REPORTER GENE *LUC*

E.B. Sharypova[1], E.V. Kashina[1], O.V. Arkova[1, 2], N.P. Bondar[1], T.V. Arshinova[1], P.M. Ponomarenko[3], M.P. Ponomarenko[1], L.K. Savinkova[1]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Vector-Best Inc., Koltsovo, Novosibirsk Region, Russia*

[3] *Children's Hospital Los Angeles, Los Angeles, CA 90027, USA*

Key words: *gene expression, regulation, SNP, TATA-box*

An empirical and computational study of the interaction of TBP with TATA boxes in promoters of reference human genes and with SNP-containing TATA boxes (that are associated with various diseases) allowed us to derive an equation of the equilibrial TBP–TATA binding for analysis of unannotated SNPs of TATA boxes and for prediction of possible functionally important polymorphisms. The identified SNPs of TATA boxes with predicted influence on the affinity for TBP were subjected to comprehensive empirical characterization by molecular biological, biochemical, and biophysical methods. For the first time, it was shown that polymorphisms of TATA boxes—in promoters of the genes associated with various human diseases—alter the TBP–TATA affinity, which at the molecular level is characterized via a change in the velocity of formation and disintegration of TBP–TATA complexes and a change in their half-life. For example, polymorphism –24T→G in the TATA box of the *TPI* gene's promoter (which is associated with an increased risk of neurological and muscular disorders and lowers the TBP–TATA affinity 25-fold) lowers the formation velocity of the TBP–TATA complexes 36-fold, increases the disintegration velocity 1.4-fold, and increases their half-life 1.4-fold. The correspondence between i) changes in the affinity of a transcription factor for an SNP-containing binding site in DNA, and ii) the influence of this SNP on the expression of a reporter gene, is not always observed. The results of our experiments revealed that SNP-35A→G (rs201381696) [6] of the TATA box in the LEP promoter (this SNP lowers the TBP–TATA affinity ~2.9-fold) also lowered the expression of reporter gene LUC more than twofold in the MCF-7 cell line (carcinoma of the human mammary gland epithelium), which is known to strongly express the LEP gene. In the culture of HCT116 colon adenocarcinoma cells (low activity of the LEP gene), no differences in the expression of the LUC gene were detected among the alleles.

This work was supported by RFFI grant # 14-04-00485a, RNF # 14-24-00123, and project # 0324-2015-0003.

THE OCCURRENCE OF SPRING FORMS IN TETRAPLOID TIMOPHEEVI WHEATS IS ASSOCIATED WITH VARIATION IN THE FIRST INTRON OF VRN-A1 GENE

A.B. Shcherban[1]*, A.A. Schischkina[2], E.A. Salina[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of General Genetics RAS, Moscow, Russia*
** Corresponding author: atos@bionet.nsc.ru*

Key words: *allelic variation, vernalization, VRN-1 gene, promoter, first intron, tetraploid, diploid, Triticum, Aegilops*

Motivation and Aim: Triticum araraticum and Triticum timopheevii are the tetraploid species of Timopheevi group (GGAA genome), and the former includes both winter and spring forms with predominance of winter forms, whereas T. timopheevii is considered as spring species. In order to clarify the origin of spring growth habit in T. timopheevii, allelic variability of the VRN-1 gene was investigated in a set of accessions of the both tetraploid species, together with diploid species Ae. speltoides, presumable donor of the G- genome to these tetraploids..
Methods and Algorithms: The plant material included tetraploid wheat species T. araraticum and T. timopheevii (46 and 4 accessions, respectively), and 23 accessions of Ae. speltoides obtained from different collections. DNA was extracted from seedlings and PCR with preliminarily designed PCR primers was performed to detect the presence of dominant or recessive alleles of VRN-A1 and VRN-G1 loci. Amplified DNA fragments were directly sequenced and the obtained nucleotide sequences were analyzed using MEGA4 software package [1].
Results: 3 different alleles were identified with large mutations in the first (1st) intron of VRN-A1, namely, VRN-A1f-del, VRN-A1f-ins and VRN-A1f-del/ins. The first allele with a deletion of 2.7 kb in a central part of intron 1 occurred in a few accessions of T. araraticum and none accessions of T. timopheevii. The VRN-A1f-ins allele with insertion of a 0.4 kb MITE element about 0.4 kb upstream from the start of intron 1 and allele VRN-A1f-del/ins having this insertion coupled with the deletion of 2.7 kb are characteristic only for T. timopheevii. Allelic variation at the VRN-G1 locus includes insertion of a 0.2 kb MITE in the promoter (VRN-G1a) found in a few accessions of the both tetraploid species. Alleles VRN-A1f-del and VRN-G1a have no association with spring growth habit, while in all accessions of T. timopheevii this habit was associated with the dominant VRN-A1f-ins and VRN-A1f-del/ins alleles. The phenogram constructed on the base of the promoter VRN-1 sequences confirmed early divergence (~3.5 MYA) of the ancestor(s) of the B/G genomes from Ae. speltoides.
Conclusion: Spring tetraploid T. timopheevii had a set of VRN-1 alleles both common for two tetraploid species (VRN-G1a), and specific (VRN-A1f-ins, VRN-A1f-del/ins). The latter alleles includes mutations in the 1st intron of VRN-A1 and shares a 0.4 kb MITE insertion near the start of intron 1. We suggested that this insertion resulted to spring growth habit in a progenitor of T. timopheevii and has probably been selected during domestication.

References:

1. Tamura K. et al. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol. 24: 1596-1599.

FREQUENCY OF GERMLINE MUTATIONS GENES CHEK, FANCL AND FANCI PATIENTS WITH BREAST CANCER IN THE REPUBLIC OF TATARSTAN

L. Shigapova[1]*, E. Shagimardanova[1], O. Gusev[1, 4], A. Nikitin[2], M.Gordiev[3]
[1] *Kazan (Volga Region) Federal University, Russia*
[2] *Genetic Laboratory, Federal Research and Clinical Center, FMBA Russia*
[3] *Molecular-diagnostic Laboratory, Kazan Cancer Centre, Russia*
[4] *RIKEN, Yokohama City, Japan*
** Corresponding author: shi-leyla@yandex.ru*

Key words: *hereditary breast cancer, mutation, next-generation sequencing (NGS)*

Motivation and Aim. Breast cancer is the most common form of cancer and the second most common cause of cancer death among women cancer among women in worldwide. In 2014 1639 breast cancer cases were diagnosed in the Republic of Tatarstan. Approximately 5–10% of breast cancer cases might be inheritable, up to 30% of which are due to BRCA1/2 mutations. Recently, new data showing the predisposition to Breast and ovarian cancer due to mutation in other reparation genes obtained. Among these genes are CHEK1/2, FANCL, FANCI. To understand frequency of occurrence of these genes mutation in tatar women population we screened 40 patients with hereditary breast cancer by NGS.
Methods. Targeted gene enrichment was performed using NimbleGen SeqCap EZ Choice (Roche) according to the manufacturer's instructions with further sequencing using an Illumina MiSeq instrument with read length 249 bp from each end of the fragment.
Results. Two FANCL patogenetic mutations c.1099_1100 insATTA and c.C112T were observed in 4 patients. Free mutations FANCI c.A1111G, c.G286A, G3541A, c.C3673T, c.A2604C and 2 mutation CHEK2 c.1100delC, c.A38G were detected in 6 and 2 women respectively.
Conclusions. The study demonstrated that breast cancer individuals in Tatar ethnos possess patogenetic founder mutations in reparation genes with high frequency.
Acknowledgements: The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

IDENTIFICATION OF NUCLEAR GENES CONTROLLING CHLOROPHYLL SYNTHESIS IN BARLEY BY RNA-SEQ

N.A. Shmakov[1, 2]*, G.V. Vasiliev[1], N.V. Shatskaya[1], A.V. Doroshkov[1], D.A. Afonnikov[1, 2], E.K. Khlestkina[1, 2]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *Corresponding author: shmakov@bionet.nsc.ru*

Key words: *barley, near-isogenic lines, chlorophyll synthesis, albinolemma, nuclear genes, gene network, RNAseq, differential expression, IonTorrent sequencing platform*

Motivation and Aim: Albinism in plants is characterized by lack of chlorophyll and results in photosynthesis impairment, abnormal plant development and premature death. These abnormalities are frequently encountered in interspecific crosses and tissue culture experiments. Analysis of albino mutant phenotypes with full or partial chlorophyll deficiency can shed light on genetic determinants and molecular mechanisms of albinism.
Methods and Algorithms: Poly-A RNA was extracted from spikelets of barley of Bowman line and i:Bw*Alm* line with tissue-specific albinism and sequenced using IonTorrent platform. Resulting short read libraries were mapped to *Hordeum vulgare* genome using cufflinks pipeline and STAR mapper. Differential expression search was conducted with cufflinks pipeline and edgeR package. Differentially expressed genes list was examined for enriched gene ontology terms from AgriGO database and significantly represented pathways from PlantCyc database. For a selected list of genes differential expression was checked with quantitative real-time PCR. Phentypic characterization and chlorophyll distribution patterns were examined using chlorophyll fluorescence microscopy. *De novo* transcriptome assembly was performed using Trinity tool.
Results: Microscopic analysis revealed segregation of cells in spikelets to chloroplast-containing and chloroplast-deficient. Our results demonstrated that *alm* mutant has decreased expression level of plastid genes. Statistically significant differential expression was observed for several plastid operons containing protein coding genes, rRNA and tRNA-coding genes. We identified nuclear genes with differential expression in two barley lines. Functional differentiation between genes with higher and lower levels of expression in i:Bw*Alm* line was detected. As was demonstrated with gene ontology analysis, genes with lower level of expression in i:Bw*Alm* line are mostly associated with photosynthesis and chlorophyll synthesis, while genes with higher expression level are functionally associated with vesicle transport. Differentially expressed genes are shown to be involved in several metabolic pathways, most represented being Calvin-Benson-Bassham cycle. Finally, *de novo* assembly of transcriptome contains several transcripts, not annotated in current *H. vulgare* genome version.
Conclusion: Our results provide the new information about genes which could be involved in formation of albino lemma and pericarp phenotype. They demonstrate the interplay between nuclear and chloroplast genomes in this physiological process. *Acknowledgements:* This work was supported by the RFBR (№ 16-34-00924, bioinformatics data analysis), RSF (№ 14-14-00734, microscopic images preparation and analysis).

NICOTIANA GENOMICS: FROM PLANTS TO GENOMES

N. Sierro, J.N.D. Battey, S. Ouadi, N. Bakaher, L. Bovet, A. Willig, S. Goepfert, M.C. Peitsch, N.V. Ivanov*

Philip Morris International R&D, Philip Morris Products S.A., Switzerland (part of Philip Morris International group of companies)

** Corresponding author: nikolai.ivanov@pmi.com*

Key words: *tobacco Nicotiana genome transcriptome plant*

Motivation and Aim: Nicotiana tabacum (common tobacco) is a major crop species and a model organism, and as a Solanaceae shares significant similarities with tomato, potato, eggplant and pepper. The tobacco plant stands out as a complex allotetraploid with a large 4.5 Gb genome, a significant proportion of which is represented by repeats. As a species, *N. tabacum* (2n=4x=48) evolved through the interspecific hybridization of the ancestors of two South American Nicotiana species about 200,000 years ago, *Nicotiana sylvestris* (2n=24, maternal donor) and *Nicotiana tomentosiformis* (2n=24, paternal donor). Efforts to sequence a reference tobacco genome started almost 15 years ago with the Tobacco Genome Initiative, and several key achievements will be briefly described.
Methods and Algorithms: An F2 mapping tobacco population was screened with SSR markers to build a genetic map. A physical map of four bacterial artificial chromosome (BAC) libraries totaling 425,088 clones from Hicks Broadleaf was constructed using Keygene's Whole Genome Profiling (WGP™) technology. Illumina whole shotgun sequencing was used to produce raw sequences of Nicotiana genomes.
Results: A genetic map with 2,317 markers and 2,363 loci was generated using an F2 mapping population derived from the intervarietal cross of Hicks Broadleaf × Red Russian. The tobacco physical map consisted of 9,750 contigs containing 330,632 BACs, and the calculated genome coverage equaled the estimated tobacco genome size. Draft genomes for the diploid Nicotiana species *N. sylvestris* and *N. tomentosiformis* were completed, covering 72-83% of the 2.3-2.6 Gb genomes in 150-250 thousand scaffolds, and in 2014, draft genomes for three varieties of the tetraploid Nicotiana species *N. tabacum* were published, covering 81-84% of the 4.4-4.6 Gb tobacco genome in 150-250 thousand scaffolds. These genomes show both the low divergence of tobacco from its ancestor genomes and display microsynteny with other Solanaceae species.
Conclusion: We anticipate that these genomes will strengthen the use of *N. tabacum* as a versatile model organism for functional genomics and biotechnology applications.
Availability: https://solgenomics.net/organism/Nicotiana_tabacum/genome
References:

1. Sierro N, et al. Whole genome profiling physical map and ancestral annotation of tobacco Hicks Broadleaf. Plant J. 2013 Sep;75(5):880-9.
2. Sierro N, et al. Reference genomes and transcriptomes of Nicotiana sylvestris and Nicotiana tomentosiformis. Genome Biol. 2013 Jun 17;14(6):R60.
3. Sierro N, et al. The tobacco genome sequence and its comparison with those of tomato and potato. Nat Commun. 2014 May 8;5:3833.

COMPARATIVE ANALYSIS OF GASTROINTESTINAL MICROBIOME IN WILD AND DOMESTIC QUAILS

M.N. Siniagina[1], M.I. Markelova[1], E.R. Kirillova[1]*, E.A. Boulygina[1], A.V. Lichoman[2], V.V. Radchenko[3]

[1] *Kazan Federal University, Kazan, Russia*

[2] *Kuban State Agrarian University, Krasnodar, Russia*

[3] *M.M. Shemyakin and Yu.A. Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, Moscow, Russia*

** Corresponding author: kittinwork@mail.ru*

Key words: *microbiome, next generation sequencing, 16S rRNA, quail, probiotics*

Motivation and Aim: The quails, including Japanese quail (*Coturnix japonica*), are an ecologically and economically important natural food resource all around the world [1]. In aviculture the birds are affected by different infections and diseases. Recent investigations have shown that the host's nutritional, physiological, and immunological processes are profoundly connected with the microbiota [2]. The aim of our investigation is to characterize the microbial community of gastrointestinal tract (GIT) in wild and domestic quails to screen the perspective probiotic strains.

Methods and Algorithms: In our research the analysis of 16S rRNA gene nucleotide sequences from 5 regions (crop, gizzard, cecum, ileum, colon) of 5 wild and 5 domestic Japanese quails was performed on the MiSeq (Illumina). Metagenomic data were analyzed by QIIME pipeline with GreenGenes database v.13.8 and RDP Classifier.

Results: The crop in wild birds contains increased amount of *Enterobacteriaceae* (88.4±3.9% vs. 8.8±3.4% in domestic) owing to their wide spreading in plant phylosphere. Domestic quails unlike to wild ones have increased content of *Lactobacillaceae* (19.5±26.2% vs. 2.95±7.5% in wild) due to feeding with compound poultry feed. It is known that lactobacilli have beneficial effect on digestive system because of their antimicrobial features against pathogens [3]. In domestic quails the microbiome consists of comparatively high proportion of *Bacteroidaceae* (9.0±12.9% vs. 2.9±7.2% in wild) that is able to fermentate starchy feed ingredients [4], and greater representation of *Clostridiaceae* (4.04±4.4% vs. 0.68±1.3% in wild), that take part in cellulose degradation. It is worth to note that the amount of *Helicobacteraceae* in GIT in domestic birds (5.96±11.2%) is greater than in wild (0.03±0.09%) indicating the increased risk of inflammatory diseases development of the mucous membranes in the stomach and intestine [5].

Conclusion: In general, microbial composition of GIT in wild quails is more diverse than in domestic ones. All bacterial families that form microbiota of GIT in the domestic birds present in the wild quails, as well. And more comprehensive analysis of microbial community in wild birds can facilitate screening of the probiotics.

References:

1. H. Su et al. (2014) Cultivable bacterial microbiota of northern bobwhite (*Colinus virginianus*): a new reservoir of antimicrobial resistance, *PloS One*, 9(6): 1-11.
2. E.G. Zoetendal et al. (2004) Molecular ecological analysis of the gastrointestinal microbiota: a review, *J. Nutr*., 134: 465–472.
3. P. Lan *et al*. (2004) Effects of two probiotics *Lactobacillus* strains on jejuna and cecal microbiota of broiler chicken under acute heat stress condition as revealed by molecular analysis of 16S rRNA genes, *Immunol*., 48(12): 917-929.
4. B. Tarakanov, T. Nikolcheva (2000) Biology of cellulolytic bacteria used for cellobacterin probiotic preparation, *Reports of the RAAS*, 3: 42-44
5. P. Ruggiero (2010) *H. pylori* and inflammation, *Cur. pharm. design*, 16: 4225-4236.

PHYLOGENETIC ANALYSES OF *MYCOBACTERIUM TUBERCULOSIS* URAL FAMILY BY WGS DATA FROM EURASIA

V. Sinkov[1]*, O. Ogarkov[1, 2, 3], Y. Bukin[4, 5], I. Mokrousov[6], S. Zhdanova[1]
[1] *Scientific Centre of the Family Health and Human Reproduction Problems, Irkutsk, Russia*
[2] *Irkutsk State Medical Academy of Continuing Education, Irkutsk, Russia*
[3] *Irkutsk State University, Irkutsk, Russia*
[4] *Limnological Institute SB RAS, Irkutsk, Russia*
[5] *Irkutsk National Research Technical University, Irkutsk, Russia*
[6] *Laboratory of Molecular Epidemiology and Evolutionary Genetics, St. Petersburg Pasteur Institute, St. Petersburg, Russia*
** Corresponding author: vsinkov@gmail.com*

Key words: *phylogeny, sequencing, tuberculosis*

Motivation and objectives. *Mycobacterium tuberculosis* has a clonal and hierarchical population structure and the Ural genetic family makes up a part of its grand Euro-American lineage. The Ural genotype is phylogeographically delimited to Northern Eurasia and has long been believed to be less virulent and less transmissible than other *M. tuberculosis* families. However, recent studies in different locations in Eastern Europe reported an alarming emergence of Ural strains with multidrug-resistance (Crudu et al. 2015; Vyazovaya et al., 2015).
Methods and Algorithms. Here we analyzed the publicly available WGS data of the 103 Russian, 40 Moldovan and 7 Belarussian Ural genotype isolates extracted from the GMTV database (Chernyaeva et al., 2014) and TB-ARC genome project (Broad Institute). Unique non-synonymous SNPs were studied by in-house Perl-written annotation tool snpMiner2 (Sinkov, 2016). ML-tree was constructed with PhyML 3.0 and divergence time of the major groups was tested in format lognormal relaxed clock with BEAST v.1.8.2.
Results. Three phylogenetically significant clades were identified and tentatively named A, B and C. Clade A consisted largely of strains from Moldova; in Clade B the majority of isolates were from Russia, clade C was represented by strains of all three countries (Russia, Moldova and Belarus). Compared to clades B and C, clade A had significantly greater number of mutations associated with multidrug resistance (MDR) $\chi^2 = 12.7$; $p < 0.01$. We propose that MDR strains of clade A are the most evolutionarily young and the most dangerous group within the Ural genotype and are emerging as a critical source of the epidemic spread of MDR-TB in Eurasia. In addition, 8 Ural-specific unique SNPs in genes *Rv1901 (cinA), Rv1966 (mce3A), Rv1967 (mce3B), Rv2345 (Rv2345), Rv2485c (lipQ), Rv2933 (ppsC)* and *Rv3498c (mce4B)* were found.
References:

1. Crudu V. et. al. (2015) Nosocomial transmission of multidrug-resistant tuberculosis, Int. J. Tuberc. Lung. Dis. 19(12):1520-1523.
2. Chernyaeva, E.N. et. al. (2014). Genome-wide Mycobacterium tuberculosis variation (GMTV) database: a new tool for integrating sequence variations and epidemiology. BMC Genomics. 15(1), 308.
3. TB-ARC initiative, Broad Institute (broadinstitute.org). URL: https://olive.broadinstitute.org/projects/tb_arc.
4. Sinkov V.V. (2016). snpMiner2: vcf files annotation tool. Zenodo. 10.5281/zenodo.51052
5. Vyazovaya A. et al. (2015) Tuberculous spondylitis in Russia and prominent role of multidrug-resistant clone Mycobacterium tuberculosis Beijing B0/W148. Antimicrob Agents Chemother. 59:2349-2357.

DEVELOPMENT OF TISSUE-ENGINEERING CELL-SEEDED CHITOSAN-POLYCAPROLACTONE BLENDS FOR VASCULAR SURGERY

A.M. Smirnova[1, 2, 3]*, I.S. Zakharova[1, 3, 4]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Novosibirsk State Research Institute of Circulation Pathology, Novosibirsk, Russia*
[4] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
** Corresponding author: ann777sm@yandex.ru*

Key words: *tissue engineering, endothelial cells, mural cells, polycaprolactone, chitosan, vascular surgery*

Nowadays, there is necessity of obtaining small-diameter vascular substitutes in vascular surgery. Tissue engineering provides the opportunity to overcome the long-term outcomes of synthetic vascular grafts. The choice of the optimal scaffold and cell source for seeding are key conditions to bring properties of vessel substitute to physiological. Previous publications have shown that a chitosan-polycaprolactone blend is a suitable biodegradable material for tissue engineering [1, 2].

In this study, for the first time, we suggest an efficient method to generate of tissue-engineered chitosan-polycaprolactone blends, cellularized by endothelial and mural cells of human cardiac explants. Cultured on the blended membranes cells demonstrate high levels of proliferation, adhesion and viability; retain their functional properties (taking up ac-LDL, forming tube-like structure in matrigel); maintain specific endothelial (CD 31, vWF) and mural markers (SMMHC, alpha-sma) and antigens and synthesis of extracellular matrix (collagen IV, fibronectin and elastin). In addition, we have found that proliferative properties of cells of human cardiac explants depends on blend ratio and neutralization conditions.

These results suggest that tissue-engineered chitosan-polycaprolactone blends seeded by endothelial and mural cells of human cardiac explants may be potential to development of substitutes for small diameter blood vessels with properties maximally close to the physiological.

The work was supported by the Russian Science Foundation (14-04-00082).

References:

1. A. Sarasam and S. Madihally (2005) Characterization of chitosan-polycaprolactone blends for tissue engineering applications, Biomaterials, vol. 26, no. 27, pp. 5500–8
2. W. Yang et al. (2010) Study on chitosan/polycaprolactone blending vascular scaffolds by electrospinning, Journal of Biomedical Nanotechnology.

DIFFERENTIAL EXPRESSION OF GLYCOLYSIS-RELATED GENES IN HILAR CHOLANGIOCARCINOMA

A.V. Snezhkina[1], D.V. Kalinin[3], M.S. Fedorova[1], O.L. Kardymon[1], I.Y. Karpova[1], A.F. Sadritdinova[1, 2], N.V. Melnikova[1], A.A. Belova[1, 2], M.M. Belyakov[2], O.S. Sudalenko[2], N.N. Volchenko[2], A.Y. Popov[2], K.M. Nyushko[2], A.D. Kaprin[2], B.Y. Alekseev[2], A.A. Dmitriev[1], G.S. Krasnov[1], A.V. Kudryavtseva[1, 2]

[1] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

[2] Herzen Moscow Cancer Research Institute, Ministry of Health of the Russian Federation, Moscow, Russia

[3] A.V. Vishnevsky Institute of Surgery, Moscow, Russia

** Corresponding author: leftger@rambler.ru*

Key words: *Hilar cholangiocarcinoma, Warburg effect, glycolysis, differential gene expression, oncomarkers*

Hilar cholangiocarcinoma (HC) is a malignant tumor with a poor prognosis; most patients with untreated cholangiocarcinoma have a median survival about 6 months. Surgical resection of HC is the only treatment option, but patients frequently present late at advanced stage of the disease and are considered inoperable. Thus, early detection of HC is crucial for treatment and survival improvement.

In this study, the goal was to find the metabolic features of HC that can be associated with HC development and/or its aggressive behavior. Using CrossHub software we have carried out the analysis of the Cancer Genome Atlas (TCGA) project RNA-Seq data derived from cholangiocarcinoma and adjacent normal tissue. We identified 34 genes involved in glycolysis and differently expressed in cholangiocarcinoma. Quantitative analysis of these genes expression in 20 tissue paired specimens of primary HC was performed using real-time polymerase chain reaction (qPCR). We detected up-regulation of HK1, HK2, ALDOA, PKM2, PFKP, PFKM, and ENO2 genes in more than 60% HC samples. The mRNA level of HK3, ALDOB, ALDOC, PGM1, and ALDOC genes was strongly down-regulated in the majority of cases.

HK, HK2, ALDOA, and PKM2 genes encode enzymes that catalyze crucial steps of glycolysis: the first step - conversion of glucose to glucose-6-phosphate and the rate-limiting step - pyruvate- and ATP formation. Increased expression of these genes could indicate about high rate of glycolysis and formation of a large amount of pyruvate in HC. Down-regulation of ALDOB was also found in hepatocellular carcinoma that can demonstrate about their common mechanisms of carcinogenesis and etiology. Decreased expression of HK3 and ALDOC gene were shown for the first time in cancer. It has been reported that PGM1 and ENO3 are up-regulated in human cancers. Our results illustrate dramatically decreased expression of these genes in HC and are not consistent with the literature.

Thus, HC is characterized by disturbance of expression of a number of glycolysis genes at the mRNA level. Our findings suggest several potential markers that may be used for HC diagnosis methods development.

This work was financially supported by grant MK-8047.2016.4 from the President of the Russian Federation. The work was performed using the equipment of EIMB RAS “Genome” center.

THE ROLE OF MIR-9 AND MIR-98 IN THE REGULATION OF HK2 GENE EXPRESSION IN COLORECTAL CANCER

A.V. Snezhkina[1], I.Y. Karpova[1], O.L. Kardymon[1], A.F. Sadritdinova[1, 2], M.S. Fedorova[1], N.V. Melnikova[1], O.A. Stepanov, K.M. Klimina[4], E.N. Slavnova, K.M. Nyushko[2], N.N. Volchenko[2], M.A. Chernichenko[2], D.V. Sidorov[2], D.V. Kalinin[3], A.Y. Popov[2], G.S. Krasnov[1], A.V. Kudryavtseva[1, 2]*

[1] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia
[2] Herzen Moscow Cancer Research Institute, Ministry of Health of the Russian Federation, Moscow Russia
[3] A.V. Vishnevsky Institute of Surgery, Moscow, Russia
[4] Vavilov Institute of General Genetics, Russian Academy of Sciences Moscow, Russia
** Corresponding author: rhizamoeba@mail.ru*

Key words: *miRNA, miRNA regulation, targets, hexokinase 2, glycolysis, colorectal cancer*

Glucose uptake and glycolytic rate increase significantly in cancer cells. Metabolic shift toward aerobic glycolysis is associated with the increased synthesis of glycolytic enzymes, mainly hexokinases (HKs), in many types of cancer. In previous study we had shown opposite effect – the decrease of HK2 gene expression at mRNA level in colorectal cancer (CRC). HK2 is the key glycolytic enzyme that can regulate rates of glycolysis in proliferating cancer cells. Thus, the mechanisms of regulating HK2 gene expression are of particular interest to study. Using CrossHub software we have carried out the analysis of the Cancer Genome Atlas (TCGA) project data to find miRNAs that potentially regulate HK2 gene expression. We identified a number of miRNAs with predicted or validated binding sites in HK2 transcripts (using TargetScan, mirSVR, PicTar, DIANA microT, and miRTarBase) and strong negative expression correlations. Using real-time polymerase chain reaction (qPCR) we estimated expression of these miRNA in CRC. We confirmed increased expression of miR-143-3p, miR-9-5p, and miR-98-5p. Spearman's rank correlation coefficients between HK2 mRNA level and miRNAs expression were r=-0.4, r=-0.32, and r=-0.36, accordingly. Thus, in present study we showed that mRNA of HK2 gene is a potential target for miR-9-5p and miR-98-5p. These miRNAs could mediate the inhibition of the HK2 gene expression in CRC. Our results on miR-143 are consistent with the literature date.

This work was financially supported by grant 14-15-01083 from the Russian Science Foundation. Part of this work was performed using the equipment of EIMB RAS "Genome" center (http://www.eimb.ru/rus/ckp/ccu_genome_c.php)

ANALYSIS OF A POWERFUL CONSTITUTIVE PROMOTER IN CULTURED CELLS OF POLYPEDILUM VANDERPLANKI

Y. Sogame[1, 2]*, Y. Miyata [1, 3], R. Deviatiiarov[4], S. Kikuta[5], R. Cornette[1], O. Gusev[4, 6], T. Furusawa[1], T. Kikawada[1, 7]

[1] *Institute of Agrobiological Sciences, NARO, Japan*

[2.] *JSPS Research Fellow,* [3.] *Center for Biological Resources and Informatics, Tokyo Institute of Technology,*

[4.] *Institute of Fundamental Medicine and Biology, Kazan Federal University, Russia*

[5] *Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology, Tokyo, Japan*

[6] *Preventive Medicine & Diagnosis Innovation Program (PMI), RIKEN, Japan*

[7] *Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Japan*

** Corresponding author: kikawada @ affrc.go.jp*

Key words: *anhydrobiosis, sequencing, CAGE*

Motivation and Aim: Larvae of an African chironomid Polypedilum vanderplanki are known as the only insect having an ability of the extreme desiccation tolerance, anhydrobiosis. As the draft genome analysis (1) of P. vanderplanki has been accomplished, the molecular mechanisms underlying anhydrobiosis has been gradually elucidated. Genome editing technology would be a powerful tool to reveal the mechanisms of anhydrobiosis, but the existing gene expression systems haven't worked in P. vanderplanki so far. In this study, we identified one of the strongest constitutive promoters in a culture cell line of P. vanderplanki, Pv11 cells, to establish effective gene expression system.
Methods and Algorithms: We found two types of constitutive expression promoter. Based on the transcriptome analysis of the larvae, glyceraldehyde3-phosphate dehydrogenase (PvGapdh) promoter was identified. Promoter of Pv.00443 gene in the scaffold no. 121 of the genome database was selected by the CAGE (cap analysis of gene expression) analysis of Pv11 cells. *Results*: The transcriptome analysis of the larvae showed that gene for PvGapdh was the most highly expressed in the larvae. We constructed a novel expression vector containing 5'-upstream region of PvGapdh as a promoter. However, the expression efficiency of the promoter was slightly weaker to express exogenous genes in Pv11 cells. There is a possibility that the difference of the gene expression system between the larvae and the cultured cell line should result in the inadequate expression. To isolate the strangest expressed genes in Pv11 cells, CAGE analysis of Pv11 were performed. As a results, Pv.00443 gene was the highest constitutively expressed gene in Pv11 cells, whereas PvGapdh was moderately expressed in the cell line. We constructed another novel expression vector containing 5'-upstream region of Pv.00443 gene. Consequently, we confirmed sufficient expression of exogenous GFP protein in Pv11 cells with a flow cytometry. *Conclusion*: In Pv11 cells, Pv.00443 gene was the highest constitutively expressed gene. We constructed an effective novel constitutive expression vector using the promoter region of Pv.00443 gene. *Availability*: This novel expression vector will be a practical tool to reveal molecular mechanisms of anhydrobiosis in Pv11.
Acknowledgements: This work was financially supported by a Basic Scientific Research Grant (#140890) from Sumitomo foundation, by a Research Fellowship from the Japan Society for the Promotion of Science (JSPS) for Young Scientists (#13J08784 and 16J09151) and JSPS KAKENHI (16K18827, 16K15073 and 25252060).

DENOVO ASSEMBLY OF NUCLEAR GENOME OF THE SMALLEST INSECT MEGAPHRAGMA *AMALPHITANUM* (HYMENOPTERA: TRICHOGRAMMATIDAE)

A.S. Sokolov[1]*, A.V. Nedoluzhko[2], F.S. Sharko[1], E.S. Boulygina[2], S.V. Tsygankova[2], A.M. Mazur[1], A.A. Polilov[3], E.B. Prokhortchouk[1, 2], K.G. Skryabin[1, 2, 3]

[1] *Federal State Institution «Federal Research Centre «Fundamentals of Biotechnology» of the RAS*

[2] *National Research Centre "Kurchatov Institute"*

[3] *Faculty of Biology, Lomonosov Moscow State University*

* *Corresponding author: sokolovbiotech@gmail.com*

Keywords: *Hymenoptera, Megaphragma amalphitanum, Trichogrammatidae*, *nulcear genome, parasitic wasp*

Motivation and Aim: De novo assembly and annotation of *M. amalphitanum* nuclear genome. Elucidation of miniaturization effects and also gene ontology of obtained assembly. Metagenomic analysis. *Methods and Algorithms:* For de novo assembly we used 204,377,666 paired-end reads (length 150 bp) obtained in the sequencer Illumina HiSeq1500. Reads have been merged using Pear software package (*Zhang et al., 2014*), followed by the error correction using SoapCorrection(*Luo et al., 2012*). De novo assembly was performed using SPAdes (*Bankevich et al., 2012*), which allowed to collect 158,907 contigs (N50 = 6757 bp). With SOAP de novo we collected 107,373 scaffolds (N50 = 10,155 bp). *Results:* Using data from genome sequencing of parasitic wasp *M. amalphitanum*, we conducted a gene ontology study of uncovered genes of the honeybee (*Apis mellifera*), as the closest organism being studied with an annotated genome. We also analyzed as a control genomic reads of large parasitic wasps of the family *Braconidae*: *Cotesia vestalis*, *Fopius arisanus* and *Trioxys pallidus*. This approach did not reveal any genomic regions of *M. amalphitanum*, which could be the key to unlocking the miniaturization process. On this basis, it was decided to conduct the analysis of positive selection using contigs and scaffolds of *M. amalphitanum*, collected de novo. During the analysis of genomic data we isolated gene sequences that encodes chemoreceptors, olfactory Odorant Receptors, flavoring (Gustatory Receptors) and inotropic (Ionotropic Receptors) receptors. Protein sequences of these receptors together with similar sequences from other *Hymenoptera* will later be used in the comparative analysis of the evolution that will make it possible to understand the changes associated with miniaturization. Metagenomic analysis of *M. amalphitanum* identified representatives of the following genera of bacteria: *Corynebacterium*, *Propionibacterium*, *Propionibacterium*, *Chryseobacterium*, *Acinetobacter*, and *Anaerococcus*. We have not found the genera *Wolbachia* (*Wolbachiaceae*) and *Arsenophonus* (*Enterobacteriaceae*) detected previously in a parasitic wasp *Nasonia vitripennis*. *Acknowledgements:* This work is supported by the Russian Science Foundation (Grant No. 14-24-00175).

References:

1. Zhang J, Kobert K, Flouri T, et al. PEAR: a fast and accurate Illumina Paired-End reAd merger // Bioinformatics 2014. 30(5):614-20.
2. Luo R, Liu B, Xie Y, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novoassembler // Gigascience 2012. 1(1):18.
3. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing // J Comput Biol 2012. 19:455-77.

GENETIC CONTROL OF CIRCADIAN RHYTHMS: AN IMPACT OF MOLECULAR CLOCK EXPRESSION PROFILE CHANGES IN LONGEVITY

I.A. Solovev[1, 2]*, E.V. Dobrovolskaya[1], A.A. Moskalev[1, 2, 3]
[1] *Institute of Biology of Komi Scientifice Center of Ural Branch of RAS, Syktyvkar*
[2] *Syktyvkar State University, Syktyvkar*
[3] *Moscow Institute of Physics and Technology (State University), Russia*
** Corresponding author: soloviev@ib.komisc.ru*

Key words: *circadian rhythms, aging, longevity, transcriptome*

Motivation and Aim: Genes of circadian rhythms change their expression during aging of different organisms. We analyzed available transcriptomic data from different online bases and compared circadian genes' expression profile changes in animals. These findings have led us to an idea of normalizing expression profiles of circadian oscillator elements to compensate potential aging-associated changes during all lifespan on Drosophila model. The aim of the present research is to investigate the role of molecular oscillator elements (cry, per, tim, clk, cyc) in aging and longevity mechanisms [1].
Methods and Algorithms: We used standard methods of Drosophila cultivation, demographic methods to investigate the lifespan changes, RU486-inducible UAS-GAL4 system inserted before genes of interest was used as a tool which up-regulates expression.
Results: We overexpressed clk, per, cry, cyc and tim using neuron-specific RU486-inducible system, this resulted in the increase of median life expectancy (10%) in tim- and cry12-overexpessing females. Median lifespan of female fruit flies overexpressing per10 was 5.4% longer than in control group. Noteworthy, overexpression of clk shortened (-10%) only female's lifespan. 4% augmentation of median life expectancy was observed for males overexpressing per24 and cyc.
Conclusion: Thus, our data has shown that compensation of circadian clock genes' potential age-dependent expression decrease (cry, per) in the nervous system during all imago's life as well as hypercompensation of increased levels of other genes (tim, cyc) extends lifespan. The analysis of the literature shows that clock genes modulate the activity of various determinants of aging, which, probably, cause life extension [1].
Acknowledgements: The research is supported by RFBR grant N16-34-00734 and by Grant of Presidium of Ural Branch of RAS N15-4-4-23.
References:
1. I. A. Solovyev, E. V. Dobrovolskaya, A. A. Moskalev (2016) Genetic control of circadian rhythms and aging, Russian Journal of Genetics, 52(4): 393-412.

KYNURENIC ACID-SENSITIZED PHOTOLYSIS OF LENS PROTEINS UNDER ANAEROBIC CONDITIONS

E.D. Sormacheva[1]*, P.S. Sherin[1,2], E.A. Zelentsova[1,2], T.G. Duzhak[1,2], Yu.P. Tsentalovich[1,2], R.Z. Sagdeev[1,2]

[1] *International Tomography Center SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *Corresponding author: kate.sormacheva@gmail.com*

Key words: *crystallins, kynurenic acid, UV-light, cataract*

A mammalian eye lens contains high concentrations of densely-packed proteins, called crystallins. They have no turnover during an individual lifespan and constantly accumulate numerous post-translational modifications (PTMs). Eventually that can result in the loss of the crystallin solubility and the formation of aggregates scattering the incident light, i.e. the development of age-related cataracts. UV radiation is traditionally considered as one of the factors for PTMs. The main chromophores of the human eye lens are kynurenine (KN) and its derivatives. These molecules effectively transfer the energy of absorbed light quanta into the heat, acting as UV filters preventing the eye tissue from photodamages. KNs can undergo thermal and photochemical degradation with the formation of products, some of which are more photochemically active than the parent molecules. One of these products is kynurenic acid (KNA), which exhibits the yield of the reactive triplet species of about 80%. The triplet KNA, ^{T}KNA, can react with proteins leading to their modifications. The goals of this work are: (1) to study the reactions of ^{T}KNA quenching by crystallins and (2) to analyze the crystallin modifications originating from photo-induced reactions.

The aqueous solutions of crystallins in the presence of KNA were subjected to steady-state and time-resolved laser flash photolyses; the analysis of modifications was carried out by gel electrophoresis and mass spectrometry.

The crystallins quench ^{T}KNA via the electron transfer mechanism with the formation of radicals of the tryptophan and tyrosine residues. Reactions of ^{T}KNA with α- and β-crystallins results in the formation of tryptophan and tyrosine radicals, while in the case of γ-crystallin only tyrosine radicals were observed.

The formation of large molecular weight aggregates is the major outcome of KNA-sensitized photolysis of crystallins under anaerobic conditions. Another modification observed for all crystallins is the formation of new absorption band with the maximum at 325 nm corresponding to products, which fluoresce in the near UV region. Mass spectrometric analysis has shown (i) monotonic degradation of crystallin monomers, (ii) degradation of tryptophan and tyrosine residuses, (iii) oxidation of methionine and tryptophan residuses and (iv) formation of an unknown modification -2 Da on the tryptophan residues.

The obtained results clearly show that the photo-induced modifications of crystallins result in significant changes of these proteins that, in turn, can play an important role in cataractogenesis.

Acknowledgements. Authors are grateful for the financial support Russian Foundation for Basic Research (project #16-33-00669).

PARAMETER FITTING INFRASTRUCTURE FOR RULE-BASED MODELLING

O.S. Sorokina[1], A.A. Sorokin[2, 3]*
[1] *Edinburgh University, Edinburgh, UK*
[2] *Institute of Cell Biophysics RAS, Pushchino, Russia*
[3] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*
** Corresponding author: ivanov@bionet.nsc.ru*

Key words: *systems biology, rule-based modelling, parameter estimation*

Motivation and Aim: Rule-based modelling is a new and highly dynamic area of research in systems biology. Model in rule-based systems, such as Kappa language [1] or BNGL [2] provide the compact and compositional description of complex biochemical and signalling networks. The majour obstacle in adoption of this technology is the relative scarcity of tool support because the area is quite young. Recently we have developed RKappa -- parameter exploration and global sensitivity analysis framework for the Kappa language. Here we are presenting the RKappa-based model fitting infrastructure.
Methods and Algorithms: We have used RKappa as a simulation infrastructure and Nelder-Mead, genetic and particle-swarm algorithms to minimize the difference between simulation results and experimental data. Taking into account the stochastic nature of simulation results each parameter set was simulated several times to get correct estimation of the model behaviour.
Results: The fitting procedure described as a KFitProject object in R language. It requires several groups of objects, such as the model definition, the parameter space description and the *in silico* experiment description. All parts of the project are converted into form suitable for simulation in general-purpose computing claster. The performance of infrastructure is demonstrated with simple ring-closure model and the model of bacterial transcription initiation.
Conclusion: We have developed infrastructure for parameter fitting for rule-based models.
Availability: source code is available from GitHub: http://github.com/lptolik/RKappaFit
Acknowledgements This work was supported by RFBR grant r_centr_a 14-44-03679. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/)
References:
1. V. Danos, et al. (2008) Rule-based modelling, symmetries, refinements. *Form. Methods Syst. Biol.*, 5054, 103–122.
2. J. R. Faeder, et al (2009) Rule-Based Modeling of Biochemical Systems with BioNetGen. Methods Mol. Biol., 500, 113–167.
3. A. Sorokin, et al. (2015) RKappa: Statistical sampling suite for Kappa models, In: *Hybrid Systems Biology*, O. Maler, et al (Eds.), 128–142 (Springer).

VIROME ANALYSIS FOR IDENTIFICATION OF VIRUSES IN BAT SPECIES FROM MOSCOW REGION

A.S. Speranskaya[1], E.V. Pimkina[1], I.V. Artyushin[2], M.V. Safonova[1], A.A. Deviatkin[1], K.V. Kuleshov[1], V.G. Dedkov[1], G.A. Shipulin[1]

[1] *Central Research Institute for Epidemiology, Russian Inspectorate for Protection of Consumer Right and Human Welfare, Moscow, Russia*

[2] *Biological Faculty, Moscow State University, Moscow, Russia*

Key words: *virome, sequencing, bat*

The majority of infectious diseases that were discovered during the last few decades are actually zooantroponosis. Bats are widely distributed in the world and recognized reservoirs of many emerging human infection viruses. Analysis of the virome of bats that are distributed in different geographical regions is an actual approach for identifying the new species of viruses that are potentially cause human infection disease. In addition, periodic monitoring of bats populations may provide important information for zooantroponosis control. Numerous studies have described viruses in different bat species from countries of Europe, Asia, Africa but no Russia.

We characterized the fecal virome of 29 wild bats. Fecal samples were collected during 2015 in Moscow Region (Zvenigorod Biological Station, ZBS) from six species: *Myotis dasycneme*, *Myotis daubentonii*, *Myotis brandtii*, *Nyctalus noctula*, *Pippistrellus nathusii*, *Plecotus auritus*. Ectoparasite analysis of animals resulted in mites (24 samples) or mites and fleas (2 samples). The eight bats were healthy and the three animals were not examinated. We used PCR assay targeted on *Astroviridae*, *Coronaviridae*, *Herpesvirus*, *Lyssavirus I*, *Lyssavirus II*, *Caliciviridae* (nairovirus), *Filoviridae*, arenavirus, rotavirus, paramyxovirus. High throughput sequencing analysis was performed using Illumina MiSeq. Data analysis was conducted as described in study Dedkov et al., 2016.

The results revealed that 13 of 29 analyzed bats (45%) contained coronaviruses. After the SARS epidemic (a few years ago) some studies enabled hypotheses of bats as reservoir hosts of coronaviruses. It was demonstrated that 6 of the 15 recognized coronavirus species were only found in bats. In this work, we found that five of six investigated bat species are hosts of different strains similar to known coronaviruses (including porcine epidemic diarrhea virus). The only *P. auritu* (single sample) was free from coronaviruses. Our results demonstrated that ZBS populations of bats are abundant reservoir of coronaviruses.

We also confirmed other viruses that had previously been reported in different bat species from other regions: astroviruses were found in *M. dasycneme* and *M. dasycneme*. The members of herpesvirus and cypovirus genera were detected in *P. nathusii* and *M. brandtii* respectively. The twelve animals were healfy.

Our result revealed the partial genome sequences of two new novel mammalian viruses. The few sequence reads of virome from *M. daubentonii* showed similarities to Ippy mammarenavirus (51% protein identity). Another new virus demonstrated ~71-78% protein identity to rhabdovirus.

Our work provides the first report about the bat viromes in Russia. It should help understanding of the viruses communities present in bat species found near human habitats.

EXPGENE – SOFTWARE FOR ANALYSIS AND PROCESSING OF GENE EXPRESSION DATA

A.M. Spitsina
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: aspitsina@bionet.nsc.ru*

Key words: *DNA microarray, RNA-Seq, bioinformatics software*

Motivation and Aim: Bioinformatics analysis of gene expression molecular mechanisms has a great fundamental importance in various fields of science, particularly in medicine and statistics. Currently, there is a rapid development of genomic and biological technologies, which are leads to the accumulation of large experimental gene expression data in publicly available databases (most popular and free is BioGPS [1], GEO NCBI). Processing of such data requires development of new computer analysis methods, what will allow solve practical biomedical problems - reveal genes, associated with various diseases (cancer, neurodegenerative diseases and other). The purpose of this work is the creation of software for the analysis and visualization of transcriptomic and microarray data, which will be easy to use and multifunctional.
Methods and Algorithms: The Affymetrix GeneChip U133A data on the human genome and genomes of model organisms (mice and rats) were used as test data. The program is written in *Python* language using *JSON* modules, and also popular libraries for processing and visualization text and numerical data (*pandas, numpy, scipy, matplotlib*).
Results: Software package ExpGene has been developed. It includes a set of options to work with a large array of microarray data - preprocessing, statistical analysis of gene expression correlations and visualization. This tool is also designed to work with gene ontologies. It allows search genes in the chromosome loops formed by pairs of CTCF transcription factor binding sites. It is versatile for any type of text databases (allows the user to pre-select processed data). The program has a user-interaction interface (menu) and is easy to handle even by an inexperienced user.
Conclusion: Using this program we performed a comparative analysis of different samples of genes [1], such as genes from gene networks annotated in the ICG SB RAS regulating cholesterol levels and circadian rhythm. We also studied genes, which are responsible for aggressive behavior of laboratory animals (rats). Gene expression correlation matrices for gene lists were reconstructed as basis for qualitative analysis of the gene network studied.
Availability: Software is available from the author upon request.
References:

1. A.M. Spitsina et al. (2015) Supercomputer analysis of genomics and transcriptomics data revealed by high-throughput DNA sequencing, Program systems: theory and applications, 6:1(23): 157–174. (In Russian)

PREDICTING SMALL RNAS FROM BACTERIAL GENOME

T. Stankovic*, J. Guzina, M. Nikolic, M. Djordjevic
[1] *Faculty of Biology, University of Belgrade, Serbia*
** Corresponding author: tamara.stankovic@bio.bg.ac.rs*

Key words: *CRISPR/Cas, small RNAs, transcription start site predictions, terminator predictions*

Motivation and Aim: Small RNAs perform important regulatory roles in bacteria. Their discovery is however complicated by still limited availability of dRNA-seq data in bacteria, and by the fact that they are often expressed under non-standard and poorly characterized conditions, which may not correspond to those of the experiment. Furthermore, small RNAs are often poorly cфonserved even between related bacterial strains, which complicate their computational discovery. As an alternative, small RNAs can be detected directly from the genome sequence, which requires accurately predicting transcription start site (TSS) and terminator signals. However, two main limiting factors in this approach are generally insufficient accuracy with which TSS are predicted in genome, and the fact that the terminator prediction parameters are trained on *E. coli* data, which may lead to suboptimal predictions in other bacteria. The aim of this work is improving TSS predictions, and assessing if the parameters for terminator predictions can be retrained to allow accurately predicting a specific group of small RNAs.
Methods: For TSS predictions, we start from accurate alignments of the promoter elements for sD.S. Johnson et al. (2008) Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets, Genome Res. 18: 393-403.
Z.D. Zhang et al., Modeling ChIP Sequencing In Silico with Applications, PLoS Computational Biology (2008), 4: 1-10.

IDENTIFICATION OF BACILLUS PUMILUS GROUP STRAINS BY MALDI TOF MS USING GEOMETRIC APPROACH

K.V. Starostin, E.A. Demidov, A.V. Bryanskaya, V.M. Efimov, A.S. Rozanov, S.E. Peltek
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Key words: *microbiology, mass spectrometry*

Geometric approach is based on representing mass spectra as points on a multidimensional space, which allows us to use geometric distances to compare the spectra. The aim of this study was to test if the geometric approach could be used to identify closely related species.

As a model for testing this approach, we used 24 closely related strains of the B. pumilus group from the collection of the Institute of Cytology and Genetics SB RAS, which had over 98% sequence similarity for the 16S rRNA gene. The studied strains were isolated from various extreme habitats in various regions of Russia, including thermal springs, saline lakes, complex ore deposits, etc.

16S rRNA sequences were used to validate the results of mass spectrometry analysis. Two groups referred to as the A and P groups were detected in our dataset with the bootstrap support of 96. The A group included strains isolated from Kamchatka thermal springs, rhizosphere of higher plants from the Novosibirsk oblast, as well as the B. altitudinis type strain (AJ831842). The P group contained strains from saline lakes of the Novosibirsk oblast, complex ore deposits of the Kemerovo oblast, and type strains of B. pumilus (AY456263) and B. safensis (AB681259).

The mass spectra centroids of the studied strains were separated into two groups, which was confirmed by the Welsh's t-test. These groups correspond to the two clusters detected on the phylogenetic tree. The dendrogram constructed by the Ward method using all 23 coordinates also yields similar results. We performed an additional analysis of mass-spectrometry data using Biotyper 3.0 as an extra check. All three methods used allowed us to reliably distinguish between two groups that correspond to two species, B. pumilus (P) and B. altitudinus (A).

The obtained reference database was used for identification of the studied microorganisms by wet-lab experiments. Identification accuracy was 98% for Jaccard coefficients and 100% for Euclidean distances.

CHANGES IN THE BRAIN TRANSCRIPTOME OF OXYS RATS AS THE SIGNS OF ALZHEIMER'S DISEASE DEVELOP AND EFFECTS OF SKQ1

N.A. Stefanova*, N.I. Ershov, N.A. Muraleva, N.G. Kolosova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: stefanovan@mail.ru*

Key words: *Alzheimer's disease, RNA-Seq, senescence-accelerated OXYS rats*

Motivation and Aim: Alzheimer's disease (AD) is a progressive, age-dependent neurodegenerative disorder, featuring progressive impairments in memory, cognition, and ultimately leads to death. In this study, using nontransgenic OXYS rats that simulate key aspects of sporadic AD, we aimed to compare the gene expression profiles of the prefrontal cortex from OXYS rats and Wistar rats (as control) to identify the molecular mechanisms and the factors underlying disease progression. The transcriptome analysis was conducted at three stages of the disease (pre-symptomatic, symptomatic and progressive stage) in OXYS rats, using RNA-Seq technique. In addition, we set out to determine a role of mitochondria in the disease pathology. OXYS rats were treated with specifically target mitochondria antioxidant SkQ1 from age 12 to 18 months, that is, during active progression of AD-like pathology in these animals.

Results: The development of the signs of AD in OXYS rats takes place during changes in mRNA expression of the genes that are mostly related to neuronal plasticity, calcium homeostasis, hypoxia, mitochondrial dysfunction, immune processes, and apoptosis. Between ages 20 days and 5 months, that is, during active changes in cognition, > 5,000 genes undergo changes in expression in both rat strains. In OXYS rats, with age, >5,500 genes change their expression, whereas in Wistar rats, only 499 genes show changes. Most of these genes have something to do with phosphorylation, neurogenesis, synaptic plasticity, and the immune system. The similarity of the mechanisms of accelerated aging of the OXYS rats' brain with the pathogenesis of this disease formed the basis for comparative analysis of transcriptomic alterations in the cortex in human AD and in OXYS rats. We found that changes in the expression of 219 genes are shared between AD patients and OXYS rats; these genes are mostly related to mitochondrial dysfunction, synaptic plasticity, and the calcium signaling pathway. We next studied the hippocampal transcriptome of 18-month-old OXYS and Wistar rats and of OXYS rats treated with SkQ1 from age 12 to 18 months, that is, during active progression of AD-like pathology in these animals. In the RNA-Seq results, SkQ1 decreased differential expression of genes in the hippocampus of OXYS rats; these effects of the drug may have something to do with improved mitochondrial function and normalization of a wide range of cellular signaling processes. This notion is supported by the results of evaluation of SkQ1's effects on progression of the key signs of Alzheimer's disease in OXYS rats: treatment with SkQ1 significantly improved mitochondrial and synaptic deficits, prevented neuronal loss and retarded structural neurodegenerative alterations, decreased amyloid β levels, and attenuated the memory deficits.

Conclusion: Thus, we obtained convincing evidence that the OXYS strain of rats matches the main criteria of the sporadic form of AD and holds promise for research into the etiology and pathogenesis of this disease and for studies on therapeutic and prophylactic interventions. This work was supported by grants from the Russian Foundation for Basic Research (projects ## 15-04-01938 and 15-04-06066)

ULTRASTRUCTURAL ANALYSIS OF MITOTIC DIVISION IN DROSOPHILA S2 CELLS

A.A. Strunov[1, 2]*, L.V. Boldyreva[2], A.V. Pindyurin[2], M. Gatti[2, 3], E. Kiseleva[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russia*
[3] *Sapienza University of Rome, Rome, Italy*
**Corresponding author: strunov.anton@gmail.com*

Key words: *TEM mitosis microtubules S2 cells k-fibers Drosophila*

Motivation and Aim: Drosophila S2 cells are a well-known system for the analysis of mitosis. These tissue culture cells have been used to conduct several RNA interference-based genome-wide screens aimed at the identification of mitotic genes. However, very few data are currently available on the ultrastructure of dividing S2 cells. Since it is virtually impossible to synchronize insect cells, obtaining S2 cells mitoses for TEM analysis is a long and difficult process. We thus developed a new approach for collecting dividing S2 cells for TEM analysis, which allowed us to obtain novel information on the organization of spindle microtubules (MTs).
Methods and Algorithms: We used asynchronously growing *Drosophila* S2 cells, which were pelleted by centrifugation, fixed and sectioned for TEM. Sections were examined for MT organization at different stages of mitosis.
Results: We found that MT organization is highly dynamic during mitosis. MTs mostly appear as single elements in early prometaphase; they then form bundles (generally containing 8-30 MTs) in late prometaphase and metaphase; these bundles interact with the kinetochores either on an end-on fashion (k-fibers) or laterally (lateral bundles). In anaphase, the MT bundles become irregular and the spaces between MTs are filled with proteinaceous material. In telophase, MTs converge to form a midbody, a large MT bundle that mediates cytokinesis.
Conclusion: Our data provide detailed information of microtubule organization during mitosis of *Drosophila* S2 cells and reveal that during late prometaphase and metaphase MTs form bundles that are end-on or laterally attached to kinetochores. A similar MT organization has not been observed in mammalian cells.
Acknowledgements: This study was supported by the grant from the Ministry of Education and Science of Russian Federation no. 14.Z50.31.0005.

ULTRASTRUCTURAL ANALYSIS OF SPINDLE AND KINETOCHORES IN AUGMIN-DEPLETED *DROSOPHILA* S2 CELLS

A.A. Strunov[1, 2]*, L.V. Boldyreva[2], A.V. Pindyurin[2], M. Gatti[2, 3], E. Kiseleva[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Molecular and Cellular Biology, Novosibirsk, Russia*
[3] *Sapienza University of Rome, Rome, Italy*
* *Corresponding author: strunov.anton@gmail.com*

Key words: *mitosis microtubules Dgt6 RNAi spindle ultrastructure*

Motivation and Aim: Spindle formation in S2 cells exploits microtubules (MTs) nucleated by the centrosomes, MTs that grow from the kinetochores, and MTs nucleated by γ-tubulin ring complexes (γ-TuRC) associated with the lateral walls of preexisting MTs. γ-TuRC association with MTs is mediated by the augmin complex, which interact with kinetochore proteins and is primarily required for kinetochore driven MT growth. RNA interference-mediated augmin depletion results in defective kinetochore fibers, metaphase arrest, and in many cells with ana-telophase spindles that contain chromosomes with joined sister chromatids. These peculiar mitotic figures, we call pseudo ana-telophases (PATs), exhibit high Cyclin B levels and are therefore metabolically in metaphase. To further define the mitotic role of augmin we analyzed the spindle ultrastructure in augmin-depleted S2 cells.
Methods and Algorithms: We performed RNA interference (RNAi) in S2 cells against the Dgt6 gene that encodes a component of the augmin complex. Cells were treated for 5 days with Dgt6 double-stranded RNA, fixed and sectioned for TEM analysis.
Results: TEM showed that in Dgt6 RNAi cells the MT bundles associated with kinetochores, either end-on or laterally, are thinner than in controls and sometimes slightly curved, a phenotype never observed in control cells. We also observed several PATs with an elongated shape and the chromosomes at the center of the cell. The kinetochores of these cells often showed a poorly defined and extended structure. In addition, some of the chromosomes at the center of the PATs were partially coated with a double membrane just like the telophase chromosomes of control cells undergoing nuclear envelope reassembly.
Conclusion: Our TEM analyses uncover new details of the mitotic phenotype of Dgt6-depleted cells and characterize the PATs at the ultrastructural level.
Acknowledgements: This study was supported by the grant from the Ministry of Education and Science of Russian Federation no. 14.Z50.31.0005.

MYC GENE FAMILY IN CEREALS: TRANSFORMATION IN THE COURCE OF THE *TRITICUM* AND *AEGILOPS* GENERA EVOLUTION

K.V. Strygina[1]*, E.K. Khlestkina[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: pushpandzhali@bionet.nsc.ru*

Keywords: *Aegilops, bHLH, gene divergence, gene duplication, flavonoids biosynthesis, MYC, Triticum, transcription factor*

Motivation and Aim: MYC family transcription factors are of an essential part of the regulatory complex «MYB + MYC + WD40», which is necessary for gene activation in flavonoid biosynthesis. In bread wheat (*Triticum aestivum* L., BBAADD, 2n=6x=42) the gene *TaMyc1* controlling the synthesis of flavonoid pigments in wheat seed was earlier isolated and characterized [1]. The aim of the current study was to identify, characterize and compare full-length sequences of duplicated homoeologous and paralogous copies of the *TaMyc1* gene.

Methods and Algorithms: The search of homologous sequences was made in databases for not annotated wheat sequences using BLAST. The cluster analysis using the MEGA software was based on UPGMA algorithm. The nucleotide substitutions rate (k) for *Myc* was calculated by the formula $Ks/2T$. The obtained value ($k=10.04\times10^{9}$) was used for the calculation of divergence time of duplicated copies.

Results and discussion: We identified 10 copies of the gene *TaMyc1* in a common wheat genome and 22 *Myc*–like genes in the genomes of related species (*T.durum*; *T.urartu*, *T.monococcum*, *Aegilops speltoides*, *Ae.sharonensis*, *Ae.tauschii*). Analysis of genetic similarity showed that the first duplication of *Myc* gene was in the diploid common ancestor of the tribe Triticeae. The duplication has undergone from two to four further acts of duplication in *Triticum* and *Aegilops* genomes. Time of occurrence of each new copy is calculated and presented in the report. Maintaining functional duplicated genes is likely due to their specialization. It is assumed that *TaMyc1* copies may be involved in the synthesis of various flavonoid compounds in different parts of the plant.

Conclusion: Polyploid genome of bread wheat carries at least 11 copies of the *TaMyc* gene involved in flavonoid biosynthesis regulation. The duplications of this gene occurred several times in the course of evolution of diploid wheat progenitors. The exon-intron organization of these genes is the same to the *TaMyc1* structure. All available sequences have a conserved bHLH domain. None significant change in the motive, which could lead to changes in gene function, has been identified among the annotated *Myc* genes of *T. aestivum* and related species.

Acknowledgements: This study was partially supported by the RAS MCB Programme (0324-2015-0016) and by RFBR (16-04-00912).

References:

1. Shoeva, O. Y., Gordeeva, E. I., Khlestkina, E. K. 2014. The regulation of anthocyanin synthesis in the wheat pericarp. *Molecules*, *19*(12), 20266-20279.

IN SILICO MODELLING OF EXPERIMENTAL CHIP-SEQ PROCESS

T. Subkhankulova*, F.M. Naumenko[2], Y.L. Orlov
[1] *Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK*
[2] *Novosibirsk State University, Novosibirsk Rissia*
** Corresponding author: subkhankul@hotmail.com*

Key words: *transcription factor binding sites, modeling, ChIP-seq*

Chromatin immunoprecipitation followed by next generation sequencing (ChIP-Seq) is recognised as an extremely power tool for study of protein–DNA interactions at a genome-wide scale. Identifying gene regulatory elements and the epigenetic modifications is crucial for reconstruction of the dynamic interplay between the epigenome and gene regulatory networks. ChIP-Seq technique is based on isolation of protein–DNA complexes followed by massively parallel DNA sequencing. Output ChIP-Seq data are analysed using a number of computational steps including *alignment of reads to genome and peak calling analysis*. Currently, more than a hundred of mapping and peak calling software together are developed. There is a high inconsistency between the results produced by different analysing programs. Therefore, the choice of appropriate software to achieve the highest standard in ChIP-Seq analysis becomes an immensely challenging task as well as optimisation of ChIP-Seq experimental conditions such as chromatin shearing, chromatin precipitation steps, the amounts of cells taken for analysis, pre-amplification of sheared chromatin, etc. The efficiency and precision of ChIP are typically assessed using spike controls or limited number of positive/negative loci followed by RT-PCR analysis, or using more advanced techniques, e.g. by placing tags onto the genome according to particular assumed distributions for the actual binding site. However, little consideration has been given to *in silico* modelling of whole process of ChIP-Seq with controlled parameters of the process and *a priori* known input.

Here, we developed unique software, isChIP, capable to fully model ChIP-Seq data in silico. As input files for isCHIP we have been using a bed-files generated from "true" biding sites well known for transcription factors or assumed to be true for the modelling purposes. We validated isChIP, demonstrating that it closely approximates real ChIP-Seq experiment: (i) location of simulated peaks coincides with experimental enriched peaks, (ii) the shape and density of the model peaks resemble the real peaks, and (iii) modelled input files represent unspecific background similar to experimental input data.

Using isChIP we have briefly compared two different mapping programs (Bowtie and BWA), and in more details a few peak calling software (CCAT, MACS, PeakRanger, Qeseq). Our analysis revealed that the tested mapping programs are prone to some byes in alignment of reads to the indexed genome. Analysis of four programs showed that the peak calling algorithm was much more essential in localization of "true" peaks rather than parameters adjusted in each program. IsChIP is applicable for other tasks such as estimation of optimal DNA shearing size, minimal cell number sufficient for successful ChIP-Seq experiment in given conditions, necessity and levels of pre-amplification prior the library construction. Finally, developed *in silico* modelling approach may serve as cost and time effective ChIP-Seq optimization.

References:

1. D.S. Johnson et al. (2008) Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets, Genome Res. 18: 393-403.
2. Z.D. Zhang et al., Modeling ChIP Sequencing In Silico with Applications, PLoS Computational Biology (2008), 4: 1-10.

SINGLE CELL EXPRESSION PROFILING OF NEURAL CREST-DERIVED CELLS

T. Subkhankulova[1], G. Aquino[2], A. Rocco[2], H. Schwetlick[1], R.N. Kelsh[1]
[1] *Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK*
[2] *Department of Microbial and Cellular Sciences, University of Surrey, Guildford GU2 7XH, UK*
** Corresponding author: ts723@bath.ac.uk*

Key words: *expression profiling, single cell, neural cells*

The zebrafish neural crest is now an established model for studies of cellular and genetic mechanisms underlying embryonic development in vertebrates. Neural crest (NC) cells are multipotent stem cells that form from the edge of the neural plate, but which delaminate and then form a great diversity of derivative cell-types. The NC is a major model for understanding stem cell differentiation, in particular how multipotent progenitors generate a balance of different derived cell-types. In this context, we use zebrafish pigment cell development as our model system, primarily because pigment cells are self-labelling so that simple microscopy allows observation of subtle phenotypic effects *in vivo*. Furthermore, in contrast to mammals, zebrafish have three NC-derived pigment cell-types – melanocytes (black), iridophores (silver) and xanthophores (yellow), all of which are thought to share a common cellular origin. We have developed an iterative approach which combines mathematical modelling with *in vivo* genetic studies to generate and formally assess gene regulatory networks (GRNs) in stem cell development. Following our initial studies of the melanocyte GRN (1), we are now focused on understanding how different pigment cell types can be generated from a common NC precursor. Specifically, we are focusing on the shared origin of melanocytes and iridiophores from a shared melanoiridoblast. We are building a quantitative model of the GRN underlying melanocyte and iridophore fate specification. To achieve this goals we have profiled the expression of single cells derived from the NC throughout a time-course of pigment cell development (18 to 72 hpf). For each time point, NC–derived GFP-positive cells from transgenic embryos were individually isolated using FACS; and expression levels for 45 genes (including known core neural crest/pigment cell genes such as *mitfa, ltk, sox10, foxd3, tfec*) were assessed using nanoString technology. Our ongoing analysis of these data give us a unique view of the diversity and quantitative gene profile of single NC cells freshly *ex vivo*, allowing unequalled insight into the state of neural crest cells *in vivo*.. Clustering analysis of the expression profiles identifies cells that show an expression pattern consistent with them being common precursors for both melanocytes and iridophores. Other cell clusters possess characteristics of more multipotent NC cells, including likely common precursors for both pigment cells and other cell types (e.g. glia or enteric neurones). Later in development, we observe cells with profiles consistent with differentiated cell-types e.g. melanocytes or iridophores. We are now developing single cell expression profiles for NC of mutants (*mitfa, sox10* and *ltk)* known to affect melanocyte or iridophore fate specification. Likewise, using single cell expression data for parameter fitting, we are now developing a quantitatively accurate mathematical model of the melanoiridoblast GRN. Initially, the parameters are fitted using the wild-type data, and the resultant quantitative model is being used to generate quantitative predictions of the mutant phenotypes. At the end of this process, we expect to have a refined GRN for the pigment cell progenitor and will be able to test the prediction that stochastic variation in gene expression levels could play an essential role in cell fate decision making.

MOLECULAR PHYLOGENETIC ANALYSIS OF THE GRASSHOPPERS OF FAMILY ACRIDIDAE BASED ON SEVERAL MITOCHONDRIAL AND NUCLEAR MARKERS

I.S. Sukhikh[1]*, A.G. Blinov[1], A.G. Bugrov[2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Systematics and Ecology of Animals SB RAS, Novosibirsk, Russia*
** Corresponding author: igor3419@gmail.com*

Key words: *Acrididae, phylogeny, mitochondrial DNA, ribosomal DNA*

Motivation and Aim: Acrididae is the biggest family of Acridoidea superfamily, including from 25 to 34 subfamilies according to different classifications, and several genus that assigned to no family. It is worth noting that the few subfamilies consists of small number of genus and species. For a long time, taxonomy, the renovation of phylogenetic relationships and understanding of the evolution of Acrididae family were based mainly on comparison analysis of key morphological structures of recent and fossil species. However, on subfamily level, the problems of convergence and parallelism are rising up for these insects. One of the most effective methods of establishing phylogenetic relationships is the analysis of various DNA markers.
Methods and Algorithms: In present work, we conducted molecular phylogenetic analysis of complete mitochondrial sequences, mitochondrial (COI, COII, CytB) and nuclear (ITS2, 28S rRNA) markers of more than 220 species of Acrididae family from 26 subfamilies, extracted from these insects experimentally and obtained from NCBI database. Phylogenetic trees for different combinations of these markers were obtained using maximize likelihood and Bayesian methods.
Results: As the result of the work done, we found out that all of locusts' species under discussion grouping up into 13 phylogenetic clusters, and phylogenetic relationships between them were established. Five out of 26 subfamilies (Acridinae, Oedipodinae, Gomphocerinae, Oxyinae and Catantopinae) found to be polyphyletic in present study, which point out the shortcoming of current taxonomy. The data obtained allowed us to clarify the classification of Acrididae species.
Availability: All DNA marker sequences will be in a free access in the NCBI database.

COMPREHENSIVE ANALYSIS OF DRAFT GENOMES OF TWO CLOSELY RELATED PSEUDOMONAS SYRINGAE PHYLOGROUP 2B STRAINS INFECTING MONO- AND DICOTYLEDON HOST PLANTS

R.I. Sultanov[1, 2]*, G.P. Arapidi[1, 2], S.V. Vinogradova[3], V.M. Govorun[1, 2, 4], D.G. Luster[5], A.N. Ignatov[6]

[1] Moscow Institute of Physics and Technology (State University), Moscow, Russia
[2] Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, the Russian Academy of Sciences, Moscow, Russia
[3] Research Center of Biotechnology, Moscow, Russia
[4] SRCC of Physical-Chemical Medicine, Moscow, Russia
[5] USDA-ARS Foreign Disease - Weed Science Research Unit, Ft. Detrick, USA
[6] R&D Center "Phytoengineering" LLS, Moscow region, Russia
** Corresponding author: rhenium112@gmail.com*

Key words: *Pseudomonas syringae, dicots, monocots, pan-genome, core genome, T3SS, virulence factors*

Motivation and Aim: In recent years, the damage caused by bacterial pathogens to major crops has been increasing worldwide. *Pseudomonas syringae* is a widespread bacterial species that infects almost all major crops. Different *P. syringae* strains use a wide range of biochemical mechanisms, including phytotoxins and effectors of the type III and type IV secretion systems, which determine the specific nature of the pathogen virulence.
Methods and Algorithms: The genomes of *Pseudomonas syringae* strains 2507 (wheat) and 1845 (sunflower) isolated on the territory of the Russian Federation were determined by pyrosequencing and compared with previously published sequences of 18 genomes of the strains belonging to the same phylogroup and affecting dicots and monocots. We analyzed seven informative genes used in MLST genotyping of *P. syringae*, calculated the average nucleotide identity (ANI), and examined the compositions of the type III secretion system (T3SS) effectors and of the elements of insertion sequences (IS).
Results: We found that strains 2507 and 1845 and strains SM and B64 form a subgroup that is stable among the other strains of phylogroup 2b. The analysis of the genome of strain 1845 indicated the recent loss of several genetic elements (the cluster of genes responsible for the synthesis of syringolin and the prophage cluster) that are present in strains 2507, B64, and SM. We found three genes (YP_234264.1, YP_234265.1, and YP_237386.1), the acquisition of which by strain 1845 could lead to the change in its host class.
Conclusion: The results obtained by comparing the strain 1845 genome with the genomes of bacteria infecting monocots can help to identify the genes that define specific nature of the virulence of *P. syringae* strains.
Availability: The genomes of *P. syringae* strains 1845 and 2507 will be available soon at NCBI.

ADAPTATION AND BIOLOGICAL TIME

V.V. Suslov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: valya@bionet.nsc.ru*

Key words: *information, invasion, evolution, adaptation*

Motivation. Operationalize simplicity is an indisputable advantage of the classical definition any adaptation in terms of fitness ω. But the definition is not universa[1] , particularly, via ω cannot explain the Osborne effect[2] requiring assessment of the own adaptation not only *a posteriori* (i), but also *in situ*[3] (ii). Both i (by I preservation which due to the information caducity requires both *I* maintaining and *I* replication) and ii (by *R/w*) allows the adaptation formalism $[R,s]\ Q(I)/(p,P)\rightarrow[Z,w]$[4] [4].
Results. If true, but not used (unnecessary) information degraded [4] and an increasing of ω requires stability of s at least during reproduction period, the increasing of ω under s=const ($\omega(s$=const)>const) will lead to a priory incomplete preservation of I and selection in favor of a temperate noise, i.e. the *I* somehow associated with a relatively large number of *Q* regardless of ω ($\omega(s$=const)≥const) and *I's* verity. Hence, even I maintaining requires expansion, for which remains the only Gause's mechanism easily blocked under $\omega(s)$>const by density-dependent regulation [2]. Adaptation in terms of ω and adaptation in terms of permanent expansion are complementary, but in situ assessment can be formalized as (time for $R\rightarrow Z$ transition)/(effective time which has an organism under the *s)*. Both biotimes organism can estimated due to interference of resources and by-products. Autoreplicator has not the simple preemptive reflection because the selection in favor of accuracy and speed of *I* replication eliminates interference organizing replication by linearly (genomes) or conformational (prions) matrix types, by the way, decreasing possibility to use the *R/w in situ* assessment.
Acknowledgements: budget 0324-2015-0003; RSF 14-24-00123; RFBR 14-04-00485.
References:
1. R. Fisher (1930), The Genetical Theory of Natural Selection. Oxford Univ. Press: 318.
2. V.V. Suslov (2014), BGRS\SB-2014: 180-184.
3. V.S. Fridman, G.S. Eremkin (2008), Urbanization of "Wild" Avian Species in Evolving Urban Landscapes. M. MSU: 139 (in Russ.)
4. V.I. Korogodin, V.L. Korogodina (2012), Phenomenon of Life. V.2. M. Nauka: 173-338.

---

[1] Given this Fisher [1], the author of the definition, also proposed the geometric interpretation of the adaptation in which, factually, rediscovered the V.Kovalevsky inadaptation and anadaptation terms indistinguishabled through ω and predicted the necessity for the formation of Eldrege and Gould's stasis.
[2] Disagreements between temps of invasion, adaptation and evolution are a quick transition from slow to rapid invasion in new econiche/biotope and *vice versa* (i), the relative irrelevance of a successful invasion of a new econiche to preadaptation to that one (ii) and discordance between evolutionary and adaptation rates of small populations during invasion (iii) [2].
[3] In situ assessment used into the habitat by the organism-invader with receptor repertoire that can evaluate aboriginal signals [3]. Without the repertoire an organism can assess own chances based on the own stress-reaction developing during invasion in novelty [2].
[4] R – resources to target action *Z; s* – environment; *Q* – operator, that is, feature for the $R\rightarrow Z$ transition built on the genome information *I*; *p*, *P* – probabilities of transition in given *s*: *p* – randomly, *P* – with *Q(I)*; *w* – transition by-products [4].
[5] Or preventive adaptation that stabilizes the area and eventually the population size (random expansion also requires a large population).
[6] For example, in course of stress-reaction tiredness resulting an interference of mobilized resources warns about the danger of resource depletion long before its exhaustion [2].

TATA-BOX AND GENE EXPRESSION NORM OF REACTION

V.V. Suslov*, M.P. Ponomarenko, D.A. Rasskazov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: valya@bionet.nsc.ru*

Key words: *TATA-box, norm of reaction, gene expression*

Motivation: In 2014, we demonstrated that the TATA-box is a molecular regulator of norm of reaction (NR). It contributes to the variability of the most common property of gene expression, its transcriptional level [1]. However, the contribution used only mechanism of the regulation based on the model of Blake et al. [3] where TBP/TATA affinity can influence variability of the gene transcriptional level through change of the ratio of initiation/preinitiation promoter stages. NR is not limited to the transcriptional level and can include many features (specificity by tissue, transcriptional factor (TF), etc.) and mechanisms of the own alteration. This study is conducted to find them.
Methods. TBP/TATA affinity ($-ln[K_{D,\,TATA}(S)]$) of the naturally TATA-boxes with flanks and composite elements (CEs) where one of the flanks shared between the TATA-box and another binding site were *in silico* investigated by the Web-service *SNP_TATA_Comparator* (http://beehive.bionet.nsc.ru) which using the equilibrium equation for the four subsequent steps of TBP/TATA-box binding [2].
Results: 1) To start the HERV-H virus transcription its GC/GT-box must bind cellular transcription factor SP1. Its protected area overlaps half of the TATA-consensus. It was shown that mutations in free flank as one in protected area infringed ($-ln[K_{D,\,TATA}(S)]$). If mutations in protected area violated SP1-binding, then they align the levels of Sp1 and Sp3 transcription. If mutations violate the ($-ln[K_{D,\,TATA}(S)]$) only, they decreased both Sp1 and Sp3 transcription levels but retained the difference between them. These data suggest that in the CE the TATA-box limits the overall level of transcription, within which realizes TF-specific levels. 2) It was shown that in a similar CEs the ($-ln[K_{D,\,TATA}(S)]$) can limit overall level of transcription within which another TFs realized different tissue-specific transcriptional patterns. 3) The complication of the previous case was found. That is two closely spaced or overlapping TATA-boxes. The competition between them affects the specific regulation of the gene expression. (Experiments where the ($-ln[K_{D,\,TATA}(S)]$) can influence to the probability of polymerase usage (polII or polIII) described in literature). 4) Model [3] is based on the additive multi-step process of the transcription activation. The ($-ln[K_{D,\,TATA}(S)]$) can influence to the any step but the only step is the actual initiation. Hence, preinitiation promoter stages can be replaced by the reinitiating stage or major/minor transcripts ratio. Both cases were found. Thus, TATA-box can affect to the transcriptional NR in various ways. This confirms earlier suggestion about TATA-box as an element, encoding gene NR. The question about the element in the literature was not compromised earlier.
Acknowledgements: budget 0324-2015-0003; RSF 14-24-00123; RFBR 14-04-00485.
References:
1. P.M. Ponomarenko et al. (2014), *Rus. Journ. of Gen.: Appl. Res*., 18: 1219-1230.
2. P.M. Ponomarenko et al. (2009), *Mol. Biol. (Mos*k), 43(3): 512-520.
3. W.J. Blake et al. (2006), *Mol. Cell*., 24(6): 853–865.

VAVILOV'S HOMOLOGOUS SERIES AS EVOLUTIONARY FORCE

V.V. Suslov*, M.P. Ponomarenko, D.A. Rasskazov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: valya@bionet.nsc.ru*

Key words: *Vavilov's homologous series, evolution, adaptation, coadaptive substitutions*

Motivation. TBP/TATA affinity ($-ln[K_{D,TATA}(S)]$) of the naturally TATA-boxes with flanks and composite elements (CEs) where one of the flanks shared between the TATA-box and another binding site (BS) was *in silico* investigated by the Web-service *SNP_TATA_Comparator*[1] (http://beehive.bionet.nsc.ru).
Results. 1) For TATA-boxes of the polII-transcribed genes similar $-ln[K_{D,TATA}(S)]$ obtained with various combinations of substitutions in the TATA-consensus and flanks. 2) On the other hand, if the TATA-consensuses and their sequence analogs in TATA-less promoters have not changed, mutations in flanks have shifted the $-ln[K_{D,TATA}(S)]$ of TATA-promoters to the one of TATA-less promoters and *vice versa*. 3) For CEs, the mutational or phylogenetical changing of the $-ln[K_{D,TATA}(S)]$ significantly negative correlated with the changes in affinity of another BS to its transcriptional factor (TF). 4) For promoters of the polIII-transcribed genes TATA-consensus with flanks formed CEs with sites of TAFs[2] but *SNP_TATA_Comparator* allows us to distinguish the contribution of $-ln[K_{D,TATA}(S)]$ to transcription, even in cases that are not distinguishable by the level of gene expression, but distinguishable in nucleotide context of TATA-boxes and flanks. That is, between the TATA consensus and flanks, as well as between the TATA-box and BS in CE there is a mutual compensation for the function. The same function[3] can obtain by different combinations of substitutions, i.e. these combinations are Vavilov's homologous series (HS). But classic – Vavilov or Sobolev's – HS consist of several combinations of features where elimination some feature cannot break the combination's function[4] [2]. In our case, it is impossible. Functional overlap is too great, and evolution of the combination is similar to the protein's coadaptive substitutions, i.e. it can go in a stable but not favorable environment depending on the mutability and degree of functional overlap. HS may be the trend of autoadaptation. Generalization to other cases of HS with high functional overlapping[5] gives the rule of divergence/convergence HS[6].
Acknowledgements: budget 0324-2015-0003; RSF 14-24-00123; RFBR 14-04-00485.

References:

1. P.M. Ponomarenko et al. (2009), Mol. Biol. (Mosk), 43(3): 512-520.
2. N.I. Vavilov (1922), J. Genet., 12(1): 47-89.

[1] Which using the equilibrium equation for the four subsequent steps of TBP/TATA-box binding [1].

[2] TBP-associated factors.

[3] In our case it is $-ln[K_{D,TATA}(S)]$, or level, or tissue specificity of gene expression.

[4] As a rule, such features are in the right – the variable – part of the Vavilov's radical. E.g. awnless, liguleless cereals [2].

[5] I.e. Gilyarov-Simpson cycle, Abel's crossing of specializations, Krenke's HS or interim case of the superblock formation, where a selection that favors the simultaneous gene network regulation by several TFs may lead to a TFs competition for gene network's space by virtue of its BS formation. The fewer BSs, the higher the probability of TF's elimination or its specialization in time, tissue etc.

[6] The more functional overlap of the features in combination, the higher the probability of HS divergence - the loss of ancestral similarity in a series of combinations (*i*). The more traits in combination, assuming its functional overlap, the higher the probability of functional suppression and HS divergence. The longer the HS diverge, the more likely the loss of any of the traits (*ii*). Ultimately, the depletion of HS may lead to its convergence or in case of loss of functional overlap – its stabilization to classic HS (*iii*).

IDENTIFICATION OF PATHWAYS ASSOCIATED WITH CELL DEATH IN THE CORTEX OF OXYS RATS AS THE SIGNS OF ALZHEIMER'S DISEASE DEVELOP

G.K. Suvorov*, D.V. Telegina, E.A. Rudnitskaya, N.A. Stefanova, N.G. Kolosova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: augustocash@gmail.com*

Key words: *Alzheimer's disease, autophagy, apoptosis, cell death, OXYS rats, RNA-seq*

Motivation and aim: Alzheimer's disease (AD) is a progressive, age-dependent neurodegenerative disorder, featuring progressive impairments in memory, cognition, and ultimately leads to death. In spite of numerous studies on AD pathogenesis, the information about the molecular genetic preconditions of events leading to the death of neurons, as well as about the pathways of death, is extremely limited. Deregulation of autophagy and apoptosis plays a pivotal role in the etiology and/or progress of many of these diseases, including AD. Autophagy and apoptosis are basic physiologic processes contributing to the maintenance of cellular homeostasis. Autophagy is a major intracellular degeneration pathway involved in the elimination and recycling of damaged organelles and long-lived proteins by lysosomes. Apoptotic processes remove old and damaged cells to maintain tissue homeostasis without harming adjacent cells. To determine the role of cell death in the pathogenesis of AD, suitable animal models are needed. In this study, using nontransgenic OXYS rats that simulate key aspects of sporadic AD, we aimed to compare the gene expression profiles of the prefrontal cortex from OXYS rats and Wistar rats (as control) to identify the molecular mechanisms and the factors underlying of neuronal cell death in disease development. The transcriptome analysis was conducted at three stages of the disease (pre-symptomatic, 20 days; symptomatic, 5 month; and progressive stage, 18 month) in OXYS rats, using RNA-Seq technique.
Results and conclusion: Our results show that the development of the signs of AD (between ages 20 days and 5 months) in OXYS rats takes place during changes in mRNA expression of the 7 genes that are mostly related to processes of autophagy, such as *Atg12*, *Atg7*, and *Atg8* (regulators of elongation). In addition, changes in mRNA expression of the 21 genes were related to apoptosis (proapoptotic genes and inhibitors) in the prefrontal cortex of OXYS rats between ages 20 days and 5 months. In OXYS rats, with progression of disease, 24 genes related to apoptosis and 7 genes related to autophagy change their expression. Importantly, Wistar rats show changes in expression of 21 genes related to apoptosis only between ages 20 days and 5 months. We also indicated the upregulation of 5 proapoptotic genes in 20-day-old OXYS rats compared Wistar rats. At the age of 5 and 18 months in OXYS rats, the balance between pro- and anti-apoptotic genes were changed (compared to Wistar rats). Accordingly, we demonstrated that the development of AD-like pathology in OXYS rats is related to the alterations in processes of autophagy and apoptosis.
Acknowledgements: This work was supported by the grant from the Russian Foundation for Basic Research (project # 15-04-06066).

GENETIC POLYMORPHISM OF GLUTATHIONE S-TRANSFERASE P1 (*GSTP1)* AMONG BURYATS, TELEUTS AND RUSSIANS

L.E. Tabikhanova[1]*, L.P. Osipova[1, 2], T.V. Churkina[1], H. Bai[3], E.N. Voronina[2, 4]
M.L. Filipenko[2, 4]

[1] *IInstitute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Inner Mongolia University for the Nationalities, Tongliao, China*
[4] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
** Corresponding author: tabikhan@bionet.nsc.ru*

Key words: *human genetics, nucleotide polymorphisms, Eastern Buryats, Teleuts, Russians, real-time PCR, xenobiotics, GSTP, rs1695, rs1138272*

Motivation and Aim: Glutathione S-transferases (GSTs) belong to a super family of detoxification enzymes, which play an important role in protecting cells from damage caused by endogenous and exogenous compounds. Genetic polymorphisms *GSTs* genes might influence the detoxification activities of the enzymes predisposing individuals to risk of oncological and other multifactorial diseases. The aim of this work is to determine the frequencies of *A1405G* (*Ile105Val, rs1695*) and *C2285T* (*Ala114Val, rs1138272*) polymorphisms of *GSTP1* gene in healthy Buryats, Teleuts and Russians.
Methods and Algorithms: This study was performed on Eastern Buryats (N=139), Russians (N=68) of Trans-Baikal area and Teleuts of Kemerovo Oblast (N=115). Genotyping was performed using Real-time PCR with competitive TaqMan allele-specific probes.
Results: The frequency of the *GSTP1 105Val* allele in the Buryats and Teleuts cohorts was 27,7% and 24.8% respectively. Those showings fall between those of the Eastern Asia populations' (10-20%) and of the European populations' (27-41%) [1]. The frequency of the *GSTP1 114Val* allele in the Buryats cohort (4,9%) and Teleuts cohort (2,2%) is also lower compared with the frequency range found in European populations (6-8%). At the same time the frequency of the said allele in the Eastern Asia populations (0,0-0,5%) is considerably lower than that of Buryats and Teleuts. The frequency of the *105Val* allele (31,3%) *114Val* allele (7,4%) in the Russian cohort corresponds to the frequency range found in other European populations.
Conclusion: Genetic polymorphism of GSTP1 among Buryats, Teleuts and Russians of Trans-Baikal area was compared with other ethnic groups worldwide. Our findings demonstrate the impact of ethnicity and reveal a characteristic pattern of three populations. The study would provide a database for future genetic studies and for further epidemiological investigations in the populations.
Acknowledgments: The work was supported by RFBR 15-54-53091.
References:
1. The 1000 Genomes Project Consortium. (2012) An integrated map of genetic variation from 1,092 human genomes, Nature. 7422: P. 56.м

THE *ILE462VAL* POLYMORPHISM OF THE CYTOCHROME *P450* CYP1A1 GENE AMONG EASTERN BURYATS COMPARED WITH RUSSIANS IN TRANS BAIKAL AREA

L.E. Tabikhanova[1]*, L.P. Osipova[1, 2], T.V. Churkina[1], E.N. Voronina[2, 3], M.L. Filipenko[2, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
** Corresponding author: tabikhan@bionet.nsc.ru*

Key words: *human genetics, nucleotide polymorphisms, Eastern Buryats, Trans Baikal area, real-time PCR, xenobiotics, CYP1A1 Ile462Val*

Motivation and Aim: New chemical substances (xenobiotics) make their way onto human environment. Many of them can act as carcinogen and mutagen. Enzymatic system of biotransformation of xenobiotics plays the most important role in organism defense. This work concerns a polymorphism of the cytochrome *P450 CYP1A1* gene, the *CYP1A1*2C* variant (*Ile462Val,* rs1048943) This substitution results in a two-fold increase in enzyme activity, which leads to accumulation of active intermediates and increases the risk of DNA mutations and chemically induced carcinogenesis. The aim of this work is to study the frequency of the *462Val* allele in two different ethnic groups.
Methods and Algorithms: This study was performed on Eastern Buryats (N=132) and Russians (N=67) in Alkhanai and Orlovskii settlements from Aginskii Buryat Region of Trans-Baikal area [1]. Genotyping was performed using Real-time PCR with competitive TaqMan allele-specific probes.
Results: The frequency of the *462Val* allele in the Buryats cohort was 28,4%, which corresponds to the frequency range found in East Asian populations and is higher than the values typical of European populations [2]. The frequency of the *462Val* allele in the Russian cohort was 3,0%, which corresponds to the frequency range found in European populations.
Conclusion: A high-frequency occurrence of the *CYP1A1 462Val* allele among Eastern Buryats of Trans Baikal area may be indicative of a higher population-wide risk of diseases influenced by this genetic polymorphism, compared with Russians.
Acknowledgments: The work was supported by RFBR 15-54-53091.
References:
1. Tabikhanova L. E. Osipova L. P. (2012) Analysis of the genetic and demographic structure of populations from Aginskii Buryat district contrasting in habitation conditions, Russian Journal of Genetics, 48: 1239-1246.
2. Tiis R.P. et al. (2016) The *ILE462VAL* polymorphism of the cytochrome *P450CYP1A1* gene among Tundra Nenets in Yamalo-Nenets Autonomous Okrug, Nganasans in the Taimyr Peninsula and Russians in Siberia, Vavilov Journal of Genetics and Breeding. 20: 16-22. (In Russian).

COMPUTER MODELLING OF INHIBITORS OF PROTEASE OF HUMAN HEPATITIS C VIRUS BASED ON KNOTTIN SCAFFOLD

A.V. Talanova, D.S. Shcherbinin, E.F. Kolesanova, A.V. Veselovsky
V.N. Orekhovich Institute of Biomedical Chemistry, Moscow, Russia

Key words: *hepatitis C, modeling, knottin*

Knottins are cystine-knot family of proteins, a class of small polypeptides (typically about 30 amino acids) usually act as proteases inhibitors. Knottins contain three disulfide bonds forming a molecular 'knot' that constrain loop regions to a core of anti-parallel b-sheets. The unique topology of the knottin fold imparts high chemical and thermal stability and resistance to proteolysis. Moreover, knottins can be chemically synthesized and folded in vitro or produced recombinantly in various expression systems.
Here we present the modification of knottin from Momordica cochinchinensis (MCOTI-II) (selective inhibitor of trypsin) toward to its ability to interact with protease NS3 human hepatitis C virus (HCV). Modifications included truncation of N-end of inhibitor and mutations of residues in inhibitory loop and outside it. The five variants of modified inhibitors were designed. Complexes of protease HCV with these inhibitors were simulated by molecular dynamics and binding energy was estimated using MM-PBSA method. Based on results of modelling two modified knottins were offered for synthesis.

THE STRUCTURE OF GENETIC PREDISPOSITION TO TYPE 1 AND TYPE 2 DIABETES

N.V. Tarasenko[1, 2]*, I.A. Goncharova[1, 3], A.V. Markov[1], V.P. Puzyrev[1, 2]
[1] *Research Institute of Medical Genetics, Tomsk, Russian Federation*
[2] *Siberian State Medical University, Tomsk, Russian Federation*
[3] *Institute for complex issues of cardiovascular diseases, Kemerovo, Russian Federation*
* *Corresponding author: nataly.tarasenko@medgenetics.ru*

Key words: *genetic predisposition, type 1 diabetes, type 2 diabetes*

Motivation and Aim: diabetes mellitus is one of the largest global health emergencies around the world. Type 1 and type 2 diabetes are having a number similar etiological factors and certain resemblance in the mechanisms of the disease. Consequently it is important try to estimate the role of common genetic component in the genesis and formation of both types of diabetes. The aim of our study was to compare the structure of genetic predisposition to type 1 and type 2 diabetes.
Methods and Algorithms: we studied 58 polymorphic variants (SNPs) of 47 genes whose products are engaged in various metabolic pathways and involved in fibrogenesis, endothelial function, immune response and inflammation. The group of patients with type 1 diabetes (T1D) included 285 individuals and the group of patients with type 2 diabetes (T2D) included 96 individuals. The control group consisted of 300 individuals from Tomsk population. Genotyping was performed using mass spectrometry on Sequenom MassARRAY (USA).
Results: we found an association with type 1 diabetes for seven markers: *MMP3* rs679620 (p=0,004), *ITGB5* rs1007856 (p=0,039), *ITGA4* rs1143674 (p=0,002), *LIG1* rs20579 (p=0,003), *ADAMDEC1* rs3765124 (p=0,014), *IFNL2* rs12980602 (p=0,029), *PARP4* rs4986819 (p=0,043). These genes are involved in signal transduction and provide cell communication (*ITGB5, ITGA4, ADAMDEC1*), participate in the metabolism of proteins and nucleic acids (*MMP3, LIG1, PARP4*) and in the regulation of the immune response (*IFNL2*). The T2D group showed association with markers *APOA2* rs5082 (p=0,035), *LDLR* rs2738446 (p=0,017), *MTAP* rs7023329 (p=0,008), *CDKN2B-AS1* rs1333049 (p=0,029), *KIAA1462* rs3739998 (p=0,003). Genetic markers associated with T2D are involved in the processes regulated metabolism and energy pathways (*APOA2, MTAP*), signal transduction and cell communication (*LDLR, CDKN2B-AS1*), the endothelial cells adhesion (*KIAA1462*).
Conclusion: we did not identify any genetic markers shared between the T1D and T2D. Thus, our findings suggest that genetic predisposition for the studied markers for T1D and T2D are specific.
The study is partially supported by the RFBR (16-04-00840).

FIGHTING WITH HIV-1 RESISTANCE TO REVERSE TRANSCRIPTASE INHIBITORS BY COMPUTER-AIDED APPROACH

O.A. Tarasova*, D.A. Karasev, D.A. Filimonov, V.V. Poroikov
Institute of Biomedical Chemistry, Moscow, Russia
** Corresponding author: olga.a.tarasova@gmail.com*

Key words: *HIV reverse transcriptase inhibitors, resistance, PASS*

Motivation and Aim: HIV reverse transcriptase (RT) inhibitors are important components of the highly active antiretroviral therapy [1]. Response to treatment with HIV RT inhibitors agents often depends on viral drug resistance development. There are a lot of clinical and biochemical data on the relationships between the occurring of the single point mutations in pol gene of HIV and the resistance of the particular variants of the RT to the nucleoside reverse transcriptase inhibitors (NRTI) and non-nucleoside reverse transcriptase inhibitors (NNRTI). The aim of our study is the development of the computer-aided approach to the search for HIV-1 RT inhibitors, active against the resistant RT variants.
Methods and Algorithms: We propose an application of PASS algorithm [2] to the (i) prediction of the amino acid changes, potentially involved in the resistance of HIV-1 and (ii) integrated approach based on usage of small molecules descriptors and the descriptors of the amino acid sequences of the protein to the search for the compounds with activity against the HIV resistant strains.
Results: We used over 3200 variants of the HIV-1 RT from the publicly accessible HIV Drug Resistance Database tested against the ten anti-HIV drugs in two susceptibility assays (Phenosense and Antivirogram). Two classes of the variants were considered: "susceptible" and "resistant". The average balanced accuracy of prediction obtained in the leave-one-out procedure for the Phenosense data set was about 82%, and for the Antivirogram data set was about 87%. For further computational experiments, we selected over 500 sequences, for which the complete amino acid sequences can be retrieved from NCBI Protein database. We have developed and tested an approach based on the integration of (i) estimated probability of the specific pentapeptide to occur in the amino acid sequence of the particular variant of HIV RT and (ii) estimated probability of the ligand descriptors (multilevel neighborhoods of atoms, MNA [2]) to arise in the particular ligand. The average balanced accuracy of prediction obtained in the leave-one-out procedure was about 84%.
Conclusion: The computer-aided approach to finding new HIV-1 RT inhibitors provides the possibility to predict the (i) amino acid changes, potentially involved in resistance and (ii) probability of the compound to be active against the particular HIV-1 variant with average balanced accuracy about 84%.
Availability: Detailed description of the algorithm description may be provided on request.
Acknowledgements: This work was supported by RFBR grant No. 16-34-60187.
References:
1. Geronikaki A. et al., *Topics in Medicinal Chemistry*, 2016, 1-59. DOI:10.1007/7355_2015_5001.
2. Filimonov D. et al., *Chemistry of Heterocyclic Compounds*, 2014, 50 (3), 444-457.

NUCLEOTIDE DIVERSITY ANALYSIS HIGHLIGHTS FUNCTIONALLY IMPORTANT GENOMIC REGIONS

T.V. Tatarinova[1, 2]*, E. Chekalin[3], Y. Nikolsky[3, 4, 5], S. Bruskin[3], D. Chebotarov[6], K.L. McNally[6], N. Alexandrov[6]*

[1] *Center for Personalized Medicine and Spatial Sciences Institute, University of Southern California, Los Angeles, CA, USA*

[2] *Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia*

[3] *Vavilov Institute of General Genetics, Moscow, Russia*

[4] *F1 Genomics, San Diego, CA, USA*

[5] *School of Systems Biology, George Mason University, VA, USA*

[6] *International Rice Research Institute, Los Baños, Philippines*

* *Corresponding authors: tatiana.tatarinova@usc.edu; n.alexandrov@irri.org*

Key words: *Oryza sativa, 3,000 Rice Genomes Project, SNP density, nucleotide diversity, promoter, TSS, TTS, nucleotide composition*

Motivation and Aim: Understanding the relationship between genotype and phenotype is a key issue in life sciences with hugely important implications in biomedical R&D, healthcare and agriculture. Innate genetic variability is both the source and consequence of selection in populations of humans, crops and animals.

Results: We analyzed functionality and relative distribution of genetic variants across the complete *Oryza sativa* genome, using the 40 million single nucleotide polymorphisms (SNPs) dataset from the 3,000 Rice Genomes Project[1], the largest and highest density SNP collection for any higher plant. We have shown that the DNA-binding transcription factors (TFs) are the most conserved group of genes, whereas kinases and membrane-localized transporters are the most variable ones. TFs may be conserved because they belong to some of the most connected regulatory hubs that modulate transcription of vast downstream gene networks, whereas signaling kinases and transporters need to adapt rapidly to changing environmental conditions.

Conclusion: In general, the observed profound patterns of nucleotide variability reveal functionally important genomic regions. As expected, nucleotide diversity is much higher in intergenic regions than within gene bodies (regions spanning gene models), and protein-coding sequences are more conserved than untranslated gene regions. We have observed a sharp decline in nucleotide diversity which begins at about 250 nucleotides upstream of the transcription start and reaches minimal diversity exactly at the transcription start. We found the transcription termination sites to have remarkably symmetrical patterns of SNP density, implying presence of functional sites near transcription termination. Also, nucleotide diversity was significantly lower near 3′UTRs, the area rich with regulatory regions.

Availability: SNPs are available from the SNP-Seek database (http://oryzasnp.org/iric-portal/).

References:

1. The 3,000 rice genomes project, GigaScience 2014 3:7,DOI: 10.1186/2047-217X-3-7

RELATIONSHIP OF CELL DEATH IN RETINA OF RATS DURING AGING WITH THE DEVELOPMENT OF RETINOPATHY

D.V. Telegina*, O.S. Kozhevnikova, N.G. Kolosova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: telegina@bionet.nsc.ru*

Key words: *RNA-seq, apoptosis, autophagy, necrosis, RPE cells, microglia, macroglia, age-related macular degeneration, retinopathy, OXYS rats*

Motivation and aim: The dysfunction of retinal pigment epithelium and the following death of photoreceptors are hallmarks of late-stage of age-related macular degeneration (AMD) - a complex disease, which becomes a major cause of irreversible vision loss in people older than 60. In spite of numerous studies on AMD pathogenesis, the information about the molecular genetic preconditions of events leading to the death of photoreceptors, as well as about the pathways of death, is extremely limited. Aging and age-related disease related are associated with changes in the expression of many genes. High-throughput genomic studies integrated transcriptomic next-generation sequencing with bioinformatic analysis of molecular pathways. Today, this approach seems to be productive at elucidation of disease development. In recent years, the data on the retinal transcriptome have accumulated substantially, but the information on its changes with age and at various stages of AMD is scarce. The main aim of this study was to identify the mechanisms of retinal cell death during aging and the development of retinopathy similar to the dry form of AMD in OXYS rats.
Results and conclusion: It is show that during the development of the first evidence of retinopathy (between 20 days and 3 mo) it was observed the changed expression of genes that regulate processes of autophagy (11 genes), necroptosis (8 genes), and apoptosis (25 genes) in OXYS rats. The data of the construction and analysis of gene interaction networks formed by differentially expressed genes between OXYS and Wistar, involved in apoptosis, shows the suppression of apoptosis in OXYS rats at all ages. We also demonstrated the significance of the extrinsic apoptotic pathway at preclinical, early, and advanced stages of retinopathy development. The largest number of apoptosis genes changed the expression profile during the manifestation of the first signs of retinopathy in rats OXYS. An increased level of apoptosis was observed in OXYS rats at age 20 day, and the number of TUNEL-positive cells in OXYS rats was 1,5 - fold greater than that in Wistar rats. By the age of three months, the number of TUNEL-positive cells in the retina of both OXYS and Wistar rats significantly decreased to solitary cases and remained at the same level in the retina of 18-month-old animals, without any interstrain differences. Also it is shown that the total number of RPE cells decreased with age at both OXYS and Wistar lines, in OXYS rats retinopathy develops against the background of the destructive changes in the RPE cells. The study of protein expression GFAP (macroglial marker) showed that GFAP decreased in the retina of 20 day-old and increased in retina of 7-month-old OXYS rats. The activity of microglia migration did not differ in OXYS and Wistar rats and did not changes with age, but we observed change of the activated microglial cells distribution in OXYS rats at the 20 day, 3mo and 18 mo.
Acknowledgements: Supported by the Russian Foundation for Basic Research (project # 15-04-02195A).

SANGER DATA PROCESSING IN UNIPRO UGENE

A.V. Tiunov*, E.A. Pushkova, Y.A. Algaer, G.A. Grekhov and the UGENE team
Unipro Center of Information Technologies, Novosibirsk, Russia,
** Corresponding author: atiunov@unipro.ru*

Key words: *UniPro UGENE, Sanger sequencing, Aligning, Raw data, pipeline, Workflow Designer*

Motivation and Aim: Sanger [1] sequencing has been used for several decades. Although more modern sequencing technologies have already appeared (e.g. NGS [2]), the Sanger technology is still widely used as a supplementary method for NGS for more precision or as an independent method for sequencing of small genomes. It makes developement of a high quality algorithm for Sanger data processing an actual objective.
Methods and Algorithms: UGENE [3] is a universal toolkit for biological data visualization and analysis. One of UGENE's features is Workflow Designer that allows anyone with no programming skills to design and run complex computational pipelines. Many algorithmic blocks are implemented in Workflow Designer to perform different kinds of tasks. These blocks have been used to build a new element that aligns Sanger reads to a reference sequence. The algorithm is based on the BLAST [4] tool that first localizes an appropriate region of the read on the reference. Then its exact location is found by the Smith-Waterman algorithm [5] that can be launched on GPU to increase the overall pipeline performance. In the end, all aligned reads are merged into a single multiple alignment data structure via UGENE's own merging algorithm.
Results and Conclusion: The algorithm has proved itself as a fast and memory efficient solution for aligning Sanger reads to a reference. The Workflow Designer tool provides pipelined data processing for several independent blocks of Sanger data. In the nearest future the algorithm will be available in the UGENE toolkit along with a new viewer for raw Sanger data.
Availability: http://ugene.net/download.html
References:

1. Sanger F., Niclein S., Coulson A.R. (1977) DNA sequencing with chain-terminating inhibitors, Proceedings of the National Academy of Sciences, USA, 74: 5463-5467.
2. de Magalhães JP, Finch CE, Janssens G (2010) Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions, Ageing Research Reviews 9 (3): 315–323.
3. Okonechnikov K., Golosova, O., Fursov, M., the UGENE team (2012) Unipro UGENE: a unified bioinformatics toolkit, Bioinformatics. 28 (8): 1166-1167.
4. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool, Journal of Molecular Biology 215:403-410.
5. Smith, Temple F.; and Waterman, Michael S. (1981) Identification of Common Molecular Subsequences, Journal of Molecular Biology 147: 195–197.

GENETIC BASIS OF AGGRESSION: CLUSTERIZATION OF EXPRESSION PROFILES

E.S. Tiys[1, 2]*, A.O. Bragin[1, 2], I.V. Medvedeva[1, 2], I.V. Chadaeva[1], A.L. Markel[1], Y.L. Orlov[1, 2]

[1] *Institute of Cytology of Genetics SB RAS, Novosibirsk, Russia*

[2] *Novosibirsk State University, Novosibirsk, Russia*

* *Corresponding author: tiys@bionet.nsc.ru*

Key words: *aggression, RNA-Seq, differential expression, functional annotation, MeV, GOexpress*

Motivation and Aim: Aggression is a complex phenomenon of behavior with physiological and genetics roots. It is important to find traces of aggression on genetic level. RNA-seq experiment is invaluable source of information on genetic component of aggressive behavior.

Methods and Algorithms: We used RNA-Seq data from brain areas of aggressive and tame rat lines obtained at ICG SB RAS [1]. MirVana™ miRNA Isolation Kit(Ambion) was used to extract mRNA. NEBNext® mRNA Library Prep Reagent Set for Illumina (NEB) was used to prepare libraries. Sequencing was conducted according to Illumina protocol. Reference genome RGSC Rnor_5.0\rn5 and TopHat2 tool were used for the alignment. The sequence reads were analyzed using Cufflink workflow [2] using FPKM to quantify expression. Clustering of samples by gene expression was done with MeV [3] using average linkage clustering. Correlation coefficient was used as a measure of similarity. Differential expression was calculated with MeV using Wilcoxon test with FDR multiple tests correction. Ranking of processes was made by GOexpress [4].

Results: Expression profiles were clustered by tissues and the tissue clusters were accurately clustered into aggressive and tame samples. Differential expression was not statistically significant. Applying GOexpress ranking allow us to prioritize the following Gene Ontology biological processes: “negative regulation of transporter activity”, “extrinsic apoptotic signaling pathway in absence of ligand”, “salivary gland cavitation”, “regulation of response to DNA damage stimulus, establishment of cell polarity”.

Conclusion: On the one hand accurate clusterization of the expression profiles confirms the presence of molecular genetic traces of aggression behavior in the RNA-seq data, on the other hand differential expression and prioritizing were not show statistical significance. Further analysis is needed to uncover molecular basis of aggression.

Acknowledgements: The work was supported by RSF grant 14-14-00269.

References:

1. Y.L. Orlov et al. (2012) ICGenomics: program complex for symbol sequence analysis in genomics, *Vavilov journal of ge*netics and breeding, 16(4/1): 732-741. (In Russian).
2. http://cole-trapnell-lab.github.io/cufflinks/
3. A.I. Saeed (2003) TM4: a free, open-source system for microarray data management and analysis, Biotechniques, 34(2): 374.
4. K. Rue-Albrecht (2016) GOexpress: an R/Bioconductor package for the identification and visualisation of robust gene ontology signatures through supervised learning of gene expression data, BMC bioinformatics, 17(1): 1.

GENEONTOLOGY BIOLOGICAL PROCESSES SENSITIVE TO SALT DIET CHANGES IN AN EXPREIMENT WITH 105-DAY ISOLATION: STATISTICAL ANALYSIS OF URINE PROTEOME

E.S. Tiys[1, 2]*, E.D. Petrovskiy[1], L.Kh. Pastushkova[3], D.N. Kashirina[3], I.M. Larina[3], V.A. Ivanisenko[1]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk state university, Novosibirsk, Russia*
[3] *Institute for Biomedical Problems RAS, Moscow, Russia*
** Corresponding author tiys@bionet.nsc.ru*

Key words: *salt intake, Gene Ontology, Random forest, GOexpress, Bioconductor*

Motivation and Aim: Conducting mass-spectrometry experiments allows to solve important problem of protein expression measurement. Promising approach to interpret mass spectrometric studies is based on the identification of biological processes, protein representation of which is the most contrast between case and control samples. The aim of the study was a computer search for biological processes, changing of functional activity of which was related with salt intake regime, based on urine proteome analysis in 105-day volunteer isolation experiment.

Methods and Algorithms: Weekly during the 105-day isolation urine samples from six volunteers were collected and chromato-mass spectrometry was conducted. Collected spectra were analysed with Mascot tool. The results consist of lists of identified proteins for each volunteer and data on salt content in diet (6, 9 or 12 g/day). GOexpress package [1] of Bioconductor was used to search for relation between salt intake and Gene Ontology (GO) processes through corresponding protein sets. Built-in package estimation of statistical significance was used with 250000 permutations of protein ranks. Processes with corrected $p<0.05$ were considered as significant. The Benjamini-Hochberg correction was used (R language, function p.adjust).

Results: Using GOexpress we identified 131 GO processes, demonstrating statistically significant relation with salt consumption. Among these, we can emphasize metabolic processes of connective tissue, such as chondroitin sulfate metabolic process, collagen catabolic process, hyaluronan metabolic process. This fact consistent with earlier hypothesis of extracellular matrix participation in sodium deposition [2, 3].

Conclusion: Mass-spectrometry data analysis of volunteer urine proteomes gave potential link between changes in salt intake and functioning of biological processes, involving in metabolism of connective tissue, that in agreement with hypothesis about its protein participation in sodium deposition.

Acknowledgements: Project № 0324-2015-0003 and RSF grant 14-24-00123

References:

1. K. Rue-Albrecht (2016) GOexpress: an R/Bioconductor package for the identification and visualisation of robust gene ontology signatures through supervised learning of gene expression data, BMC bioinformatics, 17(1): 1.
2. I.M. Larina et al., (2012) Reconstruction of associative protein networks connected with processes of sodium exchange regulation and sodium deposition in healthy volunteers based on urine proteome analysis, Human Physiology, 38(3): 316–323.
3. L.Kh. Pastushkova, (2015), Proteome profile of urine of healthy human under normal conditions and under factors of space flight, thesis [In russian].

PHSYOLOGICAL AND TRANSCRIPTIONAL CHANGES IN A BLOSSAM-END ROT RESISTANT TOMATO INTROGRESSION LINE IL8-3 FRUIT

S. Tomoki, H. Ikeda, Y. Kanayama*
School of Agricultural Science, Tohoku University, Sendai, Japan
** Corresponding author: yoshinori.kanayama.a7@tohoku.ac.jp*

Key words: *Blossom-end rot, Calcium, Solanum lycopersicum, Solanum pennellii, tomato, transcriptome*

Motivation and Aim: Tomato (*Solanum lycopersicum*) is one of the most important vegetables due to its high economic and nutrition values and is also important as an experimental model species of the Solanaceae family and fleshy-fruited plants. Tomato is sensitive to Calcium (Ca) deficiency, which causes several physiological disorders in fruits such as cracking and blossom-end rot (BER). Ca concentration, early fruit growth, and expression of Ca-movement-related genes were analyzed during early fruit development in tomato, which is the most important stage in the incidence of BER, to investigate physiological mechanisms affecting the occurrence of BER. We used tomato introgression line IL8-3 having a chromosome segment from the wild relative (*Solanum pennellii*) because the line shows lower incidence of BER compared with the parent cultivar 'M82' (*S. lycopersicum*), as described previously. In addition, we tried to analyze transcriptomic character of early fruit development in IL8-3 on microarray data and characterize tomato Ca-movement-related genes by phylogenetic analysis.
Results and Conclusion: Total Ca concentration in fruit and leaves was higher in IL8-3 than in 'M82', whereas no significant differences were observed between total Ca concentration in roots and stems of 'M82' and IL8-3. Therefore, the lower incidence of BER in IL8-3 could be not related to whole-plant uptake but to transport within the plant. IL8-3 fruit showed a lower growth rate than 'M82', which could result in preventing the occurrence of BER. The expression of genes encoding Ca^{2+}/H^{+} exchanger, autoinhibited Ca^{2+}-ATPase, Ca^{2+} channel, and Na^{+}/Ca^{2+} exchanger-like protein, was higher in IL8-3 fruit than in 'M82' fruit, suggesting active Ca movement in IL8-3. In contrast, the expression of the gene encoding Group 3 Cation/Ca^{2+} exchanger was lower in IL8-3 fruit than in 'M82'. Additionally, the expression level of boron-movement-related genes is different between IL8-3 and 'M82'. It is reported that boron deficiency may also affect the incidence of BER and Ca^{2+} transportation. All our results could be related to physiological mechanisms accounting for the lower incidence of BER in IL8-3 and one of these factors may play a primary role.

DIFFERENTIAL EXPRESSION OF *SHAGGY*, *A DROSOPHILA MELANOGASTER* GENE ENCODING GSK-3 BETA, AFFECTS LIFESPAN

M.V. Trostnikov, N.V. Roshina, E.G. Pasyukova
Institute of Molecular Genetics, RAS, Russia
** Corresponding author: egpas@rambler.ru*

Key words: *Life span, GSK-3 beta, the nervous system, Drosophila*

Motivation and Aim: Protein kinases are important proteins involved in multiple phosphorylation cascades, and their role in systemic regulation of various metabolic processes is well recognized. A serine-threonine kinase GSK-3 beta (glycogen synthase kinase-3 beta) is an important protein involved in various signaling pathways and metabolic processes, neurogenesis and neuronal function. Earlier, we have shown that several genes involved in asymmetric neuroblast division, including *shaggy* (*sgg*) encoding *Drosophila* GSK-3 beta, affect *Drosophila* lifespan. Several mutations in *sgg* increased male and female mean lifespan by 5% to 50% together with slightly reducing the total amount of GSK-3 beta protein. In order to further assess the role of GSK-3 beta in the control of lifespan, aging and neuronal function, we used transgenic lines with additional copies of the four *sgg* variants under inducible UAS promoters, to provide *sgg* overexpression; transgenic lines with hairpins, to provide *sgg* RNA-i knockdown; and special driver lines, to induce overexpression and knock-down in various tissues and cell types.
Results: We demonstrate that i) the proper expression of the main functional GSK-3 beta transcript, RB, in the nervous system and fat body is essential for survival; ii) *sgg* transcripts whose functionality has not previously been demonstrated, RA, RG and RO, appeared to be functional: alterations in their expression in the nervous system slightly decrease male lifespan; in muscles, only overexpression of RA affects lifespan; iii) *sgg* misexpression in different neurons affects lifespan in a neuron-specific manner: moderate increase and decrease in the amount of the main isoform of GSK-3 beta, PB, in peptidergic and cholinergic neurons has no effect on lifespan, while in glutaminergic and motor neurons it increases female and decreases male lifespan; iv) dopaminergic neurons are the most sensitive to *sgg* function: strong *sgg* RB overexpression is lethal, while moderate decrease in PB amount improves female lifespan. Analyses of the molecular and cellular mechanisms underlying positive effects of *sgg* misexpression on lifespan are under way.
Conclusion: Differential expression of GSK-3 beta is one of the mechanisms involved in a complex regulation of lifespan and aging.

ADDITIVITY AND NON-ADDITIVITY OF GENETIC CONTROL OF HUMAN METABOLOME

Y.A. Tsepilov[1, 2]*, S. Shin[3, 4], N. Soranzo[3], T.D. Spector[5], J. Adamski[6, 7, 8], G. Kastenmüller[9], K. Strauch[10, 11], R. Wang-Sattler[12], C. Gieger[10], Y.S. Aulchenko[1, 2], J.S. Ried[10]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom
[4] MRC Integrative Epidemiology Unit (IEU), University of Bristol Integrative Epidemiology, Bristol, United Kingdom
[5] Department of Twin Research and Genetic Epidemiology, King's College London , London, United Kingdom
[6] Institute of Experimental Genetics, Genome Analysis Center, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
[7] Institute of Experimental Genetics, Life and Food Science Center Weihenstephan, Technische Universität München, Freising-Weihenstephan, Germany
[8] German Center for Diabetes Research, Neuherberg, Germany
[9] Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
[10] Institute of Genetic Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
[11] Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany
[12] Research Unit of Molecular Epidemiology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany
** Corresponding author: tsepilov@bionet.nsc.ru*

Key words: *genome-wide association studies, non-additive models, KORA, metabolomics, genotypic model*

Motivation and Aim: Genome-wide association studies (GWAS) are widely applied to analyze the genetic effects on phenotypes. With the availability of high-throughput technologies for metabolite measurements, GWAS successfully identified loci that affect metabolite concentrations and underlying pathways. In most GWAS the effect of each SNP on the phenotype is assumed to be additive. Other genetic models such as recessive, dominant or over-dominant were considered only by very few studies. In contrast to that, there are theories that emphasize the relevance of non-additive effects as a consequence of physiological mechanisms. This might be especially important for metabolites as these traits are closer to the underlying pathways than other traits or diseases.
Methods and Algorithms: In this study we analyzed systematically non-additive effects on a large panel of serum metabolites and all possible ratios (22,801 in total) in a population based study (KORA F4, N = 1,785). We applied four different 1 df tests corresponding to an additive, dominant, recessive and over-dominant trait model and additionally a genotypic model with 2 df that allows a more general formulation of genetic effects.
Results: Twenty three loci were found to be genome-wide significantly associated (Bonferroni corrected p-value $\leq 2.19 \times 10^{-12}$) with at least one metabolite or ratio. For five of them we show evidence of non-additive effects. We replicated seventeen loci (including three loci with non-additive effects) in an independent study (TwinsUK, N = 846).
Conclusion: We proposed efficient methodology of identification of non-additive loci. We showed that genetic effects on metabolite concentrations and ratios were mostly additive; at the same time we also found significant and replicable non-additive loci.

IDENTIFICATION OF BREED-SPECIFIC SNP-MARKERS FOR *SUS SCROFA DOMESTICUS* USING SRA-DATA OF NGS PROJECTS

I.S. Tsybovsky*, V.N. Kipen, S.A. Kotova
Scientific and Practical Centre of the State Committee of Forensic Expertises, Minsk, Belarus
** Corresponding author: tsybovsky@yahoo.com*

Key words: *SNP, Sequence Read Archive, Next-Generation Sequencing*, *Sus scrofa domesticus*

Motivation and Goal: The following domestic pig breeds are widely spread in the Republic of Belarus: Large White, Duroc, Landrace, Pietrain and others. Large White pig, and based on it generated factory types dominate in the breeding livestock. The percentage of these four swine breeds represents at least 98% of the total population of *Sus scrofa domesticus* in Belarus. The breeders are oftenly faced with the objective to estimate breeding animals' breed-purity to correct inbreeding values in order to manage breed productivity indexes.

The goal of this study is to find breed-specific SNPs for domestic pigs, which are cultivated in Belarus: Large White, Landrace, Pietrain, Duroc and Meishan.(for comparison) SRA-data were analyzed by whole genome sequencing (NGS), located in the public domain of the DNAnexus cloud service (http://sra.dnanexus.com/) and NCBI-SRA server (http: //www.ncbi.nlm.nih .gov / sra). Search Basics were provided by the previously published results of Ramos A. et al. [1], produced using SNP-chip «PorcineSNP60» (Illumina).

Methods and Algorithms: Analysis was performed using SRA Nucleotide BLAST algorithm and program BioEdit v.7.2.5. The amount included in analysis SNP – 193; the number of reads for whole genome Large White pigs – 19, for other breeds – 71 (Landrace – 22, Pietrain – 6, Duroc – 28, Meishan – 15). The total number of analyzed sequences is 32 754 738 518.

Results: Finally, 55 breed-specific markers were identified among 193 SNPs, i.e. for those for which one allele is unique to only one breed. 6 specific SNPs were identified for the Large White breed (for which minor allele prevalence frequency was in the range of 7,9-23,7%); for Landrace – 5 (9,1-20,5%); for Pietrain – 7 (16,7-66,7%); for Duroc – 32 (7,1-100%); and for Meishan – 5 (23,3-100%).

Conclusion: Basing on these results it is expected in the future to develop and test SNP-markers genetic panel intended to provide the differentiation of domestic pig breeds, such as Large White, Landrace, Duroc, Pietrain and Meishan.

Availability: http://sra.dnanexus.com/, http://www.ncbi.nlm.nih.gov/sra

References:

1. A.M. Ramos et al. (2011) Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing, Animal Genetics, **42**: 613-620

IDENTIFICATION OF THE TAXA OF THE ORDER *ARTIODACTYLA* FOR CRIMINAL INVESTIGATION CASES OF ILLEGAL HUNTING

I.S. Tsybovsky*, S.A. Kotova, V.I. Rybakova, A.A. Rabcava, E.A. Spivak
Scientific and Practical Centre of the State Committee of Forensic Expertises, Minsk, Belarus
** Corresponding author: tsybovsky@ yahoo.com*

Key words*: microsatellite, wild animal, criminal identification , illegal hunting*

Motivation and Aim: Illegal removal of animals from their natural habitat represents a worldwide problem. In addition to negative economic impacts, illegal hunting can be detrimental to the environment because it can lead to uncontrollable changes in biological communities or reduce the population of protected species (e.g., European bison, *Bison bonasus*). In the Republic of Belarus, moose (*Alces alces*), wild boar (*Sus scrofa*), roe deer (*Capreolus capreolus*), and red deer (*Cervus elaphus*) are the main species subject to hunting. All these species belong to even-toed ungulates (order *Artiodactyla*) and are therefore closely related to each other as well as to domesticated animals such as the cow, sheep, goat, and pig. Investigation of illegal hunting activities may require the DNA-identification of either the species or the specific animal since biological specimens (e.g., blood) are morphologically indistinguishable.
Methods and Algorithms: We performed multiplex PCR of STR loci using fluorescently labeled primers. PCR products were analyzed with capillary electrophoresis using automated sequencers manufactured by Applied Biosystems.
*Results***:** The phenomenon of targeted cross-species microsatellite amplification within the order *Artiodactyla* was investigated by using primers designed for STR loci of one species (original species) for genotyping of genetically related species (target species). We obtained non-amplifiable, monomorphic loci (i.e., loci with a single allele in all members of a species or family) as well as polymorphic loci (i.e., loci that have more than one allelic variant) and investigated the differences in molecular size of PCR products of different species. We propose the BM1824 locus of the cattle [1] as the key element of the test system for species identification. This locus produced monomorphic PCR products of 139 b.p. for the moose, red deer, roe deer and fallow deer (identification of the family *Cervidae*); PCR products exceeding 169 b.p. for the cattle, European bison, goat, and sheep (identification of the family *Bovidae*); and no PCR products not only for the human, horse, domestic cat, and dog but also for wild boar/pig (identification of the suborders *Ruminantia/Suiformes*). A multiplex test system was developed that combines primers for STR-loci of the cattle, deer, reindeer, goat, sheep, and pig. This system allows, via the analysis of PCR products, to determine complex patterns specific for various species of the order *Artiodactyla*, which are necessary for identification of species and other taxa.
*Conclusion***:** The identification of biological specimens of wild animals without defined morphological, anatomical, and physiological attributes can be conducted by DNA analysis based on cross-amplification.
*Availability***:** Forensic DNA analysis laboratories, determination of authenticity of meat products.
*References***:**
1. Genbank: G18394; UniSTS ID: 44288

MOLECULAR EVOLUTION OF YUCCA PROTEIN FAMILY

I.I. Turnaev*, V.V. Suslov, K.V. Gunbin, D.A. Afonnikov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
**-Corresponding author: turn@bionet.nsc.ru*

Key words: *Yucca proteins, Flavin monooxygenases, molecular evolution*

Motivation and Aim: The aim of our study was to determine the origin and evolution YUCCA protein family.
Methods and Algorithms: The homologous sequences of *Arabidopsis thaliana* YUCCA proteins were identified using BLASTP and DELTA-BLAST tools at NCBI. Protein sequence alignment reconstructed by Promals. Protein 3D structures were reconstructed by I-TASSER 4.4 web-service. After that, various protein structure refinement tools were used (FG-MD, ModRefiner, i3Drefine, GalaxyRefine). The identification of protein domain boarders was based on 3D protein structures; for this purpose, we used ThreaDom system.
Results and Conclusion: Analysis on the phylogenetic tree of plant and non-plant YUCCA homologs identified three groups of proteins which are differ by patterns of functional sites, conserve domains, and whole 3D structure: i) FMO proteins (including YUCCA proteins), almost all of them contain in site FMO motif FxGxxxHxxxY/F, 3 structural domains (SD), and short highly hydrophobic inner loop in central domain tightly covering active center; ii) Baeyer–Villiger monooxygenases, which all contain FxGxxxHxxxW motif, 3 SD, and long flexible highly hydrophilic inner loop in central domain locating near active center; iii) a group of proteins characterized by predominance of FxGxxxHxxxH motif and 4 SD. FMO motives i and ii is already recorded in [1], FMO motif is not identified in the literature. Our results suggest that these three groups of proteins have different enzymatic functions. Therefore FMO motif ii and iii groups of sequences could be considered as unrelated to FMO/YUCCA superfamily and should be excluded from further analysis.
We identified that closest homologs to YUCCA proteins are YUC-like proteins of soil bacteria and cyanobacterial proteins. Analysis of 3D structures of YUCCA proteins (YUCCA1, 2, 4-6, 10 *A. thaliana*) revealed three regions: N-terminal (1-141 aa by YUCCA10 by *A. thaliana*), central (142-311 aa) and C-terminal (312-383 aa). YUCCA proteins form a two-subunit structure: the first subunit composed of central region, and the second composed of the combination between N- and C-terminal regions. Phylogenetic analysis sequences of these three regions showed that the topology of their phylogenetic trees differ. The phylogenetic tree topology of the central YUCCA region match with standard species phylogeny of plants (in contrast to the tree topology of whole protein). However, the tree topology of both N- terminal and C-terminal YUCCA regions are differ from the plant species phylogeny. Hence, that the difference between the phylogenetic tree topologies of 3 regions of YUCCA proteins could be due to either the differences in evolutionary regimes, or the presence of horizontal transfer of DNA fragments which correspond to the regions.
Acknowledgements Protein 3D structure reconstruction by I-Tasser performed by KVG. The work supported by budget project № 0324-2015-0003.
References:
1. A. Riebel, et. al.. (2013), *J Mol Catal B Enzym,* **88:** 20-25.

LOOKING FOR PROTEOMIC MARKERS OF BREAST CANCER IN BLOOD EXOSOMES

O.S. Tutanov[1], S.N. Tamkovich[1]*, Y.S. Bakakina[2], L.V. Dubovskaya[2], Y.P. Tsentalovich[3], I.D. Volotovskiy[2], P.P. Laktionov[1]

[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*

[2] *Institute of biophysics and cellular engineers NASB, Minsk, Byelorussia*

[3] *Institute "International Tomografic Center" SB RAS, Novosibirsk, Russia*

* *Corresponding author: s.tamk@niboch.nsc.ru*

Key words: *proteomics, exosomes, blood, breast cancer*

Motivation and Aim: It is known that the body cells secrete exosomes into the extracellular space, including into biological fluids, such as blood. In contrast to tumor specific proteins circulating in blood in a very low concentration exosomes of tumor origin are readily isolated providing enrichment of tumor proteins and improve efficacy of their detection. Thus, exosome based tumor-specific protein assay represent a valuable tool for non-invasive diagnosis of malignant neoplasms.

Methods and Algorithms: Exosomes from blood plasma and cell-surface-bound exosomes are obtained as described [1] from blood of healthy females (n=5) and primary breast cancer patients (n=5, T1-2N0M0). Size distribution and concentration of the microparticles was estimated by nanoparticle tracking analysis (NTA) using NanoSight NS-300 (Malvern, USA); anti CD-63, CD-24, CD-9, CD-81 antibodies (BD Biosciences, USA) were used for characterization of exosomes by flow cytometry; protein concentration was measured by NanoOrange Protein Quantitation kit (Molecular Probes, USA). Exosomal proteins were separated by 2D-SDS PAAGE and identified by MALDI-TOF mass-spectrometry.

Results: NTA demonstrate presence of 65-175 nm membrane–wrapped particles in the preparations isolated from blood. The exposure of CD-63, CD-24, CD-9, CD-81 demonstrate isolation of mainly exosomes. Proteins ranged from 10 to 250 kDa were found in exosomes by 2D-SDS PAAGE and besides 8 regions of electrophoregram differ between healthy and cancer patients in expression level and number of proteins. MALDI-TOF with a score exceeding 56 reveals 186 proteins; 63 of which are found first time (according to the data on the basis of Exocarta March 2016).

Conclusion: The data obtained demonstrate that exosomes carry a number of not yet described proteins and represent a potential source of tumor-specific proteins for detection of breast tumors. Deep analysis of the data including expression level of discovered tumor related proteins is necessary for evaluation of their diagnostic value.

Acknowledgements: The present work was supported by the integration project of the National Academy of Sciences of Belarus (grant #M15CO-025) and the SB RAS (head Tamkovich SN) "Analysis of circulating blood exosomes under normal conditions and in breast cancer".

References:

1. S. Tamkovich et al (2015) The method for isolation exosomes from blood. Russian patent # 2556825 from 18.06.2015.

MITOCHONDRIAL DYSFUNCTION IN SPORADIC ALZHEIMER'S DISEASE-LIKE PATHOLOGY IN OXYS RATS

M.A. Tyumentsev[1]*, E.V. Kiseleva[1], V.A. Vavilin[2], N.G. Kolosova[1], N.A. Stefanova[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Molecular Biology and Biophysics SB RAMS, Novosibirsk, Russia*
* *Corresponding author: landselur@bionet.nsc.ru*

Key words: *Alzheimer's disease, brain aging, senescence-accelerated OXYS rats*

Motivation and aim: Alzheimer's disease (AD) is the most prevalent neurodegenerative disease resulting in brain tissue atrophy and dementia. Amyloid cascade hypothesis posits that central event in the AD pathogenesis is the accumulation of beta amyloid peptide. However, recent data indicate that Aβ hyperproduction might be a secondary event in the AD pathogenesis as most hallmarks of the disease predate it. Mitochondrial cascade hypothesis posits that mitochondrial dysfunction is the key factor coupling the vicious cycle of the AD pathogenesis. However, mechanisms launching this cascade remain cryptic and investigation is hampered by the lack of animal models faithfully reproducing human disease. Senescence-accelerated OXYS rats constitute a model of sporadic AD developing its main hallmarks: Aβ accumulation, tau hyperphosphorylation, neuronal loss, impaired learning and memory. Aim of our research is to study the role and nature of the mitochondrial dysfunction in the development of AD-like pathology in OXYS rats.
Methods and algorithms: In order to determine pathways participating in the development of the mitochondrial dysfunction in OXYS rats, we used RNAseq data from rats' hippocampi collected at the age of 18 months when the symptoms of the AD-like pathology are the most pronounced. With 93 differentially expressed genes (DEGs) belonging to the "mitochondrion" category of the DAVID database we have reconstructed gene network using GeneMANIA plugin. To establish the role of reactive oxygen species (ROS) in the pathogenesis of the AD-like pathology we measured ROS production by isolated brain mitochondria of 3 months old Wistar and OXYS rats. Age-related changes hippocampal neurons' mitochondrial network were studied using electron microscopy.
Results: RNAseq analysis showed differences in the mRNA levels of 93 genes from the "mitochondrion" category of which 55 genes are upregulated and 38 downregulated. Ten most connected nodes of the gene network were taken for further evaluation. The nodes correspond to genes involved in the energy metabolism, lipid metabolism, Aβ detoxification, proteostasis, and mitochondrial translation. Nine of these genes are upregulated. Six genes are connected to enzymatic activity; four genes are subunits of protein complexes. It is worth noting that we have not found significant changes in the antioxidant signaling. Additionally, ROS production by the OXYS brain mitochondria do not differ significantly from that of Wistar rats. Electron microscopy showed that OXYS mitochondria unlike those from Wistar rats fail to increase their quantity with age with this trend accompanied by the shutdown of mitochondrial dynamics as evidenced by decreased number of intermitochondrial contacts.
Conclusions: We have found that the AD-like pathology is characterized by changes in a diverse set of mitochondrial functions but not oxidative stress with condition's onset marked by changes in mitochondrial content and dynamics in the hippocampal neurons.
Acknowledgements: The work was supported by the RSF grant 16-15-10005.

REGULATION OF RIPKS IN CELL SURVIVAL AND CELL DEATH BY APOPTOSIS AND NECROPTOSIS, INSIGHTS AND THERAPEUTIC POTENTIAL

P. Vandenabeele[1, 2]
[1] *VIB Inflammation Research Center, Technologiepark 927, Zwijnaarde-Ghent, 9052, Belgium*
[2] *Department of Biomedical Molecular Biology, Ghent University, Technologiepark 927, Zwijnaarde-Ghent, 9052, Belgium*

Key words: *cell death, necroptosis, gene networks, drugs*

Necroptosis was initially identified as a backup cell death program when apoptosis is blocked. However, it is now recognized as a cellular defense mechanism against infections and is presumed to be a detrimental factor in several pathologies driven by cell death. Necroptosis is a prototypic form of regulated necrosis that depends on activation of the necrosome, a protein complex in which receptor interacting protein kinase (RIPK) 3 is activated which on its turn phosphorylates a plasma membrane destabilizing protein MLKL, resulting in cellular swelling and explosion. The RIP homotypic interaction motif (RHIM) is the core domain that regulates activation of the necrosome. To date, three RHIM-containing proteins have been reported to modulate the kinase activity of RIPK3 within the necrosome: RIPK1, Toll/IL-1 receptor domain-containing adaptor inducing IFN-β (TRIF), and DNA-dependent activator of interferon regulatory factors (DAI). RIPK1 is a key molecule determining cellular fate downstream of several innate immune receptors. It is a serine/threonine kinase consisting of an N-terminal kinase domain linked by a largely unstructured intermediate domain to a C-terminal death domain. In the TNF signaling pathway, RIPK1 paradoxically promotes cell survival as well as cell death. These opposite cellular functions are mediated by 2 distinct faces of RIPK1. Upon binding of TNF to TNFR1, RIPK1 is recruited to the TNF receptor 1 complex I where it acts as a scaffold protein promoting cell survival, in part, by activating the canonical NF-kB and MAPK pathways, but also by IKK-dependent phosphorylation of RIPK1. Specific conditions (blocking IAPs, TAK1 or IKKs) changes RIPK1 from a survival scaffold to a deadly kinase, which then regulates assembly of 2 possible cytosolic cell death-inducing complexes, namely complex IIb (RIPK1-FADD-Caspase-8) leading to apoptosis and the necrosome (RIPK1-RIPK3-MLKL) leading to necroptosis, requiring in both cases RIPK1 kinase activity. The precise molecular mechanisms controlling RIPK1 survival and cell death functions are currently unknown. Similarly, how RIPK1 kinase activity exactly contributes to either of both cell death modalities is largely unknown. Despite this lack of understanding, it is evident that RIPK1 plays a dual role downstream of TNFR1 and that its kinase activity therefore needs tight repression to avoid unnecessary damage to the organism. This crucial role of RIPK1 is also illustrated *in vivo* by the observation that RIPK1 controls survival and cell death by apoptosis or necroptosis of many cell types such as epithelial cells of intestine, skin and lung, but also haematopoietic cells and liver cells.

Targeting necroptosis can occur at three levels: blocking RIPK1 and RIPK3 kinase activity, and blocking MLKL. Novel drugs and known drugs have been identified in cellular phenotypic screenings which block necroptosis. These drugs are effective in blocking inflammatory, degenerative and infectious diseases. On the other hand induction of necroptosis has been found effective in inducing immunogenic cell death. Altogether these data illustrate the crucial role of RIPKs in homeostasis and pathologies, and provide profound therapeutic perspectives for their targeting.

MUTATIONS SPECTRA OF MAJOR ONCOGENES IN PATIENTS WITH MULTIPLE PRIMARY NEOPLASIA

G.V. Vasiliev[1]*, A.V. Savkova[2], A.V. Gerasimov[3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Centre for postgraduate medical education, NGU, Novosibirsk, Russia*
[3] *Novosibirsk regional clinical oncology hospital, Novosibirsk, Russia*
* *Corresponding author: genn@bionet.nsc.ru*

Key words: *sequencing, oncogenes, mutation spectra*

Motivation and Aim: The study of genetic predisposition to cancer has a high predictive value for early detection and proper treatment. The occurrence of multiple non-metastatic primary malignancy in one patient reveal a genetic predisposition to the development of cancer.
Methods: A total of 8 patients of Novosibirsk regional clinical oncology hospital with multiple primary neoplasia were included in this study. Patient information, including sex, age, tumor type and family hystory were recorded. DNA was extracted from 5 ml of whole human blood. The next generation sequencing Ion AmpliSeq™ Cancer Hotspot Panel v2 for Ion Torrent PGM was used to investigate generative mutations spectrum in the samples from all patients. The panel used targeted 207 amplicons encompassing 2800 known cancer-relevant variants across 50 cancer-related genes. Each sample was individually barcoded, all 8 samples was pooled prior E-PCR, loaded on a 316v2 Chip and sequenced according to the Ion PGM 200 Sequencing protocol. The average depth of total coverage was >200, each nucleotide coverage was >50. Sequencing reads were analyzed using Torrent Suite software program with the 'variant caller v4.0.2' plugin and aligned to the human reference genome, hg19, which was uploaded on the Ion Reporter software v4.2 to perform variant calling and mapping.
Results: For 8 patients a total of 94 polymorphic variants in 207 regions covering "mutation hotspots" in 50 tumor-related susceptibility genes where found. Number of mutations per patient vary from 9 to 17 homo- or heterozygous SNP. Among the 50 genes included in panel, 17 genes were found mutated. All 8 patients had variations in *FGFR3* gene. Most of patients had mutations in *TP53, EGFR, PDGFRA* genes. Six patients in our cohort had at least two hotspot mutations associated with cancer according with COSMIC database.
Conclusion: Patients with multiple primary neoplasia revealed a large number of polymorphic variants in the major oncogenes. This data confirms the assumption of strong genetic predisposition to the development of cancer for patients with multiple primary neoplasia.

PARASITES OF THE GENERA *NOSEMA*, *APICISTIS, CRITHIDIA* AND *LOTMARIA* IN THE NATURAL HONEYBEE AND BUMBLEBEE POPULATIONS: A CASE STUDY IN INDIA

V. Vavilova[1]*, I. Konopatskaia[1, 2], M. Woyciechowski[3], S. Luzianin[4], A. Blinov[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Environmental Sciences, Jagiellonian University, Krakow, Poland*
[4] *Biological Faculty, Kemerovo State University, Russian Federation*
** Corresponding author: valeriya-vavilova@bionet.nsc.ru*

Key words: *bumblebees, honeybees, parasite, Nosema, Apicystis, Crithidia, Lotmaria, ribosomal gene cluster, genetic variant*

Motivation and Aim: In recent decades the populations of important pollinators honeybees and bumblebees decrease around the world. A variety of pathogens and parasites contribute to significant honeybee and bumblebee colonies loss. Fungal parasites of the genus *Nosema* (Microsporidia: Nosematidae) and protozoan parasites of the genera *Apicystis* (Apicomplexa: Ophryocystidae), *Crithidia* and *Lotmaria* (*Kinetoplastida*: Trypanosomatidae) have a significant negative impact on the honeybee and bumblebee colonies. Recent studies of nuclear DNA markers from these parasites allowed to describe new species and genetic variants. Thus, investigation of *Nosema*, *Apicystis*, *Crithidia* and *Lotmaria* species variability in honeybee and bumblebee populations from previously unstudied territories of states *Karnataka and Jammu and Kashmir* (India) will allow to identify novel genetic variants of the parasites.
Methods and Algorithms: We have analyzed distribution and diversity of parasites in Indian honeybee and bumblebee populations using bioinformatical and molecular biological methods.
Results: Identification of the parasites was conducted, using PCR with primers specific for the ribosomal RNA gene cluster of *Nosema*, *Apicystis*, *Crithidia* and *Lotmaria* species followed by sequencing. In total 80 individual honeybees specimens were tested; twenty and ten specimens revealed presence of *Nosema* and *Lotmaria* parasites, respectively. No honeybee's specimen with *Crithidia* and *Apicystis* infection was found in India. Among 39 tested bumblebee specimens four were infected with *Nosema* spp. and twelve with *Crithidia* species, while no sample infected with *Lotmaria* spp. and *Apicystis* spp. was detected. Comparative analysis of ribosomal RNA genes showed that studied honeybee samples were infected by representatives of two microsporidia species (*N. ceranae* and *N. bombi*) and one trypanosomatid species (*Lotmaria passim*). The bumblebee specimens were infected by *Nosema D* genetic variant, which was previously described in bumblebee populations from China, and two *Crithidia* species (*C. bombi* and *C. expoeki*).
Conclusion: Thus, in the present study the distribution of *Nosema*, *Apicystis*, *Crithidia* and *Lotmaria* parasites were investigated in the natural populations of honeybees and bumblebees from states *Karnataka and Jammu and Kashmir (In*dia). For the first time *N. bombi* infection was detected in honeybee populations.
Availability: The obtained nucleotide sequences of *rRNA* genes of *Nosema* spp., *Crithidia* spp. and *L. passim* were deposited to GenBank and European Nucleotide Archive.

EFFECTS OF LAMBERTIANIC ACID AMIDE ON EPILEPTIFORM ACTIVITY IN HIPPOCAMPAL SLICES INDUCED BY PICROTOXIN OR MAGNESIUM-FREE MEDIUM

S.O. Vechkapova, A.L. Proskura*, T.A. Zapara, E.D. Sorokoumov, A.S. Ratushnyak
Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia
** Corresponding author: annleop@mail.ru*

Key words: *hippocampal slices, epileptiform activity, NMDA receptor, lambertianic acid amide, picrotoxin, magnesium-free solution*

Motivation and Aim: An imbalance between excitatory and inhibitory mediator systems in CNS leads to the development of a number of neurodegenerative diseases. Both hyperactivity of glutamatergic system and decreasing the activity of GABAergic leads to a moderate neuron membrane depolarization, which relieves the magnesium blockade of the NMDA receptor and results in excitotoxicity. Currently, drugs with glutamatergic mechanisms of action are being developed for the treatment of cognitive disorders and neurodegenerative processes. The aim of the present study is to test the antiepileptic effect of lambertianic acid amide (AmLA), which has been shown to be a promising compound that could be used in the synthesis of new pharmaceutical reagents [1,2].
Methods and Algorithms: The experiments were carried out on the hippocampal slices of ICR male mice using standard electrophysiological techniques. Stimulation of Schaffer collaterals and registration of induced population spikes of pyramidal neurons in the CA1 field were made using the glass microelectrodes, filled with saline. The epileptiform activity in the pyramidal neurons was *induced* by treatment the slices by picrotoxin or by magnesium-free medium.
Results: The application of AmLA in concentration of 170 μM significantly decreased the epileptiform activity or fully terminated it. The preincubation of slices with AmLA for an hour before applying epileptiform conditions prevented the development of epileptiform activity. Also incubation of the slices in normal solution with AmLA did not affect the initiation of NMDA-dependent synaptic potentiation.
Conclusion: Thus, AmLA (produced from Siberian cedar) helps to normalize the activity of hippocampal neurons both in glutamatergic system hyperactivation and the lack of GABAergic inhibition.
Acknowledgements: The work was supported by VI.35.1.5 basic project of fundamental researches of RAS and RFBR grant 15-29-04875.
References:
1. T.G. Tolstikova et al. (2001) Nootropic activity of lambertianic acid derivatives. Doklady Biological Sciences, 376: 8-9.
2. T.G. Tolstikova et al. (2004) Neurotropic activity of lambertianic acid adducts with N-substituted maleinimides. Pharm. Chem. J., 38: 532-534.

ARGO_CUDA: A FULL-EXHAUSTIVE GPU BASED APPROACH FOR A MOTIF DISCOVERY IN THE LARGE DNA DATASETS

O.V. Vishnevsky[1, 2]*, A.V. Bocharnikov[2], N.A. Kolchanov[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: oleg@bionet.nsc.ru*

Key words: *degenerated oligonucleotide motif, transcription regulation, CUDA, GPU*

Motivation and Aim: A motif discovery in ChIP-Seq datasets remains a challenging issue. A low effectiveness of classic heuristic motif discovery approaches on a whole ChIP-Seq datasets forces the researchers to take into analysis only a fraction of top "peak" segments.
Methods and Algorithms: Argo_CUDA web service is designed to process the massive DNA data. This program for detection of degenerate oligonucleotide motifs of fixed length is based on the full-exhaustive approach and uses high-performance GPU technologies.
Results: We compared an effectiveness of Argo_CUDA and Info-gibbs [1]. Info-gibbs is a Gibbs sampling algorithm that compares well with existing heuristic methods like MEME, BioProspector, Gibbs or GAME on both synthetic and biological data sets. The sets of random sequences of 128bp in length and of 100, 1000, and 10000 sequences in size were generated. The sample motifs of different degeneracy level were placed in a 60 percent of the sequences. The similarity between a motifs obtained by the programs and the sample motifs were measured with the average Kullback-Leiber distance (KLD).
Conclusion: An effective web service for motif discovery in ChIP-Seq datasets is developed. It is not as fast as a classic heuristic approaches, but it considerably reduces the restrictions on the size of the sample under analysis.
Availability: wwwmgs.bionet.nsc.ru/mgs/programs/Argo_CUDA.
Acknowledgements: The work was supported by the budget project 0324-2015-0003.
References:
1. Defrance M, van Helden J. (2009) Info-gibbs: a motif discovery algorithm that directly optimizes information content during sampling. *Bioinformatics* 25(20):2715-22.

THE IMPACT OF HUMAN GENETIC VARIABILITY ON LIGAND-PROTEIN INTERACTIONS AND INDIVIDUAL DRUG RESPONSE

P.K. Vlasov*, O. Pich I Rosello, A.V. Vlasova, F.A. Kondrashov
Centre for Genomic Regulation, Universitat Pompeu Fabra, Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain
** Corresponding author: peter.vlasov@crg.eu*

Key words: *ligand-protein interactions, drugs, adverse drugs events, human genetic variability*

The mechanism of action for majority of modern therapeutics is directly related to the interaction of the drug molecule with its target, typically a protein. Adverse drug events (ADEs), instances when a medication causes an unintended response, contribute substantially to morbidity, the cost of treatment and often appear unpredictably [1, 2]. Genetic variability is thought to account for a substantial fraction of the individual drug response in humans [3, 4, 5] - meanwhile our understanding of the contribution of such variability to the causes of individual drug response remains fragmented. In our project we combined genome-wide data on human single nucleotide polymorphisms (SNPs) with structural data on drug-protein complexes. Using data from 1000genome project and The Cancer Genome Atlas (TCGA) consortium, at the genome-wide scale we identify all SNPs potentially affecting the proteins binding affinity for drugs, drug-like compounds and metabolites. Our results suggest that SNPs with a serious impact on ADE are present in most individuals, however, most of such polymorphisms are rare requiring a personalized approach to their identification. So, the genetic component for many ADEs may be highly personalized with each individual carrying a unique set of relevant SNPs. The reduction of ADEs may, therefore, primarily rely on the application of computational genome analysis in the clinic rather than the experimental study of common SNPs.

References:

1. Rodríguez-Monguió R, Otero MJ, Rovira J. Assessing the economic impact of adverse drug effects. Pharmacoeconomics. 2003;21(9):623-50.
2. Boeker EB, de Boer M, Kiewiet JJ, Lie-A-Huen L, Dijkgraaf MG, Boermeester MA. Occurrence and preventability of adverse drug events in surgical patients: a systematic review of literature. BMC Health Serv Res. 2013 Sep 28;13:364.
3. Evans WE, McLeod HL. Pharmacogenomics-drug disposition, drug targets, and side effects. N Engl J Med. 2003 Feb 6;348(6):538-49.
4. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. N Engl J Med. 2011 Mar 24;364(12):1144-53.
5. Relling MV, Evans WE. Pharmacogenomics in the clinic. Nature. 2015 Oct 15;526(7573):343-50.

CHARACTERISTICS OF *ACDS*-GENE OF BACTERIA *PSEUDOMONAS PUTIDA* B-37 RESPONSIBLE FOR ACC-DEAMINASE SYNTHESIS

D.S. Volkava*, S.I. Leanovich, A.A. Melnikava, E.A. Khramtsova
Belarusian State University, Minsk, Republic of Belarus
** Corresponding author: 95volkovads@gmail.com*

Key words: *ACC-deaminase, acdS-gene, Pseudomonas putida, ethylene, transgenic plants*

Nowadays one of the main problems in agriculture to be solved is plants' resistance to the numerous environmental factors expanding every year due to the active anthropogenic intervention. Plant's natural reaction on stress is the production of stress hormone ethylene that inhibits growth and development of plant organism during unfavorable periods of time. This process attends to decrease of the biomass production which is very unprofitable for agriculture. One of the most perspective ways to decline the level of stress ethylene is creating transgenic plants with *acdS*-gene coding for 1-aminocyclopropane-1-carboxylate (ACC-deaminase). ACC-deaminase is required for increasing the concentration of ethylene's precursor, ACC. This process promotes roots elongation, tuber forming, and biomass accumulation. The aim of current work was analysis of the primary nucleotide sequence of *acdS*-gene of *Pseudomonas putida* strain B-37 for further development of the recombinant plant cells *Nicotiana benthamiana* and *Nicotiana tabacum*. Primary nucleotide sequence of ACC-deaminase gene from *P. putida* B-37 was analyzed using programs available on on-line (NCBI, ExPASy, NPS@) and off-line resources. Search for homologs was conducted using BLASTn, and analysis of conserved domains – in Conserved Domains, available on-line on NCBI resource. To estimate ACC-deaminase gene its amino acid sequence was constructed using Translator on ExPASy resource (CDS size – 1017 b.p.). Molecular weight, amino acid composition, estimated half-life, theoretical pI were valued using ProtParam on ExPASy resource. Secondary structures were forecasted by the consensus prediction from the multiple alignments using SOPMA on NPS@. Phylogenetic tree was constructed using Neighbor-joining method implemented in MEGA 6.0. It was shown that analyzing sequence has high homology with ACC-deaminase genes from different species of *Pseudomonas* genus. In the protein coding for open reading frame of this gene was detected Aminocyclopropane-1-carboxylate deaminase (ACCD) domain refers to tryptophan synthase beta superfamily (fold type II). Molecular weight is above 36718.9 D, mean theoretical pI = 5.6, estimated half-life in *Escherichia coli* is above 10 hours. Among secondary structures were mostly predicted alpha helixes (31.95%) and random coils (34.91%). Assumptive 3D structure of analyzing protein was modeled using SWISS-MODEL on ExPASy. To estimate divergence and homology of analyzing protein and ACC-deaminase proteins of different species of *Pseudomonas* phylogenetic tree was constructed; the highest homology with analyzing protein was obtained for ACC-deaminases from *Pseudomonas putida* strains. Analysis of the primary nucleotide sequence from *P. putida* B-37 showed the presence of ACC-deaminase gene that has high homology with the sequences coding for *acdS*-genes from other species and strains of *Pseudomonas* genus. This gene was isolated and cloned in vector with broad host range pBI121 which was used to create the recombinant plant cells *N. benthamiana* and *N. tabacum*. Such plants are assumed to be more resistive to the unfavorable environmental factors due to the ACC-deaminase synthesis.

AMPLISEQ ™: AMPLIFICATION AND SEQUENCING

I.A. Volkov
Department of scientific and methodological support of "Khimexpert Agency", Moscow, Russia
** Corresponding author: ilyavolkov83@gmail.com*

Key words: *AmpliSeq, new generation sequencing (NGS), genome libraries, Ion Torrent ™, Ion S5 ™*

High-performance sequencing in a relatively short period of time allows obtaining large tracts of genetic information. During the period of accumulation of knowledge it is important to focus research on a limited number of targets. Ultramultiplex technology PCR followed by sequencing AmpliSeq ™ - this is the best method for mass screening of a large number of targets. AmpliSeq ™ panels are useful for sequencing indels (50 bp) and SNP, groups of genes, RNA molecules; there is a panel of sequencing exome person. The online resource Ion AmpliSeq ™ Designer (ampliseq.com) allows you to quickly create a research panel, which is a multiplex primer pool. The number of primers in a single pool may be from 12 to 6144 pairs allowing one panel genome sequence block size from 1 kb 5 mln.b.p. The primers in the same pool did not overlap, and overlapping primers are carried in different pools. Typically, the panel consists of two pools of primers. The panels for the analysis of point mutations and polymorphisms by the script HotSpot Designs, consist of 1 primer pool. Panels are generated automatically according to the Pipeline with maximum coverage of amplicons. Ion AmpliSeq ™ Designer automatically selects primers for amplicon of about 200 bp and 400 bp (Length of the readings on the Ion Torrent ™ platforms).

There is enough only 10 ng DNA / RNA on 1 reaction to search for genetic variants and evaluation of gene expression. There is a considerable amount of ready-branded and custom AmpliSeq ™ panels that are available for use: more than 20 panels designed for the study and diagnosis of different groups of hereditary diseases. Panels include from one hundred to four hundred or more genes associated with the development of diseases. Automatic analysis of the results includes data processing software Torrent Suite ™, integrated in devices Ion Torrent ™. Evaluation of the quality of the results obtained by the coating is carried out using coverage Analysis plugin. Ion Reporter ™ software helps to interpret and use data annotations. Cloud resource Ion Reporter ™ appeals to a large number of databases, annotates the data and the results of the report in the analysis of the results.

Thus, the establishment of libraries AmpliSeq ™ method, sequencing at Ion Torrent ™ platforms and analysis Torrent Suite ™ and Ion Reporter ™ makes the targeted method of sequencing a simple, convenient and fast to use in dealing with routine and research applications.

COUPLED MOLECULAR DYNAMIC AND CONTINUUM ELECTROSTATIC METHOD TO COMPUTE IONIZATION OF PROTEINS AS A FUNCTION OF PH

Yu.N. Vorobjev
Institute of Chemical Biology SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: ynvorob@niboch.nsc.ru*

Key words: *proteins, ionization constants*

Motivation: Accurate prediction of ionization constant of amino acids in proteins is a long standing problem. A computational method is developed to calculate the pKa values of ionizable residues Asp, Glu, His, Tyr and Lys of proteins.
Methods: Calculations of electrostatic energy of proteins is based on an effective version of continuum dielectric electrostatic model developed by us. A conformational flexibility is modeled by the method of molecular dynamics of 10 ns of length in an implicit water solvent.
Results: The accuracy of proposed method of calculation of pKa values is estimated for a test set of proteins with experimental pKa data for 297 ionizable residues of 34 proteins. The pKa prediction test shows that 57%, 86% and 95% of all predictions have an error lower than 0.5, 1.0 and 1.5 pK units, respectively. In total, our method of pKa prediction demonstrates a good accuracy, which it treed protein flexibility by natural way as protein molecular dynamics in water solvent. Computer program is available by request from professor Yu.N. Vorobjev.

The work is supported by grant RFFI #*15-04*-00387 and by program of RAS MCB *(6.11), R SF #16-14-10038*

METHODS TO CALCULATE P-VALUE OF RNA OF A DEFINITE SHAPE

D.G. Vorobyev*, V.V. Solovyev
Softberry Inc., Novosibirsk, Russia
** Corresponding author: denis.g.vorobiev@gmail.com*

Key words: *RNA, secondary structure, pattern, P-value, partition function*

A model of RNA structure ('*Model M*') is considered that is given by a tree topology, size ranges of elements and size ranges of total or subpattern length. Our program ('*Matcher*') is able to rapidly search *model M* in long sequences. In this work, three methods of calculating the frequency **P(*model M*)** of such pattern in a random sequence are suggested. Two of these methods use the idea of partition function. Energy is also taken into account. Our *Matcher* is similar to the thermodynamic matcher from [1], except that it uses a more specifically defined shape. For each occurrence, the program outputs a best fold, its energy and the total number of possible folds corresponding to *Model M*. How frequently *Model M* occurs in a random sequence? In a model with fixed-sized elements, $\mathbf{P}(\boldsymbol{Model\ M}) = \prod \mathbf{P}_i$, where $\mathbf{P}_i$ is the *i*-th stem probability, which values can be preliminary tabulated by sampling. When elements vary in length, the sum of probabilities of all theoretically possible folds of *Model M*, the partition function ***S***, can be calculated. The calculation is performed with a dynamic programming procedure. It is clear that different folds occur interdependently and tend to form big clusters. Intuitively, dividing ***S*** by the average number ***C*** of folds in a single cluster, we get the desired **P(*Model M*)**. It can be shown that the form of the distribution of ***C*** makes no difference. Thus, $\mathbf{P}(\boldsymbol{Model\ M}) = \boldsymbol{S} / \mathbf{M}(\boldsymbol{C})$, where **M(*C*)** is the expectation of ***C***. Because we are rather interested in stable structures, we should calculate the probability $\mathbf{P}(\boldsymbol{Model\ M,\ E<E_t})$ to meet *model M* with energy ***E*** under given threshold $\boldsymbol{E_t}$. Obviously, $\mathbf{P}(\boldsymbol{Model\ M,\ E<E_t}) = \mathbf{P}(\boldsymbol{Model\ M}) \cdot \mathbf{P}(\boldsymbol{E<E_t} \mid \boldsymbol{Model\ M})$. To obtain $\mathbf{P}(\boldsymbol{E<E_t} \mid \boldsymbol{Model\ M})$, we need a set of random occurrences of *model M*. Since, in general, we can't get random occurrences by a lengthy simulation, we have to imitate them. We generate them by allowing pairs that are usually forbidden ("non-pairs"). We assign them energies with a big positive value, 50 kcal/mol, making them extremely unfavorable. This way, *Matcher* is able to fold any random fragment into a structure given by *Model M*. *Matcher* minimizes the number of non-pairs. In the calculated structure, we substitute non-pairs with GC, AU or GU pairs and, thus, get a sequence fragment that is able to fold according to *Model M* with normal complementarity rules. Having a big number of 'random' occurrences, we obtain both the estimate of **M(*C*)** and the energy distribution, which turns out to be near Gaussian. Thus, we can estimate $\mathbf{P}(\boldsymbol{E<E_t} \mid \boldsymbol{Model\ M})$. The 2-nd way to estimate **P(*Model M*)** is to obtain the distribution **P(*X*)** of number ***X*** of non-pairs in random sequences. The distribution turns out to be Gaussian-like as well. We need its value at $\boldsymbol{X}$=0 meaning 'no non-pairs'. The value of **P(*X=0*)** will be our estimate of **P(*Model M*)**. The 3-rd way to estimate **P(*Model M*)** is to obtain the spectral decomposition of ***S***. In such a way the value of $\boldsymbol{S(E<E_t)}$ can be calculated. At high $\boldsymbol{E_t}$ values, $\boldsymbol{C(E<E_t) \rightarrow 1}$. Therefore, $\boldsymbol{S(E<E_t)}$ is our upper estimate of $\mathbf{P}(\boldsymbol{E<E_t, Model\ M})$ made without sampling of 'random' occurrences. We have suggested 3 methods of calculating P-value of RNA pattern. They can be used along with context measures to search unknown ncRNAs by clustering similar structures. The software will be available at corporate web-site soon.

References:

1. S. Janssen, R. Giegerich (2015) The RNA shapes studio, *Bioinformatics*, **31:** 423-425.

HOCOMOCO COMPREHENSIVE MODEL COLLECTION AS A PRACTICAL GATEWAY TO REGULATORY MOTIF-OME OF HUMAN AND MOUSE TRANSCRIPTION FACTORS

I.E. Vorontsov, Y.A. Medvedeva, V.J. Makeev, I.V. Kulakovskiy*
Vavilov Institute of General Genetics, Moscow, Russia
Engelhardt Institute of Molecular Biology, Moscow, Russia
** Corresponding author: ivan.kulakovskiy@gmail.com*

Key words: *transcription factor binding motifs, human, mouse, ChIP-Seq, HT-SELEX*

Motivation and Aim: Knowledge of sequence motifs resembling transcription factor binding sites is beneficial for a vast array of studies in regulatory genomics.
Methods and Algorithms: Using ChIPMunk [1] motif discovery tools we performed *de novo* motif discovery in more than two thousands data sets for human and mouse transcription factors studied by ChIP-Seq (in vivo, obtained from GTRD [2]) and HT-SELEX (in vitro [3]). The newly created binding models were benchmarked against known binding patterns for mammalian transcription factors.
Results: We present the latest release of the HOCOMOCO COmprehensive MOdel COllection [4] that provides binding models for 6 hundreds of human and almost 4 hundreds of mouse transcription factors. The primary collection provides classic mononucleotide position weight matrices (PWMs) which are linked with the hierarchical classification of transcription factors [5]. In addition, new release of HOCOMOCO includes dinucleotide position weight matrices based on ChIP-Seq data and a set of command-line java tools to facilitate motif finding with HOCOMOCO models.
Conclusion: We present a complete workflow used to build HOCOMOCO and discuss practical applications of the HOCOMOCO *motif*-ome in regulatory genomics.
Availability: HOCOMOCO and all the supporting tools are freely available online: http://hocomoco.autosome.ru and http://opera.autosome.ru.
Acknowledgements: This study was supported by RFBR grants 15-34-20423 and, partly, 14-04-01838.
References:
1. http://autosome.ru/ChIPMunk/
2. http://gtrd.biouml.org
3. A. Jolma A., J. Yan, T. Whitington, J. Toivonen, K.R. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, et al. (2013) Cell, 152:327-339.4.
4. I.V. Kulakovskiy, I.E. Vorontsov, I.S. Yevshin, A.V. Soboleva, A.S. Kasianov, H. Ashoor, W. Ba-Alawi, V.B, Bajic, Y.A. Medvedeva, F.A. Kolpakov, V.J. Makeev (2016) Nucleic Acids Res, 44(D1):D116-25.
5. E. Wingender, T. Schoeps, J. Dönitz (2013) Nucleic Acids Res, 41:D165-D170.

THE FREQUENCY, SPECTRUM AND FUNCTIONAL SIGNIFICANCE OF MUTATIONS IN CODING SEQUENCE OF TP53 GENE IN RUSSIAN PATIENTS WITH DLBCL

E.N. Voropaeva[1]*, T.I. Pospelova[2], M.I. Voevoda[1], V.N. Maximov[1, 2]
[1] *Institute of Therapy and Preventive Medicine, Novosibirsk, Russia*
[2] *State Medical University, Novosibirsk, Russia*
** Corresponding author: vena.81@mail.ru*

Key words: *diffuse large B-cell lymphoma, mutations, gene TP53, sequence analysis*

Motivation and Aim: It is shown that the current programs of diffuse large B-cell lymphoma (DLBCL) treatment including the target therapy are not able to overcome the adverse prognostic significance of mutations in *TP53* gene. Information about the frequency and spectrum of *TP53* mutations in the Russian patients with DLBCL in the current version of the IARC *TP53* Mutation Database R17 is not represented.
The goal of this work was to study the frequency, spectrum and functional significance of *TP53* mutations in Russian patients with DLBCL.
Methods and Algorithms: Presently 74 patients are included in the study. Diagnosis of DLBCL was determined according to the criteria of the latest WHO classification system 2008. Genomic DNA was isolated from formalin-fixed, paraffin embedded tissue blocks. Tissue sections contained at least 80-90% of the tumor cells. Direct sequence analysis of gene *TP53* was performed according to the IARC protocol (2010 update). Assessment of the damaging effect of amino acid substitutions on the function of the p53 protein was performed using resource Polyphen-2.
Results: 28 mutations in the coding sequence of the *TP53* gene in 74 samples were identified: 18 (64,29%) – missense, 6 (21.43%) – silent, 2 (7.28%) – nonsense, 1 (3,50%) – frameshift and 1 (3,50%) – splice. Four patients had a multiple mutations. The following findings have met several times in the cohort: p.Cys275Ser (5 patients) and p.Val272Glu, p.Thr155Ile and p.Arg212Term (in the two cases each).
The potentially protein function damaging of *TP53* mutations were found in 17 (22.9%) patients, including the p.Arg212Term, p.Ala189fs and IVS6-36 (affect the splicing according to the Human Gene Mutation Database). All of the possibly or probably damaging missense mutations with the exception of p.Pro316His are located in the region from 5 to 8 exons which encoding DNA binding region (DBR) implicated in transcriptionally-dependent and independent function of the p53 protein.
Conclusions: The mutations rate in our study is in good agreement with other studies which the frequency of the *TP53* mutations in patients with DLBCL ranged mostly from 13 to 23%. Codons 275, 155, 272 and 212 were the “hot spots” of mutation in our study. High frequency (95%) of the *TP53* mutations in regions of the gene encoding the DBR is a reflection of their functional selection. The assessment of immune-chemotherapy effect in the studied cohort of patients with DLBCL in addition to the *TP53* mutation status is planned.

HOW SEQUENCE AND STRUCTURE AFFECT THE MIRNA MATURATION

P.S. Vorozheykin[1]*, I.I. Titov[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: pavel.vorozheykin@gmail.com*

Key words: *microRNA, miRNA, secondary structure, biogenesis, SNP*

Motivation and Aim: MiRNAs expression is crucial for various developmental and cell processes; it is important to control the miRNA maturation and provide diverse miRNA expression patterns. The miRNA biogenesis proceeds in several steps, each of which is a subject to regulation by a range of mechanisms involving numerous protein-RNA and RNA-RNA interactions.
Results: Following the miRNA biogenesis scenario, we describe the influence of pri-/pre-/miRNA primary and secondary structure on the processing. We supplement these data with our statistical observations of miRNA properties. In particular, we discuss how the precursor loops change the Dicer/Drosha cleavage sites and affect the miRISC loading process and the miRNA expression level. Then we review the other regulatory elements: the conserved motifs in the terminal loop, the stem and flanks of the pre-miRNAs. Through them the miRNAs can regulate its own processing and the maturation of others using the RNA-RNA bindings. Besides, some of the motifs are involved in feedback loops and protein-RNA interactions. Then we focus on the single nucleotide polymorphisms in the pri-/pre-/miRNA sequences, their genome-wide characteristics and the influence on the miRNA functions and evolution. Finally, we propose a structure-based framework of the miRNA precursors, which can be useful for understanding the regulation of the biogenesis processes and for developing new miRNA prediction tools.

DRAFT GENOME SEQUENCE OF *STREPTOMYCES SP.* IB 2014 011-1 ISOLATED FROM LAKE BAIKAL MACROINVERTEBRATES

I.V. Voytsekhovskaya[1], D.V. Axenov-Gribanov[1]*, B.T. Tokovenko[2], Y.V. Rebets[2], E.S. Protasov[1], A.N. Luzhetskyy[2], M.A. Timofeyev[1]

[1] *Irkutsk State University, Institute of Biology, Irkutsk, Russia*

[2] *Helmholtz Center for Infectious Research, Helmholtz Institute for Pharmaceutical Research Saarland, Saarbrücken, Germany*

** Corresponding author: irina.voytsekhovskaya@gmail.com*

Key words: *Actinobacteria, Baikal Lake, biodiversity, Streptomyces sp. IB2014 011-1*

Motivation and Aim: Unique ecosystems with specific environmental conditions is a promising source for isolation of new actinobacteria strains. Lake Baikal is one of the greatest examples of ecosystem with high species biodiversity and endemicity caused by the long isolated evolution. The main aim is estimation of the *Streptomyces sp*. IB2014 011-1 strain to produce natural compounds through draft genome sequence.

Methods and Algorithms: The *Streptomyces sp.* IB2014 011-1 strain was isolated from *Trichoptera sp*. larvae collected from the bottom of Lake Baikal close to the Listvyanka settlement. The raw sequencing data was obtained using Illumina HiSeq 2500 technology. High molecular mass DNA was extracted from *Streptomyces* sp. IB2014 011-1. Manufacturer-recommended standard protocol was used to prepare two paired-end libraries. After quality control, only the 2nd library was used for genome assembly using SPAdes v3.7.A total of 3985 contigs were assembled, of them 84 longer than 1kbp. Scaffolding was performed by SSPACE 2.1 Premium using both libraries, and resulted in 49 scaffolds. Genome annotation was performed using prokka and antiSMASH v.3, followed by manual GenBank pre-submission curation. 16S rRNA delineation was performed using both ARB-SILVA database and NCBI's non-redundant database blast search. 16S rDNA sequences were multiple-aligned using MAFFT v7.222 (algorithm: auto, scoring matrix: 200PAM / k=2, gap open penalty 1.53, offset value 0.123). The phylogenetic consensus tree was built and formatted using Geneious 9.0.4 (Tamura-Nei model, NJ tree build method, *S. avermitilis* as an outgroup, 1000 bootstrap replicates).

Results: The genome of *Streptomyces* sp. IB2014 011-1 has a total length of about 8.1 Mbp, including a possible 100 kbp plasmid (scaffold STIB_19). The GC content, the number of protein coding genes, tRNA and rRNA genes are in accordance with other streptomycetes.

Conclusion: The chromosome of *Streptomyces sp.* IB2014 011-1 contains 31 putative gene clusters involved in the biosynthesis of secondary metabolites (or 84 gene clusters, if we also include ClusterFinder predictions)

Acknowledgements: This study was supported by the Ministry of education and science of Russian Federation as a part of Goszadanie projects (№6.382.2014/K, 6.734.2016 DAAD, 6.696.2016 DAAD), Russian science foundation (project N 14-14-00400), Russian foundation for basic research (projects N 14-04-00501), Grants of Irkutsk State University for researchers and Deutscher Akademischer Austauschdienst.

ACTUAL APPROACHES FOR QUALIFICATION AND QUANTIFICATION OF PROTEOME CHANGES

E.P. Vrzheschch
Bio-Rad
** Corresponding author: evgeniy_vrzheshch@bio-rad.com*

Ключевые слова: *протеомика, белок, экспрессия, гель-электрофорез, вестерн-блоттинг, нормализация, HKP, total protein normalization*

В современном научном мире все большее значение придается сложным многокомпонентным исчерпывающим исследованиям различных процессов, происходящих в живых системах. Количество публикаций, включающих “омиксные” подходы, поиск изменений экспрессии множества генов и белков, например, по ключевым словам геномика и/или протеомика, составляет несколько сотен тысяч. В то же время правильная постановка эксперимента требует учета множества факторов, которые не всегда очевидны.

В докладе будут рассмотрены различные варианты подходов к планированию и постановке эксперимента для определения качественных и количественных изменений экспрессии генов и белков, начиная со способов

- проведения пробоподготовки
- выделения и очистке белков и нуклеиновых кислот из образцов различного происхождения, в том числе и ограниченных по объему
- проведения одно- и
- двумерного гель-электрофореза
- вестерн-блоттинга

Дополнительно будут рассмотрены подходы к нормализации результатов вестерн-блоттинга по house-keeping proteins (HKP) и общему белку (total protein normalization, TPN), а также приведены рекомендации редакторов ведущих журналов по постановке и оформлению результатов эксперимента (MIQE and JBC guide-lines).

MULTIDIMENSIONAL PATTERNS OF METABOLIC RESPONSE IN ABIOTIC STRESS-INDUCED GROWTH OF ARABIDOPSIS THALIANA

B.S. Yadav[1]*, S. Freilich[2]*, E. Katz[1], A. Finkelshtein[1], D.A. Chamovitz[1]
[1] *Department of Molecular Biology and Ecology of Plants, Tel Aviv University, Israel*
[2] *Newe Ya'ar Research Center, Agricultural Research Organization, Ramat Yishay, Israel*
** Corresponding author: brijeshbioinfo@gmail.com*

Key words: *Abiotic Stress, Metabolic Response, Expression data, Arabidopsis thaliana*

The global climatic pattern is becoming increasingly unpredictable resulting in an average yield loss of more than 70% for major crops and losses estimated in hundreds of billions of US dollars each year. The awareness of stress damages leads to a demand for stress-tolerant crop varieties. The AtGenExpress experiment provided the first controlled four dimensional expression profile dataset, focusing on the effect of eight abiotic stress types on different plant parts of *Arabidopsis thaliana* at several time points, leading to the identification of stress associated genes. We aim to go beyond the gene-level characterization and further contextualized the discrete information into (1) a process-level signature of stress-specific, time-specific, and tissue-specific responses and (2) identify patterns of response progression across a time axis. To gain a functional perspective, pathways associated with the differentially expressed genes were characterized for each experiment. We find that the global response of pathways to stress is multi-dimensional and does not obviously cluster according to stress, time or tissue. Overall, early response typically involves RNA, hormone synthesis and signaling; late response typically involves metabolism of amino acids and secondary metabolites. By linking specific primary and secondary response pathways we outline specific routes of response progression. In *Arabidopsis*, the secondary metabolite glucosinolates plays a central role in plant survival during progression of abiotic stress. This analysis effectively presents a global response of pathways during progression of abiotic stress with respect to time and tissue.

OPTIMIZATION OF THE PIGGYBAC TRANSPOSON SYSTEM FOR CULTURED DROSOPHILA CELLS

L.A. Yarinich[1, 2]*, M.O. Lebedev[1, 2], A.V. Pindyurin[1, 2]
[1] *Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *Corresponding author: l.yarinich@mcb.nsc.ru*

Key words: *gene expression, chromatin position effect, DNA barcode, multiplex analysis, Drosophila cultured cells*

Motivation and Aim: The TRIP (Thousands of Reporters Integrated in Parallel) approach allows studying the influence of the local chromatin context on the gene activity simultaneously at thousands of genomic loci in cultured mouse cells. The approach is based on piggyBac-mediated transposition of the DNA-barcoded reporter constructs into the genome with the subsequent identification of their insertion sites and analysis of their transcriptional activity using high-throughput sequencing. The aim of this study is the adaptation of the TRIP approach to cultured *Drosophila* cells, which are a convenient model system for understanding the regulation of gene expression.
Methods and Algorithms: *Drosophila* Kc167 and S2 cultured cells of embryonic origin were used. Both electroporation and chemical methods were used to transfect cells. The level of transcription of the gene encoding piggyBac transposase was assessed by RT-qPCR. Efficiency of transposition of the reporter constructs into the genome was evaluated by qPCR as well as FACS. The insertion sites of the reporter constructs were identified by inverse PCR.
Results: We generated plasmid constructs encoding the piggyBac transposase under the control of four different promoters (constitutively active and inducible ones) and optimized the conditions for the transient activity of the transposase to avoid the 'rehopping' of the reporter constructs after their integration into the genome. For this purpose the following two approaches were used: [i] the chimeric version of the piggyBac transposase (PB-L3-ERT2), whose activity is regulated by tamoxifen, and [ii] the constructs encoding the transposase and the reporter gene as one transcriptional unit allowing to turn off the piggyBac transposase expression immediately after the transposition of the reporter construct from the plasmid into the genome. The system with the transposase gene and the transposon present in two separate plasmids was used to manipulate the ratio of the transposon to the transposase. We observed that 5:1 ratio of the transposon to the transposase allows to obtain the maximum copy number of the transposon insertions in the genome of *Drosophila* cells. The protocols for transfection of *Drosophila* cultured cells by electroporation as well as chemical methods were optimized to get as much as possible insertions of reporter constructs.
Conclusion: We found that the level of transcription of the gene encoding piggyBac transposase and the frequency of reporter construct integration into the genome depend not only on the selected promoter for the transposase expression, but also on the cell line, the method of transfection and the ratio of the transposon to the transposase used.
Availability: The developed plasmid constructs and protocols can be used for effective transgenesis of *Drosophila* cultured cells.
Acknowledgements: We thank Dr. S.V. Kulemzin (IMCB SB RAS) for FACS analysis of cells. This study was supported by the grant from the Russian Science Foundation no. 16-14-10288.

GTRD – GENE TRANSCRIPTION REGULATION DATABASE

I.S. Yevshin[1, 2], R.N. Sharipov[1, 2, 3], Yu.V. Kondrakhin[1, 2], F.A. Kolpakov[1, 2]*
[1] *Institute of Systems Biology Ltd., Novosibirsk, Russia*
[2] *Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
[3] *Novosibirsk National Research State University, Novosibirsk, Russia*
* *Corresponding author: fedor@biouml.org*

Key words: *ChIP-seq, BioUML, gene expression, transcription factors, workflow*

Motivation and Aim: A lot of processes in living cells are controlled by regulation of gene expression at the level of transcription. This level of regulation is mediated by chromatin modifications and cooperative action of transcription factors. ChIP-seq technology allows identification of DNA regions bound by individual transcription factors in the whole genome. The main goal of this study was to produce an unambiguous database of regulatory elements in the whole human and mouse genomes using data from ChIP-seq experiments.

Results: We collected raw ChIP-seq data known from literature, GEO, SRA and ENCODE databases for human and mouse. All the data were processed through unified workflows using BioUML platform: sequenced reads were aligned to the reference genome using Bowtie and approximate binding regions of TFs were identified using different peak finders. Based on these regions we constructed the set of position weight matrices for each TF. Position weight matrices were used for theoretical prediction of transcription factor DNA binding sites . The procedure for learning PWMs from several ChIP-seq experiments automatically selects the best model based on the analysis of receiver operating characteristic. We use hierarchical classification of TFs by their DNA-binding domain to make generalized PWMs for TF classes in the case when DNA binding sites of similar TFs cannot be distinguished by their sequences.

Using BioUML platform we also have developed web interface for access to GTRD database that provides possibility to browse and search corresponding information. Built-in genome browser provides powerful visualization of ChIP-seq data.

Conclusion: The main advantages of GTRD are following:

- it contains the most comprehensive collection of ChIP-seq data for human and mouse;
- it contains not only meta-data about ChIP-seq experiments but also raw and uniformly processed data;
- all data were uniformly processed using the same workflow. It required several months of constant work of powerful server;
- obtained PWM matrices were systematically estimated and compared;
- all data were grouped around classification of TF from TFClass database.

Availability: http://gtrd.biouml.org

References:

1. ENCODE project - https://www.encodeproject.org/
2. SRA - Sequence Read Archive - http://www.ncbi.nlm.nih.gov/sra
3. GEO - Gene Expression Omnibus - http://www.ncbi.nlm.nih.gov/geo/
4. BioUML platform - http://wiki.biouml.org
5. E. Wingender et al. (2013) TFClass: An expandable hierarchical classification of human transcription factors. *Nucleic Acids Res*., 41, D165-D170

ASSESSMENT OF TRANSLATION EFFICENCY FROM RIBOSOME PROFILING AND MRNA-SEQ DATA

I.S. Yevshin[1, 2], R.N. Sharipov[1, 2], O.A. Volkova[3]*
[1] *Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia*
[2] *Institute of Systems Biology, Ltd, Novosibirsk, Russia*
[3] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
* *Corresponding author: ov@bionet.nsc.ru*

Key words: *mammalian mRNAs, translation efficiency, ribosome profiling, Ribo-Seq, mRNA-Seq, lncRNAs*

Motivation and Aim: Ribosome profiling or Ribo-Seq technology allows to estimate the level of translation of distinct mRNAs. For the last years many studies grounded on this technology have been performed on different cell types and in various physiological conditions. To undertake a wide-scale study of regulation of translation efficiency (TE) in cell, it is necessary to control of input Ribo-Seq data and adequate assessment of TE. The goal of this work was to assess the quality of raw Ribo-Seq data available in public databases and identify the optimal parameters of their processing and TE assessment.
Methods and Algorithms: Datasets containing both, Ribo-Seq and mRNA-Seq data, were selected from a RiboSeqDB database developed by the authors. Raw data were processed using special workflows developed for a BioUML platform. To control the quality of sequenced reads, calculated TE values were compared with respective TE values but obtained using mass-spectrometry and mRNA-Seq.
Results: (1) Several methods for TE assessment were developed and their accuracy was checked. Methods for the confidence interval and TE ratio error estimation were developed. (2) The optimal parameters for adapter trimming and Ribo-Seq read alignment were adjusted. (3) The influence of PCR duplicates on TE assessment accuracy was checked: they must be taken into account like other reads. (4) The influence of Ribo-Seq reads corresponding to the mRNA regions outside of the reading frame on TE assessment accuracy was checked: they also must be taken into account like other reads.
Availability: in the frames of the BioUML platform (http://www.biouml.org)
Acknowledgements: This work was supported by the Russian Foundation for Basic Research (№ 14-04-01284).

STRUCTURAL BIOINFORMATICS OF FPG GLYCOSYLASE: SEARCH FOR SUBSTRATE SPECIFICITY IN THE SEQUENCE SPACE

A.V. Yudkina[1, 2]
[1] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: yudkinaanya@gmail.com*

Key words: *DNA damage, DNA repair, protein design, formamidopyrimidine–DNA glycosylase*

Motivation and Aim: DNA glycosylases are enzymes that remove modified bases from DNA and are a key element of base excision DNA repair. Due to their ability to cleave DNA at defined modified nucleotides, these enzymes are now finding use as tools in molecular biology, biotechnology and bioanalytics. Thus, design of DNA glycosylases with altered substrate specificities is of great primary fundamental and practical interest. In particular, bacterial formamidopyrimidine–DNA glycosylase (Fpg) removes 8-oxoguanine (oxoG) and formamidoptrimidines from DNA, and has a base-binding site built of a tetrapeptide loop. In this work we aim to design DNA glycosylases that will be able to recognize other nucleobases (e. g., 6-thioguanine, thioG, an important antiproliferative drug that acts through incorporation into DNA), on the protein scaffold of Fpg.
Methods and Algorithms: We have used molecular dynamics modeling to analyze a sample of 102 randomly mutagenized Fpg base-binding loops built into otherwise intact Fpg scaffold to compare several estimators of "closeness" of the mutant sequences to wild-type Fpg. Additionally, using quantum mechanical calculations, we have analyzed the mechanism of *N*-glycosidic bond cleavage in solution for several purine nucleotides structurally related to G and oxoG.
Results: In the Fpg molecule, the site of substrate recognition and the catalytic center are separated in space. The active site contains invariant Pro1 and Glu2 residues and some highly conserved residues that form contacts with phosphates in DNA. In the substrate recognition site, a loop of four amino acid residues binds the O^6 atom of 8-oxoguanine, with the static structure not particularly informative of the enzyme's base specificity. We constructed 102 structures of Fpg/8-oxoguanine–DNA complexes where this tetrapeptide was randomly changed and analyzed them by molecular dynamics. On the basis of the local geometry ranking, the function of model quality was built, which determines the multidimensional distance between the conformation of the active site of mutant Fpg and the catalytically competent conformation of the active site of wild-type Fpg. This function will be further used to predict the ability of random-loop Fpg structures to recognize other purines, including thioG. Several algorithms of walk in the sequence space to minimize the number of computationally expensive models have been compared.
Conclusion: We propose a scheme for rational design of DNA glycosylases with altered substrate specificity based on the Fpg scaffold. This approach will be used to construct enzymes able to cleave DNA at modified purine lesions, in particular, thioG.
This work was supported by Russian Science Foundation (14-24-00093).

IN SILICO DESIGN OF APTAMERS CONTAINING G-QUADRUPLEXES

A.O. Zalevsky[1,2]*, A.O. Demkiv[2], A.V. Golovin[1, 2]
[1] *Apto-Pharm LLC, Moscow, Russia*
[2] *Faculty of bioengineering and bioinformatics, Lomonosov Moscow State University, Moscow, Russia*
** Corresponding author: aozalevsky@fbb.msu.ru*

Key words: *aptamers, molecular dynamics, rational design, g-quadruplex*

Motivation and Aim: Aptamers are promising molecules for therapy and diagnostics. DNA-made aptamers combine high specificity of monoclonal antibodies with low immune response and ease of production [1]. However despite of all possible advantages aptamers are still not as widespread as one could thought. Major limitation related to complicated aptamers tuning process. Knowledge of aptamer's structure is crucial for succesfull optimiziations but due to nature of SELEX pipeline (method for obtaining aptamers), it is usually unknown.
Here we try to use computational approaches to overcome typical SELEX limitations and design G-quadruplex containing aptamers. G-quadruplexes are known for their increased stability compared to duplex analogs [2].
Methods and Algorithms: Our approach combines several methods for molecular modeling such as molecular docking (AutoDock Vina) and molecular dynamics (Gromacs). All tools are integrated in one pipeline written in Python with critical parts being optimized (Cython) and parallelized (MPI). Experimental validation of G-quadruplex presence is done with circular dichroism spectroscopy and binding to target is tested with capillar electrophoresis.
Results: For testing purposes, we applied our approach to thrombin. After several rounds of ranking and filtration, several candidates with relatively stable structures and preserved designed protein-dna contacts were selected and tested experimentally.
Conclusion: Here we present new computational approach for in silico design of aptamers containing G-quadruplexes based solely on target's 3D structure.
Availability:
Acknowledgements: This study was supported by the Russian Foundation for Basic Research project no. 16-34-01362. Computations were done at the supercomputer systems "Lomonosov" and "Lomonosov-2" of Lomonosov Moscow State University, Moscow, Russia.
References:
1. H. Sun et al. (2014), Oligonucleotide Aptamers: New Tools for Targeted Cancer Therapy, *Molecular Therapy Nucleic Acids,* **3:** e182.
2. A.Lane et al. (2008) Stability and kinetics of G-quadruplex structures, *JNucleic Acid Research,* **36:** 5482–5515.

DISTINCT TYPES OF EIN3-DNA INTERACTIONS IN VARIOUS FUNCTIONAL REGIONS OF *A. THALIANA* L. GENOME

E.V. Zemlyanskaya[1, 2]*, D.Yu. Oshchepkov[1], V.G. Levitsky[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: ezemlyanskaya@bionet.nsc.ru*

Key words: *transcription factor, ethylene insensitive 3 (EIN3), ChIP-seq data analysis*

Motivation and Aim: Plant hormone ethylene regulates a wide range of physiological processes during plant development and coordinates different stress responses [1]. Transcription factor (TF) EIN3 is a master regulator of ethylene signaling pathway [2]. The emergence of chromatin immunoprecipitation followed by genome-wide sequencing (ChIP-Seq) stimulated the description and characterization of motifs for many transcription factors. Availability of the whole-genome maps of chromatin types and TF binding sites derived from ChIP-seq data allows finer characterization of TF-DNA interactions. We use these data to investigate EIN3-DNA interactions to reveal details of EIN3 action.
Methods and Algorithms: We analyzed ChIP-seq data for EIN3 binding in *Arabidopsis thaliana* [2]. DNA motifs, specific for certain types of *A. thaliana* chromatin [3], were identified with Homer tool [4] for *de novo* motif search.
Results: We found the heterogeneity of peaks distribution in gene promoters at various distances from transcription start sites (TSS), within bodies of genes, in introns and intergenic spacers. EIN3 peaks were enriched within 300 bp upstream TSS and in intergenic spacers. To finer characterize EIN3-DNA interactions, we distinguished EIN3 peaks located in chromatin domains of states 2 and 4 [3]. In [3] these domains were referred to as proximal and distant 5'regulatory regions of genes. The nucleotide context of EIN3 binding sites was more conserved in distant chromatin domains than in proximal ones. *De novo* motif search in close proximity to EIN3 motifs revealed several distinct motifs in proximal and distant chromatin domains. We believe that they represent binding sites for EIN3 partner TFs. We found essential difference between the sets of partner TFs corresponding to different types of chromatin domains.
Conclusion: EIN3 regulation of gene expression is mediated by at least "proximal" and "distant" types of EIN3-DNA interactions.
Acknowledgements: The work was supported by RFBR 15-34-20870.
References:
1. F.B. Abeles et al. (1992) Ethylene in plant biology. San Diego: Acad. Press.
2. K.N. Chang et al. (2013) Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in Arabidopsis. *Elife*, 2:e00675.
3. J. Sequeira-Mendes et al. (2014) The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell,* **26**: 2351–2366.
4. S. Heinz et al. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*, **38**:576-589.

ELEMENTAL METABOLOMICS – LINKING ENVIRONMENTAL, FOOD, NUTRITION AND HEALTH SCIENCES

P. Zhang[1], I. Giannenas[2], C.A. Georgiou[3], V. Brusic[1, 4]*
[1] *Menzies Health Institute Queensland, Griffith University, Australia*
[2] *Aristotle University of Thessaloniki, Thessaloníki, Greece*
[3] *Department of Food Science and Nutrition, Agricultural University of Athens, Greece*
[4] *School of Medicine and Bioinformatics Center, Nazarbayev University, Kazakhstan*
* *Corresponding author: vladimir.brusic@nu.edu.kz; v.brusic@griffith.edu.au*

Key words: *elemental signatures, metabolome, metallomics, mineralomics, ionomics*

Motivation and Aim: The advancements in instrumentation, measurement methods, and standards in mass spectrometry enabled us to precisely measure the quantities of more than 70 trace and ultra-trace elements (measurable elements) in biological materials. Very recently, the precision of measurements and lower cost enabled us to measure elemental matrices on a large scale. Measurable elements are bioavailable from the environment, including air, water, and soil from where they are absorbed and accumulated by plants. Measurable elements enter animal and human bodies through environmental exposure or through food chain. Elemental metabolome profiles can be used for identification of environmental exposure, food authentication, effects of nutritional intervention, food quality assessment, and health monitoring.
Methods: Two studies were performed. Samples of three sets of chicken eggs (conventionally grown, organic, and free-range) were measured for elemental profiles of 12 trace elements (As, Cd, Co, Cr, Cu, Mn, Mo, Ni, Se, Tl, V, and Zn). The comparison of elemental profiles was used to classify eggs by the production system. These data were compared with the eggs of another four species (turkey, goose, duck, and pigeon) that were given exactly the same diet as was given to the courtyard chickens.
Results: Elemental profiles enabled highly accurate classification of eggs by production method. Using sensitivity and specificity measures we assessed classification accuracy to be SE=0.958, SP=0.99 (conventional); SE=0.979, SP=0.979 (organic); and SE=1.0, SP=1.0 (free-range). The comparison with elemental profiles of other species has shown that chicken egg elemental profiles are more similar to other chicken eggs elemental profiles than to those representing eggs of other species, irrespective of the production system applied.
Conclusion: Elemental metabolomics provides a new method suitable for biomonitoring. It can be used for classification of samples by their origin (geographic, production system, genetic origin), nutritional status, or health status.

SPEED READING AT THE MOLECULAR SCALE: HOW ENZYMES FIND TYPOS IN A DNA TEXT

D.O. Zharkov*
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
** Corresponding author: dzharkov@niboch.nsc.ru*

Key words: *DNA damage, DNA repair, substrate specificity, target search*

Motivation and Aim: Many enzymes involved in genome surveillance, maintenance and editing face a task of finding rare targets within a huge excess of non-specific DNA. Such targets may represent specific DNA sequences, either exactly defined or consensual, or modified nucleotides, or non-canonical DNA structures. There is always a trade-off between the rate of search and the rate of errors (not recognizing a target or mistakenly accepting a non-target), with the errors possibly bearing a huge cost for the cell (1–3). The mechanisms by which DNA-dependent enzymes achieve an optimal balance between speed and fidelity are largely obscure.
Methods and Algorithms: We have used a combination of methods including molecular modeling, fast enzyme kinetics, DNA and protein melting, and solution NMR to address the mechanisms underlying search and recognition of damaged bases by several DNA repair enzymes: DNA glycosylases and abasic site endonucleases.
Results: DNA repair enzymes search for their targets by fast one-dimensional scanning along the DNA contour intermittent with rare events of transfer to another site remote in 1D but close in 3D relative to the current location. This search is facilitated by extended N- or C-terminal domains present in some repair enzymes and apparently evolved from rapidly diverged short sequence repeats. The one-dimensional scanning function may sometimes be exploited for other purposes, such as uracil–DNA glycosylases adopted as DNA polymerase processivity factors by some viruses. The enzymes first probe DNA indirectly, relying on DNA deformability for the primary damage assessment. Once a "soft spot" is found, the sampled base is everted from the double helix stack, with several energy barriers along the eversion pathway allowing the enzyme to reject normal bases. Recognition within the active site contributes little to the specificity, with the rate of the chemical step mostly dependent on the interactions with DNA moieties other than the sampled base.
Conclusion: DNA repair enzymes find optimum between speed and fidelity through multigated recognition of the lesion, which allows for kinetic competition between diffusion and productive substrate encounters.
This work was supported by Russian Science Foundation (14-24-00093)
References:
1. D.O. Zharkov, A.P. Grollman (2005) The DNA trackwalkers: Principles of lesion search and recognition by DNA glycosylases, *Mutat. Res.,* 577: 24-54.
2. D.O. Zharkov, G.V.Mechetin, G.A.Nevinsky (2010) Uracil-DNA glycosylase: Structural, thermodynamic and kinetic aspects of lesion search and recognition. *Mutat. Res.*, 685: 11-20.
3. G.V. Mechetin, D.O.Zharkov (2014) Mechanisms of diffusional search for specific targets by DNA-dependent proteins. *Biochemistry (Mosc.)*, 79: 496-505.

PHYLOGENETIC RECONSTRUCTION WITHIN MYCOBACTERIUM TUBERCULOSIS BEIJING GENOTYPE IN NORTHEASTERN RUSSIA

S. Zhdanova[1]*, O. Ogarkov[1, 2, 3], G. Alexeeva[4], M. Vinikurova[4], E. Savilov[1, 2], V. Sinkov[1]
[1] *Scientific Centre of the Family Health and Human Reproduction Problems, Irkutsk, Russia*
[2] *Irkutsk State Medical Academy of Continuing Education, Irkutsk, Russia*
[3] *Irkutsk State University, Irkutsk, Russia*
[4] *Research Practice Center for Phthisiatry, Yakutsk, Russia*
** Corresponding author: svetnii@mail.ru*

Key words: *M. tuberculosis, MIRU-VNTR, epidemic subtypes, Beijing*

Motivation and objectives: Mycobacterium tuberculosis Beijing strains cause wide concern because of their speedy emergence and increasing prevalence in world. The phylogenetic reconstruction of Beijing strains evolution within territory with limited migration could provide us a model to understand the dynamics of some epidemic subtypes expansion.
Methods and Algorithms: We used 153 stains to compare the population structures of the *M. tuberculosis* between different birth-year cohorts of patients in Sakha (Yakutia). MIRU-VNTR genotyping, RD105/RD207 subtyping [1], classification of patterns on clonal complexes Merker M. et al. [2] and phylogenetic modeling were implemented. Profiles of 153 strains of the 24th MIRU-VNTR loci using MS Excel 2007 have been converted to a binary format and used to construct NJ tree, programs Ugene [3] and FigTree [4].
Results: The structure of the population of Mycobacterium tuberculosis and shift that occurred during the previous fifty years have been described and analyzed. The results revealed that the spread of modern subtype of Beijing genotype that have high transmissibility and multiply drug resistance currently in Russia, is significantly higher in young (post-1990 birth) than in the older generation (born before 1959) (χ^2=8.3, p<0.01). The hypothesis of later emergence of epidemic subtypes genotype Beijijng in the Sakha (Yakutia) than in other regions of Russia (around fifty years ago) is substantiated.
Acknowledgements: This work has been supported by grant from RFBR #15-04-00632 A
References:
1. M. Reed et al. (2009). Major Mycobacterium tuberculosis Lineages Associate with Patient Country of Origin, J. Clin. Microbiol, 47(4): 1119-1128.
2. M. Merker et al. (2015). Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage, Nat Genet., 47(3): 242-249.
3. K.Okonechnikov et al. (2012). Unipro UGENE: a unified bioinformatics toolkit, Bioinformatics, 28: 1166-1167.
4. V.I. Morariu et al (2008). Automatic online tuning for fast Gaussian summation. Advances in Neural Information Processing Systems. V. 21. P.1113-1120.

CHEMORESISTANCE OF LUNG ADENOCARCINOMA IS REGULATED BY TUDOR STAPHYLOCOCCAL NUCLEASE

B. Zhivotovsky[1, 2]
[1] *Lomonosov Moscow State University, Moscow, Russia*
[2] *Karolinska Institutet, Stockholm, Sweden*
** Corresponding author: Boris.Zhivotovsky@ki.se*

Key words: *cancer, gene expression, gene networks, pathway analysis*

Lung cancer is the main cause of all cancer-related deaths in the world, with lung adenocarcinoma (ADC) being the most common subtype of this fatal disease. Lung ADC is often diagnosed at advanced stages involving disseminated metastatic tumors. This is particularly important for the successful development of new cancer therapy approaches. The high resistance of lung ADC to conventional radio- and chemotherapies represents a major challenge to treatment effectiveness.

Earlier we found that Tudor staphylococcal nuclease (SND1 or TSN) is overexpressed in ADC lines and tissues, and is important for maintaining of ADC chemoresistance. Downregulation of TSN by RNAi in ADC cells led to strong potentiation of cell death in response to cisplatin.

In order to identify potential molecular targets involved in ADC sensitization to cisplatin, the global gene expression analysis was performed. Widespread transcriptional changes were observed upon TSN knockdown: 391 unique genes demonstrated a greater than twofold average change in expression, with 234 transcripts under- and 157 transcripts overexpressed compared to scrambled transfected control samples. Using the Ingenuity Pathways Analysis (IPA) program and gene ontology category enrichment analyses, we selected several major networks containing genes that were closely associated with autophagy and apoptotic cell death as well as survival, DNA damage response and Ca^{2+} signaling. The expression of shortlisted genes was further analyzed by q-RT-PCR, confirming microarray data. ON the top of the list was S100A11. Silencing of TSN was accompanied by a significant decrease in S100A11 expression at both mRNA and protein level. Downregulation of S100A11 by RNAi resulted in enhanced sensitivity of NSCLC cells to cisplatin, oxaliplatin and 5-fluouracil.

S100A11 interactions were analyzed using Interactive pathway analysis of complex'omics data and S100A11-related pathways involved in apoptosis and cell resistance to cytotoxic treatment were selected for further analysis. We found that in cell cytoplasm S100A11 interacts with Annexin A1 and Annexin A2 and inhibit phospholipases A_2 (PLA_2), a superfamily of enzymes involved in arachidonic acid (AA) release. A PLA_2 inhibitor or silencing with siRNA strongly abrogated chemosensitization upon silencing of S100A11 suggesting that PLA_2 inhibition by S100A11 governs the chemoresistance of ADCs. Thus, we present the novel TSN-S100A11-PLA_2 axis regulating superoxide-dependent apoptosis, triggered by platinum-based chemotherapeutic agents in ADCs that may be targeted by innovative cancer therapies.

This study was supported by grant from the Russian Science Foundation (14-25-00056).

COMPUTER SIMULATION OF TRICHOME PATTERNING ON GROWING WHEAT LEAF TAKING INTO ACCOUNT THE BIOMECHANICS OF CELLS

U.S. Zubairova[1]*, S.V. Nikolaev[1], A.V. Penenko[2], N.L. Podkolodnyy[1], S.K. Golushko[3], D.A. Afonnikov[1], N.A. Kolchanov[1]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia*

[3] *Design and Technology Institute of Digital Techniques SB RAS, Novosibirsk, Russia*

** Corresponding author: ulyanochka@bionet.nsc.ru*

Key words: *wheat leaf, trichome patterning, cell mechanics*, s*ymplastic growth, L-systems, computer simulation*

Motivation and Aim: Trichome patterning in wheat leaves serves as a model system to understand mechanisms of pattern formation on growing plant tissue. The question about the coordination of cell growth in plant tissue remains open up to date. In vertex-based plant tissue models an autonomous cell growth is usually assumed and generalized potentials are used for describing the changes in the tissue geometry. Meanwhile, the biomechanics is considered as an important factor of the morphogenesis of tissues and even organs. We consider this issue in the investigation of changes in the cellular structure of wheat leaf epidermis during growth. The tissue structure change occurs due to activity of growth zones containing regions of dividing and differentiating cells. The epidermis of wheat leaf is established by parallel files of cells originating from the leaf base, where specialized cells, trichomes, are formed in separate files. The aim of our work was to develop a mechanical cell-based model for growth of linear leaf blade and explore the mechanism of trichomes pattern formation.

Methods and Algorithms: A mathematical model based on the extension of L-systems approach and its implementation for computational simulation was proposed. We assumed a unidirectional growing cell ensemble starting from a meristem-like layer of generative cells and then generating parallel cell rows from every cell of the initial layer. We considered the growth zone of the leaf included division and elongation zones; in addition the division zone included a zone of asymmetric divisions where trichomes were formed. We applied a modification of Ortega's augmented growth equation [1] to the description of plant cell growth mechanics.

Results of computer simulation demonstrate that the proposed model can describe (i) the experimental cells' lengths distribution along the wheat leaf (data from [2]), and (ii) the experimental trichome spacing pattern in separate cell files. *Acknowledgements:* This work was supported by the RSF according to the research project № 14-14-00734.

References:

1. Ortega, J.K. (2010) Plant cell growth in tissue, *Plant Physiology* **154(3)**:244-1253.
2. Beemster, G. T., Masle, J., Williamson, R. E., & Farquhar, G. D. (1996). Effects of soil resistance to root penetration on leaf expansion in wheat (Triticum aestivum L.): kinematic analysis of leaf elongation, *Journal of Experimental Botany,* **47(11):** 1663-1678.

AN IMAGEJ PLUGIN FOR DETECTION OF WHEAT LEAF EPIDERMIS CELLULAR STRUCTURE FROM CONFOCAL LASER SCANNING MICROSCOPY

U.S. Zubairova[1]*, P.Yu. Verman[2] and A.V. Doroshkov[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: ulyanochka@bionet.nsc.ru*

Key words: *wheat leaf, epidermal patterning, laser scanning microscopy, cell segmentation, ImageJ plugin*

Motivation and Aim: The epidermis of wheat leaf is a complex tissue consisting of different cell types forming a certain cell pattern from parallel cell rows. Microscopic images are widely used as an important source of information on the morphometric characteristics of the cells and the statistical characteristics of the cellular structure. 3D confocal images allow to determine characteristics of the cell structure of the leaf epidermis. However, to obtain large amounts of statistical data methods of high throughput computer based image segmentation are need. Our aim was to develop a plugin for detection of structural properties of leaf epidermis from 3D-images obtained from confocal laser scanning microscopy. These characteristics of the cell structure and patterns further will act as a basis for the development and verification of spatial models of plant tissues formation mechanisms accounting for structural features of monocot leaves.
Methods and Algorithms: Methods of fluorescent staining and laser scanning microscopy were used to obtain 3D images for visualizing the cell walls and cell nucleus in two different color channels. Each image contains information about the structure of a large fragment of wheat leaf epidermis and is composed of several frames which represent the series of a single lsm scans. The plugin "LSM_Worker" allows one (i) to merge frames into a single image, (ii) to improve the image quality by removing offsets and noise, (iii) to segment the cells and nucleus for each cannel, respectively and (iv) to compare the information from both channels and generate statistical characteristics of cell and nucleus volumes.
Results and Conclusion: The plugin was used for obtaining statistical characteristics of the cellular pattern of leaf epidermis of several varieties of bread wheat characterized by different morphological features of epidermal cells. Obtained data provide a material for formulation hypotheses and development of models of variety leaf growth mechanisms.
Availability: The plugin is available from the authors upon request.
Acknowledgements: The reported study was funded by RFBR according to the research project № 16-34-00968.

ALTORFEV: A NOVEL TOOL FOR PREDCITION OF ALTERNATIVE ORFS BASED ON THE LINEAR SCANNING MODEL

B.S. Zuraev[1]*, A.V. Kochetov[1, 2], A.I. Klimenko, S.A. Lashin[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: bulatzuraev@gmail.com*

Key words: *Cytoscape plugin; ortholog; paralog; metabolic pathway; gene regulatory network; evolution; phylostratigraphy; evolution*

Motivation and Aim: The ribo-seq and proteomics techniques have revealed a large number of alternative ORFs (altORFs) within eukaryotic mRNAs. Some bioinformatics resources were developed to explore the available ribo-seq data to locate altORFs within mRNAs of interest (e.g., Ribotools, RiboGalaxy, GWIPS-viz). Indeed, knowledge on the full set of polypeptides encoded by a eukaryotic gene under study is essential for detailed investigation of its functions. However, published ribo-seq data are still very limited and conventional nucleotide sequence databanks do not provide information on the altORFs. In addition, the individual genetic variants may cause changes in mRNA coding potential: if a nucleotide sequence of mRNA under study is non-identical to the available ribo-seq-checked reference sequence, the positions of altORF(s) and their relative translation rates may differ. Thus, development of new tools for altORFs prediction remains quite actual. However, an accurate prediction of altORFs is very complicated because of a large number of various parameters influencing their recognition and translation efficiency. *Methods and Algorithms:* The altORFev is based on the linear scanning model of translation [1]. It also considers the leaky scanning and reinitiation mechanisms. In brief, 40S ribosomal subunits bind to 5'-end of mRNA and move linearly along mRNA until start AUG codon is found. The probability of AUG recognition depends on its nucleotide context: start codon in the optimal context is recognized by the majority of 40S ribosomal subunits. Thus, if AUG codon is located in the optimal context and its ORF is larger than 30 codons, this ORF is defined as "terminal" since the majority of incoming 40S ribosomal subunits can't move beyond it. If AUG codon is located in a suboptimal context, some 40S ribosomal subunits will recognize it and initiate translation, whereas others skip it and may initiate translation downstream (leaky scanning). Finally, if AUG is located in the optimal context but the ORF size is small (lesser than 30 codons), the reinitiation is possible: in this case, some 40S ribosomal subunits after termination of translation of small ORF remain connected to mRNA and may continue movement in 3'-direction. During scanning they restore their initiation competence by acquiring the lacked eIFs and met-tRNAi and may initiate translation further downstream. *Results:* We have implemented two versions of the altORFev: (1) web application (Java 1.8, Vaadin); (2) desktop application (Java 1.8, Swing). *Conclusion:* The altORFev may be used to get additional information on eukaryotic genes taking into consideration alternative coding abilities of their mRNAs.

Availability: web-version: http://www-bionet.sscc.ru:7780/AUGWeb/, desktop version: upon the requests to the authors. *Acknowledgements:* The study is supported by the RSF 14-24-00123 grant.

References:

1. Kozak M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, 361:13–37.

Author index

C

D

E

F

G

L

M

N

O

P

Q

R

S

ХИМЭКСПЕРТ

Агентство Химэксперт – официальный дистрибьютор компании Thermo Fisher Scientific (Life Sciences), являющейся основным мировым поставщиком приборов, реактивов и расходных материалов для молекулярной биологии. В компанию входят такие известные бренды, как: Applied Biosystems, Invitrogen, Ion Torrent, Ambion, Gibco, Molecular Probe, Novex, TaqMan, Pierce, Fermentas, Nunc, Finnpipette, Ленпипет и другие.

Мы предлагаем оборудование для анализа ДНК и РНК, фундаментальных протеомных и цитологических исследований, фармацевтики и биотехнологий, прикладного тестирования, включая идентификацию личности и установление родства в криминалистике и судебно-медицинской экспертизе, тестирование пищевых продуктов и многое другое.

Наши клиенты выбирают Химэксперт потому что:

– Мы предоставляем комплексные решения для оснащения молекулярно-генетических и медико-генетических лабораторий: от дозаторов до секвенаторов.

– Мы осуществляем полную техническую и методическую поддержку наших клиентов. Всегда готовы помочь заказчикам с подбором реагентов под поставленные задачи. Регулярно проводим профильные обучающие школы и семинары.

– Мы всегда осуществляем своевременные поставки заказов.

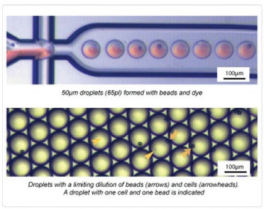
Компания Химэксперт существует 15 лет и давно зарекомендовала себя, как надежный поставщик оборудования и реактивов для широкого спектра исследований.

ООО «Агентство Химэксперт»
125009, г. Москва, Страстной б-р, д. 4, оф. 101
Тел: +7 (495) 629 28 69, 650 36 66
info@khimexpert.ru, www.khimexpert.ru

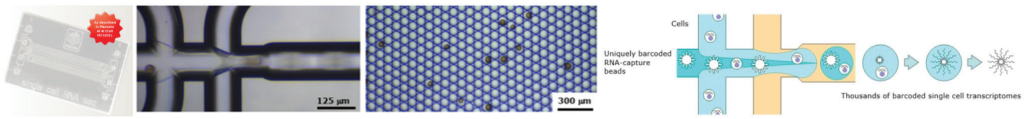
Использование микрофлюидики Dolomite для секвенирования транскриптомов отдельных клеток

Система для инкапсуляции клеток или нуклеиновых кислот в капли µEncapsulator 1

Отличная пробоподготовка для изучения экспрессии генов, ПЦР, сортировки и др.; инертные, биосовместимые материалы; возможность поддержания жизнеспособности клеток; пропускная способность: 300 000 клеток в 3 млн капель за 15 минут; инкапсуляция 100 мкл образца и реагента, или двух последовательных образцов по 100 мкл.



Система для создания библиотек единичных клеток для последующего секвенирования



Транскриптомика одиночных клеток; высокая точность, воспроизводимость, надежность, инертность материала чипа; скорость инкапсуляции, капель/сек – 4000 - выше чем у чипов из PDMS1.

Микрофлюидика Dolomite — технология, позволяющая работать с очень малыми объемами жидкостей, газов, с кристаллическими и полимерными частицами, клетками животного, растительного и бактериального происхождения, пузырьками и каплями с возможностью наблюдать за ними, манипулировать ими и контролировать процессы, протекающие с ними. Это дает возможность проводить «традиционные» исследования в миниатюрном формате, а также проводить исследования, которые ранее были невозможны.

Особенности и возможности микрофлюидики Dolomite:
работа с микрообъектами (капли, клетки, частицы, пузырьки); работа с микро- и нанообъемами (диаметр канала от 10 нм); высокая воспроизводимость: точность дозирования — порядка пиколитра; точный контроль параметров процесса: температуры, скорости потоков, давления, смешивания; большая библиотека «стандартных» чипов; чипы произвольной конфигурации и геометрии: многослойные и составные чипы с разными свойствами поверхности каналов, интеграция на одном чипе различных стадий процессов для ускорения и автоматизации методик исследований; интеграция с приборами, детекторами, системами пробоподготовки и сенсорами (хроматографами, масс-спектрометрами, лазерами, спектрофотометрами, микроскопами и т.д.); автоматизация процессов: удобство, высокий выход, воспроизводимость, точность; объединение разных стадий методик в одном приборе; уменьшение размеров приборов; появление новых методов и приборов.

Технология микрофлюидики Dolomite находит применения в таких областях как:
химический синтез, аналитическая химия, физико-химические исследования; разработка лекарственных препаратов, определение эффективности и цитотоксичности; биология, диагностика и медицина; экология, производство, приборостроение.

Биология, диагностика и медицина:
качественный и количественный анализ фрагментов НК на чипе капиллярного электрофореза; чипы для секвенирования НК; цифровая капельная ПЦР для количественной ПЦР-диагностики с высокой точностью; анализы крови (биохимические, ИФА, на глюкозу и т.д.); изоляция ДНК из цельной крови; наблюдение за иммобилизованными эмбрионами и клетками.

Компания Eppendorf занимается разработкой, производством и продажей продукции премиум-класса и услуг для лабораторий по всему миру и является одним из лидеров на рынке высокотехнологичного оборудования. Продукция Eppendorf используется в лабораториях самых различных профилей: академических, отраслевых научно-исследовательских, клинико-диагностических, экологических, криминалистических, а также в фармацевтической, биотехнологической, химической и пищевой промышленности и в лабораториях на промышленных предприятиях, которым необходим контроль качества и анализ производственного процесса.
Головной офис компании находится в Гамбурге, Германия.

Продукция Eppendorf представлена в трех направлениях:

LIQUID HANDLING:
- Механические и электронные дозаторы;
- Наконечники для дозаторов;
- Механические и электронные диспенсеры;
- Бутылочные дозаторы, цифровые бюретки;
- Станции автоматического дозирования.

SAMPLE HANDLING:
- Миксеры, термомиксеры и термостаты;
- Пробирки и планшеты;
- Центрифуги и роторы;
- Амплификаторы и расходные материалы для ПЦР;
- Низкотемпературные морозильные камеры.

CELL HANDLING:
- Фотометры, спектрофотометры;
- Микроманипуляторы, микроинъекторы и вспомогательное оборудование;
- Расходные материалы для культивирования клеток и микроскопии;
- CO2-инкубаторы, шейкеры;
- Биореакторы, ферментеры.

ООО «Эппендорф Раша» является представительством немецкой компании Eppendorf AG на территории Российской Федерации. Головной офис компании находится в Москве. Компания осуществляет всестороннюю поддержку пользователей, занимается продвижением продукции на территории РФ, участвует в тематических выставках, конференциях, проводит демонстрацию продукции Eppendorf в шоу-руме в московском офисе и семинары в Москве и других городах России для всех интересующихся продукцией Eppendorf.

ООО «ЭППЕНДОРФ РАША»
115114, Москва,
Дербеневская набережная, д. 11, оф. Б301
Т.: (495) 743-51-23, Ф.: (495) 743-51-22
URL: www.eppendorf.ru,
E-mail: info@eppendorf.ru

Компания предлагает оборудование и реагенты для ПЦР в реальном времени и NGS:

- Реагенты для секвенирования нового поколения 454 (до сентября 2016 года)
- Реагенты для обогащения целевых фрагментов перед NGS – NimbleGen SeqCap (экзомы, метиломы, транксриптомы, панели генов)
- Системы для ПЦР в реальном времени – LightCycler 96 и LightCycler 480
- Системы экстракции нуклеиновых кислот – MagNA Pure LC 2.0, MagNA Pure Compact
- Система для гомогенизации тканей и других твердых образцов – MagNA Lyser

Компания предлагает покупателю не просто свою продукцию, а единую систему, включающую технический сервис, обучение персонала, постоянную методическую поддержку, своевременную доставку реактивов и расходных материалов.

Контакты:
ООО «РОШ ДИАГНОСТИКА РУС»
115114, Москва, ул. Летниковская, д. 2 стр. 2
(БЦ «Вивальди Плаза»)
Т.: (495) 229-69-99
Ф.: (495) 229-62-64
URL: www.roche-applied-science.ru
E-mail: russia.ras@roche.com

Agilent Technologies

Компания Аджилент Текнолоджиз является мировым лидером в разработке и производстве контрольно-аналитического оборудования. На рынке аналитического оборудования Agilent Technologies представляет приборы для молекулярной спектроскопии, элементного анализа, системы для газовой и жидкостной хроматографии, хромато-масс-спектрометрические системы, ПЦР, ПЦР в реальном времени, системы анализа микрочипов, автоматизированные системы дозирования, системы капиллярного электрофореза. Также Agilent Technologies представляет широкий круг продукции для молекулярной биологии, биотехнологии и геномного анализа. Клиентами компании Agilent Technologies являются лаборатории контроля качества и научно-исследовательские лаборатории. Компания Agilent Technologies в России предоставляет консультации по вопросам подбора оптимальной комплектации оборудования, проводит запуск оборудования и обучение работе на приборах, оказывает услуги гарантийного, постгарантийного обслуживания, услуги по квалификации и валидации систем.
Более подробную информацию о компании Аджилент Вы можете найти на сайте компании:

ООО «АДЖИЛЕНТ ТЕКНОЛОДЖИЗ»
115054, Москва,
Космодамианская набережная 52, строение 1 Б
Т.: (495) 664-73-00,
Ф.: (495) 664-73-01
URL: www.agilent.com,
www.your-analytical-solution.com

Компания Qvadros-Bio предлагает современные комплексные решения для исследовательских и медицинских лабораторий, в том числе:

- оборудование для организации биобанков любого масштаба – от ручных систем хранения, сканеров и принтеров штрих-кодов для небольших лабораторий до надежных автоматизированных систем хранения LiCONiC;
- оборудование для организации криобанков – криохранилища, криорезервуары и криогенные трубопопроводы CryoTherm, система хранения образцов C+CRYO, криозамораживатели и 2D криопробирки;
- многозадачные роботизированные платформы для автоматизации различных лабораторных операций: выделение ДНК, постановка ПЦР, пробоподготовка для NGS, масс-спектрометрии и хроматографии, проведение ИФА и аликвотирование;
- современное оборудование для наблюдения за клеточными процессами от пролиферации до ангио- и нейрогенеза, формирования 3D культур.
- Высококвалифицированные специалисты компании QvadroS-Bio окажут максимальную поддержку и проконсультируют по всем вопросам. Мы подберем оптимальное решение под Ваши задачи, составим спецификацию и предложим необходимые расходные материалы и дополнительное оборудование с учетом Ваших требований и пожеланий.

Сертифицированные сервисные инженеры компании QvadroS-Bio, прошедшие обучение на заводах производителей, проводят установку оборудования, обучение персонала, и обеспечивают качественную техническую поддержку, гарантийное и постгарантийное сервисное обслуживание оборудования.

Контакты:
ООО «Квадрос-Био»
127287, г. Москва,
Петровско-Разумовский пр., д.29, стр.4
Тел: +7 (495) 981-80-35
e-mail: info@qvadrosbio.ru

АЛЬБИОГЕН

ООО «АЛЬБИОГЕН» – молодая, активно развивающаяся компания, специализирующаяся на предоставлении полного комплекса услуг, связанного с продажей, всесторонней поддержкой пользователей и сервисным обслуживаем продукции компании Illumina.

- Наша команда состоит из специалистов самого высокого уровня с большим опытом работы как с продукцией компании Illumina, так и в области продаж, продвижения и поддержки.
- Мы с гордостью можем предложить Вам высочайший уровень логистики, который позволяет бережно и в кратчайшие сроки доставлять нашу продукцию до конечного пользователя.

Компания Illumina Inc. (Сан-Диего, США) инновационная, стремительно развивающаяся компания, являющаяся мировым лидером в области секвенирования нового поколения (NGS). Секвенаторы Illumina позволяют осуществлять генетические исследования для науки, медицины, сельского хозяйства, ветеринарии и криминалистики. Более 90% научных статей, связанных с технологиями секвенирования нового поколения, сделаны при помощи оборудования Illumina.

Компания АЛЬБИОГЕН является единственным официальным представителем Illumina в Российской Федерации, Республике Беларусь и Казахстане. Нашей задачей является обеспечение полного доступа клиентов к передовым технологиям и сервисам Illumina, включая наиболее современные системы NGS, биочиповые технологии и весь спектр реактивов.

ООО «АЛЬБИОГЕН»
тел. +7 (499) 550 15 25
info@albiogen.ru, www.albiogen.ru
Адрес: 129085, г. Москва, ул. Звездный бульвар дом 21, строение 3, пом. I, ком.5

Genotek

Компания Genotek оказывает услуги по высокопроизводительному секвенированию и генотипированию различных организмов, начиная с 2010 года и является ведущим поставщиком на рынке генетических услуг.

Наши клиенты – ведущие научно-исследовательские институты и фармкомпании России и стран СНГ, такие как Институт Общей генетики им. Вавилова, Институт цитологии и генетики РАН, Московская медицинская академия им. И.М. Сеченова и др.

Широкий спектр возможностей

Только в Genotek Вы можете выбрать любую из платформ для секвенирования в зависимости от поставленных задач – Illumina HiSeq2000, Illumina MiSeq, SOLID, Ion Proton.

Высокое качество результата

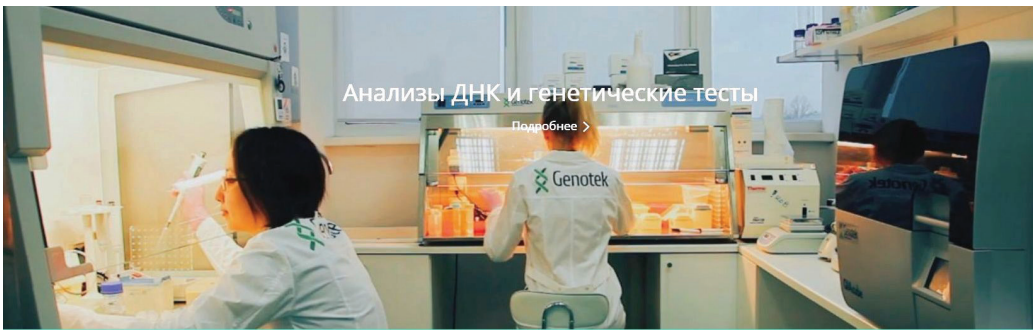
Строгое соблюдение протоколов производителей, использование только оригинальных реактивов и применение многоуровневой системы контроля качества позволяют нам гарантировать высокий уровень достоверности полученных данных.

Оптимальная стоимость

Значительный объем заказов, получаемых компанией, позволяет нам оптимально использовать производственные мощности оборудования, чтобы всегда предлагать Вам самые низкие цены.

Бесплатное консультирование

Не важно – заказываете Вы секвенирование одного ПЦР-продукта или полногеномное исследование сотен человек – Вы можете быть уверены, что получите ответы на любые возникающие вопросы совершенно бесплатно. Просто напишите или позвоните нам!



ООО «ГЕНОТЕК»
105120, г. Москва, Наставнический пер., д. 17,
стр. 1, пом. I, эт. 2, ком. 9.
Т.: (495) 215-15-14
URL: www.genotek.ru
E-mail: info@genotek.ru

Научное издание

THE TENTH INTERNATIONAL CONFERENCE ON BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE\SYSTEMS BIOLOGY

Abstracts
на английском

Printed without editing

ДЕСЯТАЯ МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ «БИОИНФОРМАТИКА РЕГУЛЯЦИИ И СТРУКТУРЫ ГЕНОМА\СИСТЕМНАЯ БИОЛОГИЯ»

Тезисы докладов

Составители: О.В. Петровская, Ю.Л. Орлов, С.В. Зубова

Выпуск подготовлен информационно-издательским отделом ИЦиГ СО РАН.
Начальник отдела Т.Ф. Чалкова
Дизайн и компьютерная верстка: А.В. Харкевич, О.Н. Савватеева

Подписано к печати 15.08.2016. Формат 70 × 108 $^{1}/_{16}$. Усл. печ. л. 31,85.
Тираж 420 экз. Заказ № 205

Федеральный исследовательский центр
«Институт цитологии и генетики Сибирского отделения Российской академии наук»
630090, г. Новосибирск, проспект Академика Лаврентьева, 10

Отпечатано в типографии ФГУП «Издательство СО РАН»
630090, г. Новосибирск, Морской проспект, 2