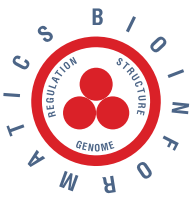

BGRS\SB-2018

The 11th International Conference

Systems Biology and Bioinformatics, SBB-2018

The 10th International Young Scientists School

SBB-2018
NOVOSIBIRSK, RUSSIA
27–31 AUGUST, 2018

CONF.BIONET.NSC.RU/SBB2018/

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

SYSTEMS BIOLOGY AND BIOINFORMATICS (SBB-2018)

The Tenth International Young Scientists School

Abstracts

27–31 August, 2018
Novosibirsk, Russia

Novosibirsk
ICG SB RAS
2018

УДК 575
S98

Systems Biology and Bioinformatics (SBB-2018) : The Tenth International Young Scientists School (27–31 Aug. 2018, Novosibirsk, Russia); Abstracts / Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences. – Novosibirsk: ICG SB RAS, 2018. – 57 pp. – ISBN 978-5-91291-043-2.

ORGANIZERS

Institute of Cytology and Genetics SB RAS (ICG SB RAS)
Novosibirsk State University (NSU)
Chair of Information Biology NSF NSU
Genetics and Breeding Scientific Advisory Board
The Vavilov Society of Geneticists and Breeders (VSGB)

SPONSORS

Russian Foundation
for Basic Research
Grant No. 18-01-20052 Г

International program committee

Prof. Nikolay A. Kolchanov, Full Member of the RAS, Institute of Cytology and Genetics of SB RAS, Novosibirsk, Russia (Chairman)
Prof. Ralf Hofestädt, University of Bielefeld, Germany
Prof. Inna Lavrik, Otto von Guericke University, Magdeburg, Germany
Prof. Evgenii Rogayev, University of Massachusetts, USA
Dr. Maxim Zakhartsev, Hohenheim University, Institute of Plant Physiology and Biotechnology, German, Norwegian University of Life Sciences, Ås, Norway
Dr. Yuriy Orlov, Novosibirsk State University, Russia
Prof. Ivo Grosse, Halle-Wittenberg University, Halle, Germany
Dr. Dmitry Afonnikov, Institute of Cytology and Genetics of SB RAS, Novosibirsk State University, Novosibirsk, Russia
Dr. Sergey Lashin, Institute of Cytology and Genetics of SB RAS, Novosibirsk State University, Novosibirsk, Russia
Dr. Olga Krebs, Heidelberg Institute for Theoretical Studies, Germany
Prof. Yuriy Aulchenko, Institute of Cytology and Genetics of SB RAS, Novosibirsk State University, Novosibirsk, Russia
Dr. Fyodor Kondrashov, Institute of Science and Technology, Austria
Prof. Peter Joshi, University Hospital of Lausanne, Lausanne, Switzerland; University of Edinburgh, Edinburgh, Great Britain

Local organizing committee

Olga Petrovskaya, ICG SB RAS (chair)
Svetlana Zubova, ICG SB RAS
Tatyana Chalkova, ICG SB RAS
Nadezhda Glebova, ICG SB RAS
Tatyana Karamysheva, PhD, ICG SB RAS
Sergey Lashin, PhD, ICG SB RAS
Alina Morkovina, ICG SB RAS
Erlan Tokpanov, ICG SB RAS
Andrey Kharkevich, ICG SB RAS

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
e-mail: sbb2018@icg.sbras.ru
popik-olga@bionet.nsc.ru

ISBN 978-5-91291-043-2

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

Director: Corr. Member of the RAS Alexey V. Kochetov
Science Head: Professor, Full Member of the RAS Nikolay A. Kolchanov
Academic Secretary: Candidate of Sciences (Biology) Galina V. Orlova
Phone: +7(383) 363-49-85 ext. 1336; e-mail: OrlovaGV@icg.sbras.ru, gorlova@bionet.nsc.ru

The Institute was founded in 1957, among the first institutions of the Siberian Branch of the Russian Academy of Sciences. Presently, ICG SB RAS is an interdisciplinary biological center, which ranks among the leading biological institutions in Russia. The second step of the restructuring of the Federal Research Center Institute of Cytology and Genetics was completed in May 2017. Presently, ICG includes three branches:

- Siberian Research Institute for Plant Industry and Breeding,
- Institute of Clinical and Experimental Lymphology, and
- Institution of Internal and Preventive Medicine.

Tasks of ICG SB RAS: Solution of top-priority problems in the development of the Russian science and technology sector in the fields of plant genetics and breeding, animal genetics and breeding, human genetics, biotechnology, and fundamental medicine by applying methods of molecular genetics, cell biology, and computational biology.

Strategic goal: Integrated studies in plant genetics and breeding, animal genetics and breeding, human genetics, fundamental medicine, and biotechnology by applying methods of molecular genetics, cell biology, and computational biology from the generation of academic knowledge to the solution of top priority problems set by Russian agricultural, biotechnological, biomedical, and pharmaceutical industries.

Staff: As on January 1, 2018, ICG included 88 scientific units, which employed 1397 members, of them 492 researchers, 2 RAS Advisors, 8 Full Members of the RAS, 4 Corresponding Members of the RAS, 98 Doctors of Science, and 305 Candidates of Science.

Postgraduate education and residency training: As on January 1, 2018, ICG trains 87 postgraduates and 18 interns.

Publications: The Institute intensely contributes to Russian and foreign journals and ranks among acknowledged leaders in Russian biology. In 2017, 344 papers were enrolled in the international online service Web of Science (WoS). The overall number of publications in peer-reviewed journals exceeds 450. In 2017 papers of the researchers of the Institute received over 4719 WoS citations.

Auxiliaries: Core facility "Center for Genetic Resources of Laboratory Animals", which includes the unique research unit "SPF vivarium", and seven shared access centers www.bionet.nsc.ru/uslugi.

The Federal Research Center Institute of Cytology and Genetics is looking to cooperate with scientific and commercial enterprises.

Address: Prosp. Akad. Lavrentiev 10, Novosibirsk, 630090 Russia
Phone/fax: +7(383) 363-49-80 / +7(383) 333-12-78
URL: www.bionet.nsc.ru
e-mail: icg-adm@bionet.nsc.ru

Contents

DOI 10.18699/SBB-2018-01

CpG islands' clustering uncovers early development genesin the human genome

R.O. Babenko[1], A.V. Tsukanov[1, 2], A.G. Galieva[1], Y.L. Orlov[2], V.N. Babenko[2, 3]
1 *Novosibirsk State University, Novosibirsk, Russia*
2 *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** e-mail: rbab@yandex.ru*

Key words: non-coding RNA, transcriptomics, stress response, crop plants

Motivation and Aim: We address the problem of the annotation of CpG islands (CGIs) clusters in the human genome. CpG dinucleotide rich genome regions, also known as CpG islands (CGIs), are important functional elements of vertebrate genomes [1]. In particular, in the majority of vertebrate genes, CpG islands coincide with gene promoter areas. In some cases, the transcription from CpG-island containing promoters is bidirectional, this is related to self-complementarity of CG dinucleotides. CGIs are the key contributors to global methylation landscapes. Degenerate content of CGIs (biased CG frequency) assumes a higher probability of tandem repeats and palindromes inside a CGI.
Methods and Algorithms: We have used own program scripts for tandem repeats and CGI counts. We used a CGI clustering method that is robust relative to the tandem duplication search A set of CGIs was retrieved from the table cpgIslandExt (www.genome.ucsc.edu; version hg19). To identify significant CGI clustering, the human genome was split into 10Kb non overlapping segments (bins) (243 785 bins in total). The number of CGIs per bin (CGI density) was assessed as a total number of CGIs divided by the number of bins. The expected number of CGIs per segment was approximated using a Poisson distribution
Results: Upon analyzing gene content within CGIs clusters, piRNA, tRNA, and miRNA-encoding genes were found as well as CpG-rich homeobox genes reported previously. Chromosome-wide CGI density is positively correlated with replication timing, confirming that CGIs may serve as open chromatin markers. Early embryonic stage expressed KRAB-ZNF genes abundant at chromosome 19 were found to be interlinked with CGI clusters.
Conclusion: We detected that a number of long CGIs and CGI clusters are, in fact, tandem copies with multiple annotated macrosatellites and paralogous genes. This finding implies that tandem expansion of CGIs may serve as a substrate for non-homologous recombination events [1].
Acknowledgements: The research has been supported by RFBR. Computing done at Siberian Supercomputer center SB RAS was supported by budget project 0324-2018-0017.

References

1. Babenko V.N., Bogomolov A.G., Babenko R.O., Galieva E.R., Orlov Y.L. (2018) CpG islands' clustering uncovers early development genes in the human genome. Computer Science Information Systems. 15(2):473-485.

DOI 10.18699/SBB-2018-02

Establishing free-living flatworm *Macrostomum lignano* as a model to study links between regeneration and cancer

M. Biryukov[1, 2]*, I. Sukhikh[1], V. Vavilova[1], K. Ustyantsev[1], E. Berezikov[1, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *European Research Institute for the Biology of Ageing, Groningen, The Netherlands*
* *e-mail: mikebirykov@yandex.ru*

Key words: *Macrostomum lignano*, flatworm, carcinogens, toxicity, lethality

Motivation and Aim: The restriction of regeneration is thought to evolve as a measure against cancer. Highly regenerative flatworms do not have cancer, suggesting that they evolved some other regulatory mechanism to keep their cells under control while allowing the process of regeneration. However, it is possible to induce carcinogenesis in flatworms with chemical mutagens [1, 2]. *Macrostomum lignano* is a free-living flatworm rapidly emerging as a model organism to study regeneration and stem cell biology [3]. $CdSO_4$ and 12-O-tetradecanoylphorbol 13-acetate (PMA) are established chemical carcinogens with known tumorigenic effect in planarians [1, 2]. The aim of this study is to find optimal concentration for $CdSO_4$ and PMA induced carcinogenesis in *M. lignano*, which should be a first step in dissecting conserved regulatory gene networks interplayed between regeneration and cancer.
Methods and Algorithms: *M. lignano* worms were grown in artificial seawater enriched with f/2-medium. Various concentrations of $3CdSO_4 \cdot 8H2O$, PMA (diluted in DMSO), and their combinations were tested for toxicity. Twenty adult worms were continuously exposed to various concentrations of the carcinogens for 2 weeks (the medium was refreshed weekly), and the percentage of lethality was calculated for each of the conditions.
Results: Thirty different carcinogen concentrations were tested for CdSO4, PMA, and $CdSO_4$+PMA, respectively. The percentage of the worm lethality was determined for each of the conditions. The toxicity was established as lethal concentration 50 (LC_{50}). Concentrations of the carcinogens suitable for chronic exposure of worms and tumor growth induction were selected.
Conclusion: The results obtained can serve as a starting point to the carcinogenesis study in *M. lignano*.
Acknowledgements: The project was supported by the Russian Foundation for Basic Research (Grant No. 18-04-01011).

References

1. Hall F. et al. (1986) Neoplastic transformation in the Planarian: I. Cocarcinogenesis and tiistopathology. Journal Experimental Zoology. 240:211-227.
2. Voura E.B. et al. (2017) Planarians as models of cadmium-induced neoplasia provide measurable benchmarks for mechanistic studies. Ecotoxicology Environmental Safety. 142:544-554.
3. Wudarski J. et al. (2017) Efficient transgenesis and annotated genome sequence of the regenerative flatworm model Macrostomum lignano. Nature Communications. 8:2120.

DOI 10.18699/SBB-2018-03

Genomic variation in Austrian and Leningrad populations of snail *Arianta arbustorum* L.

O. Bondareva[1]*, A. Kasianov[2], N. Abramson[1]
[1] *Zoological Institute RAS, St.-Petersburg, Russia*
[2] *Vavilov institute of general genetics, Moscow, Russia*
** e-mail: olga.v.bondareva@gmail.com*

Key words: invasion, genetic paradox, *Arianta arbustorum* L., RNA-seq

Motivation and Aim: The invasion occurs in extremely short periods, which allows us to consider them as "natural experiments", which provide a unique opportunity to observe the ecological and evolutionary processes in real time. Among the species of terrestrial ecosystems, one example of the extremely successful invasion of Russia in the North-West region of recent years is the tree snail *Arianta arbustorum* (Linneus, 1758).
Methods and Algorithms: In order to analyze the overall variability that potentially could lead to the success of *A. arbustorum* snail migration, we made a comparative analysis of single nucleotide polymorphysm (SNP calling) in transcripts of individuals from different populations: from Austria, which are part of the original range of *A. arbustorum*, and from the vicinity St. Petersburg, where this species was discovered relatively recently. We assembled a transcript using all samples with the following statistical parameters: Contig N50: 713, Median contig length: 372, Average contig: 564.45. After the functional annotation of the transcriptome with Transdecoder programm, there were 54 thousand contiguous verified, which we used as a reference for the search for single nucleotide substitutions.
Results: The final analysis confirmed the much lower variability of individuals from the Leningrad Region in comparison with the Austrian population, that we showed before using *COI* partial sequence. The main characteristics of the substitutions remain unchanged between the populations – a much larger number of transitions than transversions, the prevalence of substitutions in the 3 positions of the codon compared with the replacement in other positions and the predominance of the number of synonymous substitutions over non-synonymous ones.
Conclusion: The data obtained during the meta-analysis confirm our hypothesis that the settlement of the Leningrad region occurred by a group of founders with low haplotypic variability, who then proceeded to expansive reproduction.
Acknowledgements: This study was conducted in Zoological Institute RAS under the research theme No. AAAA-A17-117042410167-2 and supported by the RFBR 16-34-00958 and RFBR 15-29-02526.

DOI 10.18699/SBB-2018-04

Identification of loci determining resistance of spring barley to spot and net blotch, using association mapping approach

I.V. Bykova[1]*, S.A. Gorobets[1], N.M. Lashina[2], V.M. Efimov[1], O.S. Afanasenko[2], E.K. Khlestkina[1, 3]

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

[2] *All-Russian Research Institute for Plant Protection, St.-Petersburg, Russia*

[3] *N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources* (*VIR*)*, St.-Petersburg, Russia*

** e-mail: bykova@bionet.nsc.ru*

Key words: *Hordeum vulgare*, SNP, *Cochliobolus sativus*, *Pyrenophora teres F. teres*

Motivation and Aim: Spot blotch, caused by *Cochliobolus sativus*, and net blotch, caused by *Pyrenophora teres F. teres* are two of the most widespread and harmful diseases in barley. Identification of genetic loci associated with resistance to both *C. sativus* and *P. teres F. teres* is of importance for future marker-assisted selection. The goal of the current study was to identify loci conferring seedling resistance to different pathotypes of *C. sativus* and *P. teres F. teres* in the Siberian spring barley core collection.

Methods and Algorithms: A total of 96 spring barley cultivars and lines were phenotyped at the seedling stage with two *C. sativus* isolates (Kr2 and Ch3) and four *P. teres* isolates (S 10.2, K 5.1, P 3.4.0, A 2.6.0). About 16–17 % and 10–23 % of genotypes were resistant to spot blotch and net blotch isolates, respectively, and 26–30 % and 5–17 % were moderate-resistant to spot and net blotch isolates, respectively. A total of 94 genotypes were analyzed with the barley 50K Illumina Infinium iSELECT assay. From 44,040 SNPs, 40,703 were scorable, from which 39,140 were polymorphic. 27,319 SNPs passed filtering threshold and were used for association mapping. The data were assessed using Microsoft Excel, PASS, Tassel 5.

Results: Data analysis by GLM revealed 3 and 27 SNPs for Kr2 and Ch3 spot blotch isolates, respectively, and 2, 27, 2, 26 SNPs for S10.2, K5.1, P3.4.0, A2.6.0 net blotch isolates, respectively. A total of three genomic regions were associated with the resistance to spot blotch on chromosomes 1H, 2H, 3H and three genomic regions were associated with resistance to net blotch on chromosomes 2H, 3H and 6H.

Conclusion: Three genomic regions associated with the resistance to one or both isolates of *C. sativus* and to one of isolates of *P. teres F. teres* were identified via screening of the Siberian spring barley core collection. Comparison of their location with QTLs revealed previously either with biparental mapping populations studies or with GWAS of distinct germplasm and other isolates, demonstrated that resistance to isolates of all pathogens is conferred by known spot blotch and net blotch resistance loci. Information on SNPs related can be used further for development of DNA-markers convenient for diagnostics of resistance-associated alleles in barley breeding programs.

Acknowledgements: This study was supported by RSF No. 16-14-00086.

DOI 10.18699/SBB-2018-05

Founder effect in prevalence of hereditary hearing loss in indigenous Siberian populations

V. Danilchenko[1, 2]*, M. Zytsar[1, 2], O. Posukh[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: mikhalskaya@bionet.nsc.ru*

Key words: hereditary hearing loss, founder effect, *GJB2*, *RAI1*, indigenous populations of Siberia

Motivation and Aim: Prevalence of many monogenic diseases can be determined by specific demographic and population factors (ethnic composition, migration, isolation, founder and bottleneck effects, proportion of consanguineous and assortative marriages). Nonsyndromic hearing loss (HL) is one of the most common monogenic disorders and several dozen genes contribute to its pathogenesis. It is well known that pathogenic variants in gene *GJB2* (MIM 121011, 13q12.11) encoding connexin 26 (Cx26) account for a significant portion of hereditary HL and their spectrum and prevalence are highly specific for various populations. There are more limited data on prevalence of pathogenic variants in other genes associated with HL. We earlier found predominance of three major recessive *GJB2* pathogenic variants (c.-23+1G>A, c.235delC, p.W172C) in indigenous populations of Tuva and Altai and revealed common haplotypes for each of them that implies founder effect in their prevalence. Whole exome sequencing was applied to identifying candidate causal variants for undiagnosed hereditary HL cases in Altai [1]. Homozygous novel variant c.5254G>A in gene *RAI1* (MIM 607642, 17p11.2) was revealed in association with HL in several Altaian families and its carrier frequency was estimated as 3.33 % in Altaian control sample while c.5254G>A was not found in Tuvinian patients and controls. This study aims to evaluate the role of founder effect in prevalence of major pathogenic variants in genes associated with nonsyndromic HL among indigenous populations of Siberia.
Methods and Algorithms: To investigate genetic background of pathogenic variant c.5254G>A in the *RAI1* gene, Sanger sequencing was applied for analysis of the *RAI1* gene coding region encompassing exons 3, 4, 5 and part of exon 6 with flanking intronic regions in all Altaians homozygous or heterozygous for c.5254G>A.
Results: All studied Altaian individuals with c.5254G>A share a specific allele C-A-Q_{13}[CAG CAA $(CAG)_{10}$ del(CAG) CAA]-G-c.5254G>A that implies the common origin of this pathogenic variant in Altaians.
Conclusion: Identification of uniform allele bearing pathogenic variant c.5254G>A in gene *RAI1* along with specific common genetic background for major pathogenic variants c.-23+1G>A, c.235delC, p.W172C in gene *GJB2* confirms important role of founder effect in prevalence of nonsyndromic HL among indigenous populations of Siberia.
Acknowledgements: Supported by the Project 0324-2018-0016 and by the RFBR (18-34-00166_mol-a, 17-29-06016_ofi-m).

References

1. Churbanov A.Y. et al. (2016) Whole exome sequencing reveals homozygous mutations in *RAI1*, *OTOF*, and *SLC26A4* genes associated with nonsyndromic hearing loss in Altaian families (South Siberia). PLoS One. 11(4):e0153841. DOI 10.1371/journal.pone.0153841.

DOI 10.18699/SBB-2018-06

Genetic diversity of facultative bread wheat from the VIR collection

A.V. Dementiev*, E.K. Potokina, O.P. Mitrofanova
N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St.-Petersburg, Russia
** e-mail: alek.dementiew2010@yandex.ru*

Key words: facultative wheat, *Vrn*- and *Ppd*-genes, PCR analysis, heading time

Motivation and Aim: Facultative bread wheat (*Triticum aestivum* L.) is cultivated in different countries of the world and adapted to wide range of environments. In Russia this wheat is used as an insurance crop. For expanding of genetic basis for breeding of facultative wheat not only in the North-Caucasian region, where this wheat is cultivated, but also in other regions, it was essential to assess its world genetic diversity, to understand, how it is represented in the VIR collection, and to form core-set for usage in breeding and basic researches. The aim of our study was to use molecular markers to characterize the major alleles of genes responsible for vernalization response and photoperiod sensitivity and to elucidate the relationships between some biological or agronomic traits and allelic variations of these genes.
Methods and Algorithms: According to GRIS [1], 2116 accessions of facultative bread wheat are known in the world, among them 651 – in the VIR collection. We have formed a representative sample of 345 accessions, which were evaluated in spring- and autumn-sownin the North-West region of Russia (t. Pushkin). From this sample 243 accessions (256 biotypes) were chosen for molecular studies. Total DNA was extracted from the six seedlings, descendants of the one plant (genotype). PCR was performed using allele-specific molecular markers for determination the main allele-types for Vrn- and Ppd-genes [2].
Results: A total 15 alleles for *Vrn-A1*, *Vrn-B1*, *Vrn-D1*, *Vrn-B3*, *Ppd-D1* genes, and for *Vrn-D1* promoter region were observed. Frequencies of dominant alleles were varied from 0.8 % (*Vrn-B3*) to 36.7 % (*Vrn-B1a*), whereas of recessive alleles – from 43.8 % (vrn-b1) to 94.5 % (wild type of *Vrn-D1* promoter region). The amplified products of other sizes for genes *Vrn-A1* and *Vrn-B3* were found in 18 biotypes. No products characteristic for individual genes were detected in 52 biotypes. Four biotypes were heterozygous. In total, 64 allelic compositions were identified in 256 biotypes. The most frequent composition was *vrn-A1 Vrn-B1c vrn-D1 vrn-B3 Ppd-D1b* wild type of *Vrn-D1* promoter region, which was revealed for biotypes from Russia, Kursk region. The genotypes with alleles *Vrn-B1c* had shorter heading than *Vrn-B1a*.The degree of overwintering of biotypes was mainly related to geographic origin of accessions, from which the biotypes were isolated.
Conclusion: The core-set of facultative bread wheat comprising 256 biotypes is maintained the wide genetic diversity not only alleles of *Vrn*- and *Ppd*-genes but also the other traits. It can be used for wheat improvement and conducting basic researches.

References

1. http://wheatpedigree.net/
2. http://maswheat.ucdavis.edu/protocols/Vrn/index.htm

DOI 10.18699/SBB-2018-07

Search of single-nucleotide polymorphisms associated with accelerated senescence in OXYS rats

V. Devyatkin[1, 2]*, N. Muraleva[1], O. Redina[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: devyatkin@bionet.nsc.ru*

Key words: aging, age-related diseases, OXYS rats, cataract, SNP

Motivation and Aim: Aging is the single largest risk factor for chronic disease, still little is known about a genetic overlap between complex age-related diseases. For search pathways that modulate the onset and progression of multiple age-related diseases here we used senescence-accelerated OXYS rats derived from Wistar rats in the ICG SB RAS (Novosibirsk) and developing a phenotype similar to human geriatric disorders including cataract, age-related macular degeneration-like retinopathy and neurodegenerative pathology of the brain with features of Alzheimer's disease. We hypothesize that the senile cataract development can serve as a biomarker of systemic changes associated with aging. The purpose of our work is to identify in the genome of OXYS rats mutations in genes associated with cataract, which can potentially contribute to the development of signs of accelerated aging.
Methods and Algorithms: We used the data of RNA-Seq from prefrontal cortex, retina and hippocampus of senescence-accelerated OXYS and Wag (control) rats. Positions of SNPs within the aligned reads relative to the reference genome (Rnor 6.0) were identified using SAMtools (v. 0.1.17) utilities. Each mutation was present in at least 3 OXYS rats in homozygous state and was not present in any of the Wag rats. The effect of an amino acid substitution on protein function was predicted by the Variant Effect Predictor Web service; the consequence type, SIFT score, and prediction were obtained for each variant. The list of genes associated with cataracts was compiled according to NCBI, Cat-Map, KEGG Disease databases.
Results: In the genome of OXYS rats 52539 SNPs in the homozygous state, not presented in the genome of Wag rats, overlapped with 8012 genes and 11684 transcripts were revealed. In 328 cases the substitutions can result in significant structural rearrangements (high impact effect) of the transcripts. Among the non-synonymous substitutions 254 have a deleterious effect on the structure or function of the protein product according to the SIFT algorithm. We revealed SNPs in 255 genes that can be associated with cataract development in OXYS rats and contained 543 described and 614 novel SNPs. 4 of this genes, *Pex2*, *Nbn*, *Rab18* and *Prss56* have SNPs (rs198310567, rs105362013, rs106234270 and rs106604882, respectively) with a deleterious result according to SIFT, although these polymorphisms are described for SHR/OLAIPCV and SD rat strains without signs of cataracts. However, it is known that mutations in these genes are associated with a number of mitochondrial diseases, nervous and cardiovascular disorders, consistent with the complex manifestation of the senile phenotype in OXYS rats against the background of cataract development.
Conclusion: Genes with mutations revealed in OXYS rats are promising for further verification of the contribution of found polymorphisms to the development of complex age-related diseases.
Acknowledgements: Supported by the RFBR (project No. 18-015-00336, No. 18-315-00216).

DOI 10.18699/SBB-2018-08

Investigation of genetic polymorphism of *Avena sativa* varieties and *Avena sterilis* samples using SSR-markers

N.I. Drobot
Institute of Genetics and Cytology NASB, Minsk, Belarus
e-mail: n.drobot@igc.by

Key words: *Avena sativa, Avena sterilis*, SSR-marker, PIC, MI

Motivation and Aim: Hexaploid oat (*Avena sativa*) is a valuable cereal crop worldwide. The main directions of modern oats breeding are improvement of grain quality and resistance to biotic and abiotic environmental factors, among which the most significant are insects and diseases. The most effective way to control plant diseases is to create resistant plant varieties. Wild-growing hexaploid species that have a genomic structure similar to *Avena sativa* L. (AACCDD), which guarantees the production of fertile hybrid forms in crosses are of a particular interest for breeding.
Methods and Algorithms: The material for this study was 15 varieties of *Avena sativa* L. (Fax, Myrt, Debut, Freestyle, Lydia, Zapavet, Stramec, Zolak, Ivory, Bingo, Stoper, Sprinter, Gagubatori kh, AC Goslin, AC Fracis), 12 samples of *Avena sterilis* L., and four highly productive hybrids. Genomic DNA purification was carried out from green leaves using the Plant DNA Preparation Kit (Jena Bioscience). The genetic polymorphism was determined using a set of 8 SSR-markers: *AM1*, *AM3*, *AM4*, *AM5*, *AM7*, *AM15*, *AM22*, *AM83* with standard fragment analysis (ABI 3500).
Results: As a result, 54 alleles were detected in 8 SSR-loci. The number of alleles ranged from 2 to 10, and an average amount was 6.75 per locus. The frequency of occurrence of alleles in the analyzed sample ranged from 0.06 to 0.8. At the same time, the informative index (PIC) was from 0.2 (AM5) to 0.89 (AM3), the marker index (MI) – from 0.01 to 1.6 (*Table*).

Characteristics of the SSR-markers

SSR-marker	Alleles number	Obtained size, b.p.	Expected size, b.p.	Frequency of occurrence	**PIC** (pol.inf. content)	**MI** (marker index)
AM1	9	152–213	157–225	0.07–0.25	0.88	1.31
AM3	9	257–316	243–325	0.06–0.1	0.89	1.32
AM4	7	129–150	133–227	0.08–0.21	0.83	0.75
AM5	2	131–134	172	0.1–0.8	0.2	0.01
AM7	10	149–191	155–198	0.06–0.23	0.87	1.6
AM15	4	223–229	229	0.1–0.6	0.57	0.17
AM22	8	168–304	138	0.14–0.2	0.82	0.97
AM83	4	179–189	187–197	0.001–0.08	0.58	0.17

Conclusion: The highest level of polymorphism was observed for loci AM7, AM3, AM1 and AM22 (PIC – 0.87; PIC – 0.89; PIC – 0.88; PIC – 0.82 respectively). At the same time loci AM5 and AM15 were characterized by low polymorphism (PIC is 0.2 and 0.57, respectively) and the minimum number of alleles, that makes their using for studying genetic diversity inexpedient.
Acknowledgements: Supported by project 2.08 GPNI “Biotechnology”.

DOI 10.18699/SBB-2018-09

Early and late transcriptional responses to the low concentration of salicylic acid in *Arabidopsis thaliana* L. root

E. Elgaeva[1]*, E. Zemlyanskaya[1, 2], D. Novikova[1, 2], V. Mironova[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** e-mail: elgaeva.liza@yandex.ru*

Key words: SA, RNA-Seq, development, bioinformatics, transcriptome, concentration-dependent response

Motivation and Aim: Salicylic acid (SA) is a stress hormone in plants, also it was shown that SA regulates plant growth and development [1]. In majority of studies, the influence of high SA concentrations (> 100 μM), mainly on the aboveground part, and after a prolonged exposure to SA (~24 hours) has been studied. As a result, molecular-genetic mechanisms of low SA concentrations remain obscure, the same is true for the mechanisms of SA response in the root and early transcriptional response to SA. The goal of this study was to describe transcriptome response to low SA concentrations in early and late response in *A. thaliana* root.
Methods and Algorithms: We used RNA-Seq data, generated from roots of 3 dag seedlings of *A. thaliana* after SA-treatment (20 μM, 1h and 6h) and sequenced on Illumina-HiSeq4000 (BGI Tech, Hong Kong). RNA-Seq data analysis included quality control (FASTQC), preprocessing of the reads (Trimmomatic), mapping (STAR) and differential expression analysis (DESeq2, EdgeR). Functional annotation of differentially expressed genes (DEGs) was made with agriGO (http://systemsbiology.cau.edu.cn/agriGOv2) and DAVID (https://david.ncifcrf.gov/).
Results: We detected 1873 and 3662 DEGs in early and late response in *A. thaliana* root respectively. Low SA concentrations induced three times more genes than inhibits in both early and late response. Intriguingly, among the DEGs we found many genes providing for SA signaling under its high dosage. Consequently, functional enrichment analysis of DEGs showed many GO terms associated with defense and immune system even in the samples exposed to low-level SA for 1h only. After 6h-treatment was found a term «SA metabolism», that assumed activation of SA catabolism in late SA response.
Conclusion: Here we identified molecular targets of low SA concentrations in early and late response to low SA concentrations. Interesting that low level SA activates many defence reactions (through oxylipin, cytochrome, flavonoids, etc.). Next question is to compare the transcriptome response to low and high SA concentrations.
Acknowledgements: The work was supported by RFBR 18-04-01130.

References

1. Vlot A.C., Dempsey D.A., Klessig D.F. (2009) Salicylic Acid, a Multifaceted Hormone to Combat Disease. Annu. Rev. Phytopathol. 47(1):177-206.

DOI 10.18699/SBB-2018-10

Phenotype characters statistical analysis for selection perspectives

T.V. Erst[1]*, V.K. Khlestkin[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: erst@bionet.nsc.ru*

Key words: potato, starch, granule, morphology

Motivation and Aim: Potato is one of the most important crop plants. In addition to the traditional use for food purposes, potato is a renewable resource of such technical material as potato starch. Potato starch properties depend on its composite polysaccharides ratio and structure – linear amylose and branched amylopectin – and are regulated by starch biosynthesis genes. Despite starch biosynthesis genes are sequenced and studied, effects of allelic variations on polysaccharides molecular structures and, consequently, starch physicochemical properties, are not known. We carry out chemical phenotyping of potato starch varieties for developing scientific approaches to the most rational production and use of new varieties producing the optimal starch for a particular application. Contemporaneously, potato varieties genotyping and further "genome – starch property" correlation determination is carried out.
Methods and Algorithms: Microscopy of starch granules and study of their morphology is one of the way to phenotyping starch and to selection of the contrast forms. It is known, that amylopectin chains branching, size [1] and irregular granule shape [2] are affected by genes *SBE I*, *II* (starch branching enzyme). Tuber starch content and irregular granule shape [2] likely to be affected by genes *SSI-SSIV* (starch synthase). Microscopy images processing using ImageJ software [3] allow quickly get the data on the granules' area projection in the frame, Feret's diameter, circularity, roundness, aspect ratio simultaneously for hundreds and thousands granules. The morphology data obtained can be processed by mathematical statistics methods and treated as biological traits. We clustered our varieties in the Past software using the Word's method.
Results: We have treated 57 varieties of the 2017 year. Using the average values of listed morphological parameters, preparative yield and information on ripeness we performed clustering of the data and obtained four clusters. Using principal component analysis we obtained a varieties distribution which shows that there are significantly different varieties in our selection. Based on the obtained varieties distribution it can be concluded there is some heterogeneity in our selection, which gives us opportunities in further selection. Thus, for new varieties creations with altered values of interesting features when crossing, it is necessary to use a selection of varieties with different values of the features of interest.
Acknowledgements: Supported by the RFBR (17-29-08006).

References

1. Hofvander P., Andersson M., Larsson C.-T., Larsson H. (2004) Field performance and starch characteristics of high amylose potatoes obtained by antisense gene targeting of two branching enzymes. Plant Biotechnol. J. 2:311-320.
2. Schwall G.P., Safford R., Westcott R.J., Jeffcoat R., Tayal A., Shi Y.-Ch., Gidley M.J., Jobling S.A. (2000) Production of very-high-amylose potatostarch by inhibition of SBE A and B. Nat. Biotechnol. 18:551-554.
3. Khlestkin V.K., Erst T.V. (2017) A practical guide to the starch granules' morphology study by microscopy. Vavilov Journal of Genetics and Breeding. 21(6):728-734 (in Russian).

DOI 10.18699/SBB-2018-11

The GAME genes as a target for *de novo* domestication of wild potato

S.V. Gerasimova[1]*, K.A. Ivanova[1], E.K. Khlestkina[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources* (*VIR*)*, Saint-Petersburg, Russia*
** e-mail: gerson@bionet.nsc.ru*

Key words: domestication, potato, Solanaceae, steroidal glycoalkaloids, GAME genes, genome editing

Genome, transcriptome, and metabolome analysis of wild and domestic plants reveals so-called domestication genes associated with different traits of crop plants. *De novo* domestication via targeted modification of these genes is a new approach to accelerate domestication of wild plants, create new donors of valuable traits and increase a diversity of existing domestic forms. In the domestication process, humans have selected plants with desired traits. One of the main directions in domestication of edible Solanaceae species was selection against bitterness and toxicity. Wild potato species have bitter tubers and wild tomato has bitter fruits due to high steroidal glycoalkaloid (SGA) content. Recent study of genome changes associated with domestication of Solanaceae sp. revealed signatures of selection in genes controlling SGA biosynthesis and regulation (glycoalkaloid metabolism, GAME genes). In both potato and tomato genomes the GAME9 gene was found to be associated with domestication process [1, 2]. This gene encodes AP2/ERF transcription factor, shown to be the key transcriptional regulator of other GAME genes encoding enzymes in the SGA-specific pathway. Strong domestication signatures were observed in squalene synthase (SQS) gene of potato [1]. This gene controls early steps of SGA synthesis. For tomato genome, it was shown that selection of five major loci reduced the accumulation of SGA in fruits [2]. These loci contained few known GAME genes and a co-expression gene cluster potentially involved in SGA biosynthesis. This data allows us to consider the GAME genes as domestication genes for Solanaceae sp. The SGA biosynthesis is realized via the cytosolic mevalonate pathway and consists of three stages. The first two stages are required for the synthesis of primary metabolites, and lead to cycloartanol and cholesterol, respectively. At the third stage (the synthesis of glycoalkaloids from cholesterol), about 20 enzymes participate. In the potato genome, 14 corresponding genes were identified [3]. The reduction of SGA content in wild potato species via targeted modification of different GAME genes will create a new experimental model of domestication and provide a donor material for potato breeding. The both modeling of existing domestication events and induction of alternative changes in the GAME genes may provide essential information about regulation of SGA metabolism and genetic background of Solanaceae domestication.
Acknowledgements: The study is supported by the RFBR (18-316-00068).

References

1. Hardigan M.A. et al. (2017) Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato. Proc. Natl. Acad. Sci. 201714380.
2. Zhu G. et al. (2018) Rewiring of the Fruit Metabolome in Tomato Breeding. Cell. 172:249-255.e12.
3. Ivanova K.A. et al. (2018) The biosynthesis regulation of potato steroidal glycoalkaloids. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 22(1):25-34 (in Russian).

DOI 10.18699/SBB-2018-12

Identification of genes, associated with black pigmentation of seeds in cereals, based on transcriptomic analysis

A.Yu. Glagoleva[1, 2]*, N.A. Shmakov[2], O.Yu. Shoeva[2], G.V. Vasiliev[2], N.V. Shatskaya[2], A. Börner[3], D.A. Afonnikov[1, 2], E.K. Khlestkina[1, 2, 4]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3] *Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany*
[4] *N.I. Vavilov All-Russian Research Institute of Plant Genetic Resources (VIR), St.-Petersburg, Russia*
** e-mail: glagoleva@bionet.nsc.ru*

Key words: barley, near-isogenic lines, phytomelanin, Blp, RNA-seq, differetial expression

Motivation and Aim: Some plant species have 'melanin-like' black seed pigmentation. However, the chemical and genetic nature of this 'melanin-like' black pigment has not yet been fully explored due to its complex structure and ability to withstand almost all solvents. Nevertheless, identification of genetic networks participating in trait formation is key to understanding metabolic processes involved in the expression of 'melanin-like' black seed pigmentation. The aim of the current study was to identify differentially expressed genes (DEGs) in barley near-isogenic lines (NILs) differing by allelic state of the *Blp* (black lemma and pericarp) locus.
Methods and Algorithms: We used FastQC to estimate sequencing quality, cutadapt to remove adapter sequences and prinseq to filter sequences by quality and length. Filtered libraries were mapped to H. vulgare genome assembly version 32,608 v 1.33 from the Ensembl Plants database (http://plants.ensembl.org). Mapping was performed with TopHat2 tool. The resulting alignments were processed with the Cufflinks v 2.2.1 pipeline. Pathway analysis was performed using PlantCyc database (http://www.plantcyc.org/). After that selected DEGs were verified using qRT-PCR.
Results: Firstly, we excluded involvement of metabolic pathways, known for other types of pigments (i.e. widespread flavonoid pigments). Our data demonstrated that none of the key flavonoid biosynthesis genes is expressed significantly in black-grained barley during the pigment formation. Then, we used RNAseq approach to reveal DEGs in black-colored and uncolored-grained barley. A total of 957 genome fragments had statistically significant changes in expression levels between lines BLP and BW, with 632 fragments having increased expression levels in line BLP and 325 genome fragments having decreased expression. Genes with high level of differential expression in BLP line were identified in following pathways: suberin monomer biosythesis, diterpene phytoalexins precursors biosynthesis, cutin biosynthesis, cutular wax biosynthesis, and phenylpropanoid biosynthesis, initial reactions.
Conclusion: The number and diversity of metabolic pathways confirms our earlier assumption about the pleiotropic nature of the *Blp* locus. Besides forming a black color of the seeds, *Blp* locus is associated with the resistance to oxidative stress (possibly due to the increased content of antioxidants, including ferulic acid), as well as increased resistance to pathogens (due to the synthesis of phytoalexins and cuticular waxes). The black color of the seeds is presumably associated with the action of the enzyme polyphenol oxidase (PPO).
Acknowledgements: The study was supported by the Russian Science Foundation (No. 16–14-00086).

DOI 10.18699/SBB-2018-13

Ecotoxicity prediction of the new imidazole liquids with a phosphorus containing anion using QSAR

N. Gogoleva[1, 2]*, O. Bernachuk[3], A. Bernachuk [3]
[1] *A.E. Favorsky Irkutsk Institute of Chemistry SB RAS, Irkutsk, Russia*
[2] *Irkutsk Scientific Centrum SB RAS, Irkutsk, Russia*
[3] *Irkutsk State University, Irkutsk, Russia*
** e-mail: googo@mail.ru*

Key words: ionic liquids, imidazolium hypophosphites, ecotoxicity, quantitative structure-activity relationships (QSARs)

Motivation and Aim: The ionic liquids are broadly used in various fields of science and technology. In A.E. Favorsky Irkutsk Institute of Chemistry, Siberian Branch of the Russian Academy of Sciences the approaches to directed synthesis of new functional organophosphorous compounds from elemental phosphorus (using Trofimov-Gusarova reaction) [1] are developed. These approaches successfully develop including using the phenomenon of defect formation in a solid structure of this chemical element [2–4]. Recently, previously unknown hypophosphites were received. These hypophosphites are 1-*n*-, 1-alkyl- and 1-vinyl-3n-imidazolium, which are ionic liquids at room temperature. The aim of our work was to calculate the forecast of ecotoxicity of these new phosphorus-containing ionic liquids.
Methods and Algorithms: The calculations were carried out using quantitative structure-activity relationships (QSARs) and quantitative structure-property relationships (QSPRs) to estimate the physics-chemical and ecotoxic properties of new ionic liquid.
Results: The model shows a group contribution method that considers three main groups of descriptors in the ionic liquid structure: the anion, the cation and the substitutions (carbon chains linked to the cation). Based on these descriptors, their contribution to the ecotoxicity of the ionic liquid has been evaluated by means of a multilinear regression model.
Conclusion: The received results using quantitative structure–activity relationships approach (QSAR) for estimating the ecotoxicity correlate with the literature data for similar compounds [5].

References

1. Abakumov G.A. et al. (2018) Perspective growth-points and challenges of element-organic chemistry. Russ. Chem. Rev. 87(5):93-507.
2. Gusarova N.K. et al. Reaction of vinylpyridines with active modifications of elemental phosphorus in KOH/DMSO (2001) Chem. Heterocycl. Comp. 37(5):576-580.
3. Sukhov B.G. et al. (2003) Radiation defect formation processes as a method for activation of red phosphorus in the Trofimov-Gusarova reaction. Arkivoc. 13:196-204.
4. Trofimov B. et al. (2006) Directed synthesis of phosphines and their derivatives based on elemental phosphorus: design of novel ligands for metal-complex catalists. Catalysis Industry. 4:18-23.
5. Luis P. et al. (2010) Quantitative structure–activity relationships (QSARs) to estimate ionic liquids ecotoxicity EC50 (Vibrio fischeri). J. Mol. Liquids. 152:28-33.

DOI 10.18699/SBB-2018-14

Estimated prognosis of ecological toxicity of new imidazolium ionic liquids with phosphorus-containing anion

N. Gogoleva[1, 2]*, O. Bernachuk[3], A. Bernachuk[3]
[1] *A.E. Favorsky Irkutsk Institute of Chemistry SB RAS, Irkutsk, Russia*
[2] *Irkutsk Scientific Centrum SB RAS, Irkutsk, Russia*
[3] *Irkutsk State University, Irkutsk, Russia*
** e-mail: googo@mail.ru*

Key words: ionic liquids, imidazolium hypophosphites, ecotoxicity, Quantitative structure–activity relationships (QSARs)

Motivation and Aim: Ionic liquids are increasingly used in a wide variety of fields of science and technology. In the Irkutsk Institute of Chemistry. A.E. Favorski SB RAS developed and successfully develops approaches to the directed synthesis of new functional organophosphorus compounds directly from elemental phosphorus (the Trofimov-Gusarova reaction) (see, for example, [1]), including the use of the phenomenon of defect formation in the solid structure of this chemical element [2–4]. Recently, the previously known hypophosphites 1-H-, 1-alkyl- and 1-vinyl-3H-imidazolium, which are ionic liquids at room temperature, have been obtained. The aim of the work was the calculated prediction of the environmental toxicity of these new phosphorus-containing ionic liquids.

Methods and Algorithms: Calculations were carried out using the "quantitative structure-activity relationships (QSARs)" model to evaluate the physico-chemical and environmental-toxicological properties of new ionic liquids.

Results: The model shows a group contribution method that considers three main groups of descriptors in the ionic liquid structure: the anion, the cation and the substitutions (carbon chains linked to the cation). Based on these descriptors, their contribution to the ecotoxicity of the ionic liquid has been evaluated by means of a multilinear regression model.

Conclusion: The results obtained using the QSARs model for assessing the environmental toxicity of new ionic liquids are completely correlated with the literature data for such compounds [5].

References

1. Abakumov G.A. et al. (2018) Perspective growth-points and challenges of element-organic chemistry. Russ. Chem. Rev. 87(5):393-507.
2. Gusarova N.K. et al. Reaction of vinylpyridines with active modifications of elemental phosphorus in KOH/DMSO (2001) Chem. Heterocycl. Comp. 37(5):576-580.
3. Sukhov B.G. et al. (2003) Radiation defect formation processes as a method for activation of red phosphorus in the Trofimov-Gusarova reaction. Arkivoc. 13:196-204.
4. Trofimov B.A. et al. (2006) Directed synthesis of phosphines and their derivatives based on elemental phosphorus: design of novel ligands for metal-complex catalists. Catalysis Industry. 4:18-23.
5. Luis P. et al. (2010) Quantitative structure-activity relationships (QSARs) to estimate ionic liquids ecotoxicity EC50 (Vibrio fischeri). J. Mol. Liquids. 152:28-33.

DOI 10.18699/SBB-2018-15

A database and analytical platform for mining billions of genetic associations

D.D. Gorev[1, 2, 3]*, T.I. Shashkova[1, 2, 3], E. Pakhomov[2], A. Torgasheva[4], L. Klaric[5, 6], A. Severinov [1, 3], S. Sharapov [2, 4], D.G. Alexeev[1, 2], Y.S. Aulchenko[2, 3, 4]

[1] *LLC 'Knomics', Moscow, Russia*

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

[3] *Moscow Institute of Physics and Technology, Moscow, Russia*

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

[5] *Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK*

[6] *Genos Glycoscience Laboratory, Zagreb, Croatia*

** e-mail: gorev.d@phystech.edu*

Key words: genome-wide association study, snp, database, web-service

Motivation and Aim: Hundreds of genome-wide association studies (GWAS) of human traits are performed each year, and are, together with results from tens of thousands of previously reported GWAS, freely available. These results are published in the form of summary statistics (where for each SNP the allelic frequencies, estimates of the coefficients of regression and their standard errors are typically reported). This information can be used for multiple purposes – from research in fundamental biology and genetics, to biomarker and therapeutic intervention of targets discovery. While the amount of information accumulated by the scientific community is very large, the use of this valuable information is restricted by lack of reporting guidelines and facilities that would allow for quality control (QC), storage, and analysis of such data. This situation forces researchers to spend a lot of time and efforts on data collection, data preprocessing to accommodate different analytical tools, and QC. In this work, we designed a platform for storage, QC, and analysis of GWAS summary statistics.

Results: The original data harmonization algorithm was developed to effectively store and quickly access GWAS data. For data storage and manipulations on our platform we use two database management systems, ClickHouse, to store harmonized GWAS results and PostgreSQL for meta-data storage. Clickhouse provides us with rapid-access storage accessible via powerful and flexible SQL interface. The platform implements several GWAS QC algorithms. It also embeds several methods often used for analysis of GWAS summary statistics, such as LDsr and MRbase libraries that facilitate genetic correlations and mendelian randomization analyses, respectively, and our own implementation of the summary data-based mendelian randomization and heterogeneity in dependent instruments (SMR-HEIDI) testing that allows for analysis of pleiotropy. As a proof of concept, 429 GWASs, totaling to ~3 billion of SNP-trait associations, were uploaded to the platform. On average, the selection of all SNPs in a 500 kbp range from all GWAS in the database to a Python Dataframe takes ~17 seconds, while selection of a specific GWAS ~34 seconds.

Conclusion: We have developed a platform for storage, quality control and analysis of summary GWAS data. The platform is capable of storage and high throughput retrieval and analysis of results of thousands of GWAS.

Availability: GWAS-MAP will soon be available as a web-service platform.

DOI 10.18699/SBB-2018-16

Overexpression of *Gclc* in the *Drosophila melanogaster* thorax

Z. Guvatova[1]*, M. Shaposhnikov[2, 4], E. Lashmanova[3], G. Krasnov[1],
A. Kudryavtseva[1], A. Moskalev[1, 2, 3, 4]
[1] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*
[2] *Institute of Biology of Komi Science Center of Ural Branch of RAS, Syktyvkar, Russia*
[3] *Moscow Institute of Physics and Technology, Dolgoprudny, Russia*
[4] *Syktyvkar State University, Syktyvkar, Russia*
** e-mail: guvatova.zulfiya@mail.ru*

Key words: *Drosophila melanogaster*, gene expression; glutathione, lifespan

Motivation and Aim: The enhancement of glutathione biosynthetic capability can determine longevity and delay aging. Our recent studies demonstrated that *Gclc* overexpression slows down the age-dependent decline of locomotor activity and circadian rhythmicity without effect on fecundity [1]. Here we analyzed the effects of neuronal *Gclc* overexpression in the thorax of *Drosophila melanogaster* on the transcriptomic changes.
Methods and Algorithms: Transcriptomic analysis was performed using control UAS-Gclc flies and flies with *Gclc* overexpression at the age of 1 (young), 4 (matured) and 6 weeks (old). Processing of transcriptomic data was performed using PPLine toolkit [2] including read preprocessing (trimmomatic), mapping (STAR) and counting (HTSeq-count). The further analysis was done with R programming language (R core Team). The edgeR package was used for analysis of differential expression [3]. KEGG gene set enrichment analysis (GSEA) was performed using clusterProfiler [4].
Results: We derived RNA sequencing expression profiles for 12000 genes (after eliminating low expression ones). The expression of 760 genes (108 of 760 have FDR < 0.05) demonstrated association with *Gclc* overexpression in all of the groups: young/mature/old or males/females. When the selected threshold of expression was 2-fold or more (FDR < 0.05), *Gclc* overexpression down-regulated 42 genes and up-regulated 14 genes, such as *SMC2* (*Structural maintenance of chromosomes 2*), *w* (*white*), *CG4293*, *Gclc*, *Cyp4p2* (*cytochrome P450 4p2*), *Ipk1* (*Inositol-pentakisphosphate 2-kinase*), *CG8157*, *CR45457*. The *Gclc* level demonstrated associations with expression of genes involved in a variety of cellular processes.
Conclusion: Transcriptome analysis of the thorax of Gclc transgenic flies revealed pathways that may contribute to the longevity and prevent the age-dependent decline of biological functions.
Acknowledgements: This work costs were funded by the Russian Science Foundation grant No. 14-50-00060.

References

1. Moskalev A.A. et al. (2016) The influence of pro-longevity gene Gclc overexpression on the age-dependent changes in Drosophila transcriptome and biological functions. BMC Genom. 17:273-289.
2. Krasnov G.S. et al. (2015) PPLine: An Automated Pipeline for SNP, SAP, and Splice Variant Detection in the Context of Proteogenomics. J Proteome Res. 14:3729-3737.
3. Robinson M.D., McCarthy D.J., Smyth G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26:139-140.
4. Yu G., Wang L.G., Han Y., He Q.Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 16:284-287.

DOI 10.18699/SBB-2018-17

Adaptation of the brain tumors classifier, constructed on the direct mass spectrometric profiling data, to MALDI mass spectrometry analysis

D. Ivanov[1]*, A. Sorokin[1], E. Zhvansky[1], V. Shurkhay[2], S. Pekov[1, 3], I. Popov[1, 3], E. Nikolaev[1, 3, 4]

[1] *Moscow Institute of Physics and Technology* (*State University*), *Dolgoprudny, Russia*
[2] *N.N. Burdenko Scientific Research Neurosurgery Institute, Moscow, Russia*
[3] *Institute of Energy Problems of Chemical Physics of RAS, Moscow, Russia*
[4] *Skolkovo Institute of Science and Technology, Skolkovo, Moscow region, Russia*
** e-mail: daniil.ivanov@phystech.edu*

Key words: mass-spectrometry, lipidomics

Motivation and Aim: The localization of the exact boundaries of malignant brain neoplasms is an important problem of the modern neurosurgery. One of the promising approaches to intraoperative tumor identification is the determination of the type of tissue based on the information, achieved by molecular profiling of brain tissues. Various mass spectrometric approaches can be used to determine feature ions for tumor tissues: for intraoperative analysis the fastest method is the application of direct ionization methods. In contrast, MALDI mass spectrometry allows not only to obtain molecular profiles, but also to analyze the structure of tissue with a high spatial resolution (less than 100 μm). This method has the ability to identify the different types of tissues in the sample, determine their relative position, relative sizes and the structure of the boundaries between them.

Previously, to complete the database of molecular profiles of neoplasms in the human brain, we developed a method for online extraction of lipids from brain tissues and their subsequent profiling. We constructed a classifier that allows determining the type and degree of malignancy of the tumor according to its molecular profile [1]. However, this classifier was teached to work precisely only with the data received as a result of analyzing the lipid fraction using direct electrospray ionization mass spectrometry.

The purpose of this study is to adjust the existing classifier to analyze spatial molecular profiles obtained by MALDI MS.

Methods and Algorithms: To estimate similarity between the profiles obtained by online extraction method and MALDI MS, a correlation analysis of two sets of molecular profiles for each series of samples was conducted. After optimization of the existing classifier parameters, the samples were classified according to MALDI profiles, and then compared with analyze of profiles obtained during online extraction

Results: We have shown that the classifier, that was teached on the online extraction data, is applicable to molecular profiles, obtained by MALDI mass spectrometry. The analysis of the main features, used in the classifier, allowed optimizing the parameters of mass spectrometric experiment to improve efficiency classification and visualization of the spatial structure of the samples.

Acknowledgements: This work was supported by the RCSF (16-15-10431).

References

1. Sorokin A. et al. (2017) Feature selection algorithm for spray-from-tissue mass spectrometry. European Journal Mass Spectrometry. 23(4):237-241.

DOI 10.18699/SBB-2018-18

A study of microbial spectra of soils potentially relevant for the formation of stable soil foci of anthrax

A.V. Kalinin*, O.I. Tsygankova, E.A. Koteneva, A.V. Abramovich, G.A. Pechkovsky
Stavropol Antiplague Institute of the Rospotrebnadzor, Stavropol, Russia
** e-mail: jugask@mail.ru*

Key words: bacilli, *Bacillus anthracis*, phenotypic variability, heterogeneity of population, pathogenic properties, morphofunctional forms

Motivation and Aim: The members of the genus *Bacillus* constitute a group of widespread microorganisms which are being isolated from the soil. This heterogeneous taxonomic group of spore-forming bacteria contains more than 65 species with new species being discovered constantly [1]. Under favorable conditions of the environment the physiological and metabolic universality of bacilli ensures fast germination of spores [2]. Basically, the genus *Bacillus* consists of saprophytic organisms, however the greatest interest of medical microbiology is focused on the taxonomic group *Bacillus cereus* (sensu lato), since it comprises three species causing diseases in mammals: *Bacillus anthracis*, *Bacillus* cereus and *Bacillus thuringiensis*. *B. anthracis* is the etiological agent of anthrax. Many mammals and some birds are sensitive to this agent, but basically this species affects herbivores. After the death of the infected animal the process of spore formation begins in organic remains of the dead animal in the presence of oxygen. The soil in the places of death and burial of animals serves as a primary source for further infections during many tens of years. An assumption has been made that in "the areas of incubation" in soils rich with mineral and organic substances with pH above 6.0 and temperature above 15 °C which topographically correlate with the location of areas in which infections of animals with *B. anthracis* take place from time to time, *B. anthracis* can not only persist during the interepizootic period, but also propagate [1]. It is known that *B. cereus* can exist in the root zone of plants, and it is necessary to take into account the probability of germination of *B. anthracis* spores in the rhizosphere and propagation of vegetative cells of *B. anthracis* in this ecological niche. It has been confirmed on modelling organisms that viable bacilli of anthrax can exist in the rhizosphere [2]. Recent ecological studies have shown that anthrax bacilli can interact with earthworms, soil amoebas and bacteriophages. It has been revealed that the influence of bacteriophages on *B. anthracis* can cause phenotypic changes in bacilli; these changes make them capable of endosymbiont existence in the organisms of earthworms and of functioning as saprophytes in the soil and water. Hence, it may be assumed that amoebas and probably other soil protozoa can serve as potential host organisms for anthrax bacilli, and the life cycle of this bacteria is not limited by the standard paradigm of the development of anthrax. The major factors of counteraction to protective mechanisms of the macro organism during the infectious phase – the polyglutamic capsule and exotoxins – cannot ensure persistence and viability of the culture which comes into the environment. At this stage strains capable of forming the high-grade spores, highly resistant to adverse physical, chemical and biological factors of the environment, more often of soil, gain advantages.

Conclusion: Inhomogeneity of populations of strains of the causative agent of anthrax with respect to many properties in the absence of selective factors operating in organisms of sensitive animals may lead to the appearance and accumulation of strain subcultures which lost pathogenic properties or which possess steady complexes of atypical phenotypic and genetic properties. The estimation of roles of various soil foci of anthrax in preservation of the reservoir of the infection may be carried out taking into account studies of population changes in *B. anthracis* strains in case of their long persistence in soil and preservation of the potential of pathogenicity of such "soil" populations.

References

1. Hugh-Jones M. (2009) The ecology of Bacillus anthracis. J. Mol Aspects Med. 30(6):356-67.
2. Mullins J.C. et al. (2013) Ecological niche modeling of *Bacillus anthracis* on three continents: Evidence for genetic-ecological divergence? PLoS ONE. 8(8):e72451.
3. Bazinet A.L. (2017) Pan-genome and phylogeny of *Bacillus cereus* sensu lato. BMC Evol Biol. 17:176.

DOI 10.18699/SBB-2018-19

Differences in association of the genes with cognitive function and symptom severity of Belarusian schizophrenia patients

H. Kandratsenkan[1]*, A. Nestsiarovich[2], I. Goloenko[1], N. Danilenko[1], A. Makarevich[1], V. Obyedkov[2], O. Davydenko[1]

[1] *Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Belarus*

[2] *Republican Research and Practice Center for Mental Health, Minsk, Belarus*

[*] *e-mail: amber070790@gmail.com*

Key words: WCST, Stroop, PANSS, schizophrenia, *MIR137*, *COMT*, *DRD2*, *MTHFR*, *DNMT3B*

Motivation and Aim: Cognitive and symptom impairments of schizophrenia patients are the base of their individual disability. Patients are treated using antipsychotic therapy but in majority of the cases the prescription process is conducted in trial and error way. The analysis of the genes that may be involved in cognitive and symptomatic changes in patients could result into better schizophrenia therapy progress.

Methods and Algorithms: Our sample consisted of 150 Belarusian individuals diagnosed with schizophrenia (age 46.8 ± 9.4; 54 % females) assessed using the PANSS (symptom severity), the WCST and Stroop test (cognitive functioning). The *MIR137* rs1625579, *COMT* rs4680, *DRD2* rs1800497, *MTHFR* rs1801133, *DNMT3B* rs2424913 polymorphisms of schizophrenia patients were analyzed using RFLP and TaqMan assay. Our outcome of interest was the association of gene polymorphisms with symptom severity and cognitive parameters analyzed using univariate ANOVAs. All statistical analyses were conducted in R v.4.3.2.

Results: *MTHFR* rs1801133 was associated with positive symptom severity ($p = 0.02$), *COMT* rs4680 ($p = 0.03$) and *MIR137* rs1625579 ($p = 0.03$) – with general symptom severity for both males and females. *MIR137* rs1625579 ($p = 0.02$) and *DNMT3b* rs2424913 ($p = 0.002$) were associated with negative and general symptom severity respectively for females only. Overall, the number of minor alleles across the five SNPs of interest was correlated with negative symptom severity ($r = 0.20$, $p = 0.09$) only for males. Furthermore, there were sex-specific differences in the combined *DNMT3B C/C* genotype and *COMT* G-allele ($p = 0.0001$). Males with the *DNMT3B C/C* genotype and *COMT* G-allele (N = 13) showed greater negative symptom severity (M = 27.2 ± 6.7) in comparison to female patients with the same genotypes (N = 18, M = 15.6 ± 5.6). For cognitive function, we found the association of *COMT* rs4680 with a number of variables of WCST (categories completed, nonperseverative errors, total errors) and Stroop test (total errors) but mostly for females only. Furthermore, female patients with A2/A2 (*DRD2/ANKK1*) and A-allele **(***COMT)* demonstrated a lower number (2.97) of WCST categories completed compared to A1-allele (*DRD2/ANKK1*) and A-allele (*COMT*) (4.94; $p = 0.03$). Females with T-allele (*MTHFR*) and T-allele (*DNMT3B*) have made twice as many Stroop test total errors in comparison with C/C (*MTHFR*) and T-allele (*DNMT3B*) ($p = 0.08$).

Conclusion: We found that genetic predictors involved in symptom severity and cognitive functioning of schizophrenia patients may have an impact on psychopathology of the disease in a different way for males and females.

Acknowledgements: Supported by the BRFFI (No. M13M-065), by Presidium of NASB (No. 2015-28-069).

DOI 10.18699/SBB-2018-20

Human blood proteins and correlations with biochemical parameters after long duration space flights

D.N. Kashirina[1]*, L.Kh. Pastushkova[1], A.M. Nosovsky[1], A.J. Percy[2], K.S. Kireev[3], C.H. Borchers[2], E.N. Nikolaev[4], I.M. Larina[1]

[1] *Institute for Biomedical Problems – Russian Federation State Scientific Research Center RAS, Moscow, Russia*

[2] *Genome British Columbia Proteomics Centre, University of Victoria, Victoria, BC, Canada*

[3] *Yu.A. Gagarin Research and Test Cosmonaut Training Center, Star City, Moscow Region, Russia*

[4] *Skolkovo Institute of Science and Technology, Skoltech, Moscow region, Russia*

** e-mail: daryakudryavtseva@mail.ru*

Key words: blood plasma, cosmonauts, mass spectrometry, biochemical parameters

Motivation and Aim: The peculiar features of space radiation and microgravity pose a risk to the cardiovascular system (CVS) and add to the risks of manned space exploration. Accelerated 'ageing-like' changes develop in the CVS during spaceflight. Over a 6-month stay on the ISS, vascular and cardiac structural and functional modifications occur, including increased vascular wall thickness and stiffness, which might predispose astronauts to atherogenesis [1]. However the molecular mechanisms driving physiological changes remain unknown.

Methods and Algorithms: Blood samples from 18 Russian cosmonauts who had conducted long-duration space flights were analyzed. The samples were collected 30 days before launch, on the first and seventh days after landing. Biochemical parameters of blood were measured at the same time points. Using mass spectrometry approach involving multiple reaction monitoring in conjunction with stable isotope-labeled standards at the University of Victoria, 125 plasma proteins were detected and quantitated. To determine the dependencies between proteins and biochemical parameters of blood, a correlation was made. To increase the accuracy of the analysis, the group of cosmonauts was divided into 2 homogeneous groups, correlations were determined in each group.

Results: It was revealed that in both groups there was a correlation between total cholesterol and apolipoprotein B-100, LDL and cystacin-C and apolipoprotein E, triglycerides and coagulation factor VII and apolipoprotein C-II. These proteins could be a potential risk factors for cardiovascular disease. It was found that paraoxonase-1 correlated positively with HDL and apolipoprotein A-I00, and negatively with the coefficient of atherogenicity. Therefore this protein could have a protective effect on blood vessels.

Conclusion: The correlations found are partially confirmed by other authors and methods, and in part – it allow to construct hypotheses about the unexplored mechanisms of cardiovascular dysfunction after space flight.

Acknowledgements: The work was supported by RFBR grant No. 18-34-00524. We express special thanks to Bruker Daltonics for financial and technical support.

References

1. Hughson R.L., Helm A., Durante M. (2018) Heart in space: effect of the extraterrestrial environment on the cardiovascular system. Nat Rev Cardiol. 15(3):167-180. DOI 10.1038/nrcardio.2017.157.

DOI 10.18699/SBB-2018-21

The variability of the amylose / amylopectin ratio and the preparative yield of *Solanum tuberosum* tuber starch

Y. Khoroshavin[1, 2]*, V. Khlestkin[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: khoroshavin@bionet.nsc.ru*

Key words: potato, starch, physical and chemical properties, DNA markers

Motivation and Aim: The optimal set of physicochemical properties of starch varies significantly for different applications and depends on the content and structure of its constituent polysaccharides: amylose and amylopectin. The content and structure of these molecules are regulated by the genes of their biosynthesis and can be considered as phenotypic features, according to which potato plants can be selected. Today many of these genes already known, but the connection of their allelic variants, loci with genes not yet described, and the physics-chemical characteristics of starch requires a deeper study. The purpose of this paper is to analyze the variability of amylose / amylopectin ratio as well as the content of potato starch and locating the «locus-feature» associations, which will subsequently allow to identify the markers of the corresponding genes.
Methods and Algorithms: The study was carried out on tubers of 90 varieties *S. tuberosum* from the collection "GenAgro", mainly blighty selection. Starch was isolated according to the procedure [1] and its preparative yield was calculated. To determine the amylose / amylopectin ratio, a new technique developed by us was used. Also DNA was isolated from the studied tubers and sent for genotyping to SNP chip of Illumina company [2]. Further, an analysis will be made of data on genetic variability of varieties and the finding of "locus-features" associations.
Results: We developed a new technique for measuring the ratio of amylose / amylopectin and obtained data for several dozen varieties of potato. Contrastive varieties (from 14 to 34 %) on the content of amylose in the starch the was detected. Data on the yield of starch for 90 varieties were also obtained and contrastive varieties (from 7 to 21 %) for this feature was found.
Conclusion: The data obtained show a significant variability in the features of tuber starch, and, consequently, the possibility of breeding on them. These data will be used to find the DNA markers of the relevant features.
Acknowledgements: Supported by the RFBR (17-29-08006).

References

1. Khlestkin V.K., Erst T.V. (2017) A practical guide to the starch granules' morphology study by microscopy. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 21(6):728-734. DOI 10.18699/VJ17.290 (in Russian).
2. Vos P.G. et al. (2015) Development and analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding history. Theoretical Applied Genetics. 128(12):2387-2401. DOI 10.1007/s00122-015-2593-y.

DOI 10.18699/SBB-2018-22

Computer analysis of transcriptomes in glioma cells

S.S. Kovalev[1]*, N.V. Gubanova[2], A.V. Tsukanov[2], A.O. Bragin[2], Y.L. Orlov[2, 3]
[1] *Novosibirsk State Medical University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: sergey.kovalev.1994@list.ru*

Key words: cancer, transcriptomics, database, gene expression regulation, splicing, glioma

Motivation and Aim: Glioma is the most common malignant tumor in the central nervous system. Here we consider problem of detection of genes responsible for glioma progression in primary cell cultures. The identification of sensitive and specific biological markers that would help identify patients at a higher or lower risk of death from glioma is of crucial importance [1]. We revealed set of differently expressed gene in normal brain and glioma cell cultures and analyzed alternative splicing variants.
Methods and Algorithms: The primary cell culture samples from normal brain and secondary glioblastoma were processed for RNA extraction. This was followed by RNA-sequencing, data processing and filtration of reads. The algorithm of the pipeline consists of several stages: 1) pre-processing of the reads; 2) mapping reads to the reference genome; 3) identification of differential expression and cases of alternative splicing; 4) annotation of obtained results. For assessment of gene expression level and finding differently expressed genes we used Cufflinks. Set of computer tools were used for sequencing data processing and alternative splicing analysis [2].
Results: We found set of hormone transporter genes overexpressed in the glioblastoma cell culture. SLCO1C1 mediates the Na+-independent high affinity transport of organic anions such as the thyroid hormones thyroxine. Multiple alternative splicing transcripts have been identified as progression markers. We found set of differentially alternatively spliced transcripts. 73 of the differentially expressed genes were found in OMIM as related to glioma (of 193 loci) including GLI1, GLI3 (GLI-Kruppel family member 3); GLM4 (GLIOMA SUSCEPTIBILITY 4), GLTSCR1 (Glioma tumor suppressor candidate region gene 1) and others. We have developed database of genes with alternative splicing in glioma including APP, CASC4 and TP53.
Conclusion: The RNA-seq analysis of the cells cultures of normal brain and glioma confirmed association of these genes with tumor progression. The results provide an experimental basis for the observation that hypothyroidism induction by administration of propylthiouracil is associated with improved survival in glioblastoma patients. We work on integration of data on glioma genes using available data sources developing own database prototype.
Acknowledgements: The research has been supported by ICG SB RAS budget project (0324-2018-0021). Computing at Siberian Supercomputer center SB RAS (SSCC) was done in frames of the budget project 0324-2018-0017.

References

1. Liu M. et al. (2017) The identification of key genes and pathways in glioma by bioinformatics analysis. Journal Immunology Research. Article ID 1278081.
2. Babenko V.N. et al. (2017) Computer analysis of glioma transcriptome profiling: alternative splicing events. J Integr Bioinform. 14(3). pii: /j/jib.2017.14.issue-3/jib-2017-0022/jib-2017-0022.xml.

DOI 10.18699/SBB-2018-23

NGS data processing method for the mixture of chloroplast and mitochondrial DNA of barley

A. Makarevich*, O. Pankratov, M. Sinyavskaya, N. Lukhanina, A. Shymkevich, A. Liaudansky, I. Goloenko, N. Danilenko, O. Davydenko
Institute of Genetics and Cytology, National Academy of Sciences of Belarus, Minsk, Belarus
** e-mail: bio.makarevich@gmail.com*

Key words: NGS, data processing, chloroplast genome, mitochondrial genome, barley phylogeny

Motivation and Aim: Phylogenetic connections between cultivated barley and its wild forms are insufficiently explored field in evolution genetics [1]. Chloroplast and mitochondrial genome sequencing was performed for wild and domesticated barley lines. The fraction of chloroplasts obtained by differential centrifugation was used for DNA extraction. Such a DNA sample contains an admixture of mitochondrial DNA, which allows studying both chloroplast and mitochondrial genomes. However, the genome assembly is complicated by presence of extended areas of homology within chloroplast and mitochondrial genomes as well as between them. Thus, the aim of this research was to develop the NGS-data processing method, which allows assembling complete sequences of each genome from the mixture of chloroplast and mitochondrial DNA.
Methods and Algorithms: The algorithm of sequencing data processing included the following steps: trimming of row reads (Trimmomatic); aligning reads to the "double" reference, containing full sequences of chloroplast and mitochondrial barley genomes (Bowtie2); obtaining mapping statistics (bash scripts, BCFtools); alignment visualization (Tablet); generating VCF files (Samtools); filtering VCF files (VCFlib). The algorithm was tested on artificial Illumina reads synthesized using the ART program. Ultimately, VCF files containing all polymorphic loci of the chloroplast and mitochondrial genomes were obtained.
Results: Testing of the developed algorithm showed its applicability for the mixtures of barley chloroplast and mitochondrial DNA with different ratios of concentrations.
Conclusion: The developed algorithm allows obtaining complete sequences of chloroplast and mitochondrial barley genomes from the mixtures of plastid and mitochondrial DNA using the next-generation sequencing data. Modifications of this method can be used to study the organelle genomes of other cereals.
Acknowledgements: The study was conducted within the framework of the State Program of Scientific Research "Biotechnologies 2016–2020".

References

1. Pankin A., von Korff M. (2017) Co-evolution of methods and thoughts in cereal domestication studies: a tale of barley (*Hordeum vulgare*). Current Opinion Plant Biology. 36:15-21.

DOI 10.18699/SBB-2018-24

The metabolic pathways, metabolites and signal transduction

C.K. Mitra
Department of Biochemistry, University of Hyderabad, 500046 Hyderabad, India
e-mail:ck.mitra@gmail.com

Key words: dynamics, metabolites, pathways, enzyme kinetics, quasi-steady state

Motivation and Aim: Although the metabolic pathways have been extensively studied we only have a static picture of the complete system. It is also true that we know reasonably well the acceptable concentration limits (physiologically meaning) of a large number of metabolites, but we do not know the dynamics, i.e., the fluxes (including sources and sinks) of most of the important metabolites. For a living system, the fluxes are more relevant compared to their static concentrations. Most of the metabolite concentrations must stay within a well-defined limit (safe volume in the multidimensional hyperspace described by the metabolite concentrations).

All metabolic fluxes are products of biochemical reactions that are catalyzed by enzymes. Practically all biochemical reactions are catalyzed by enzymes and each enzyme is characterized by at least two parameters determined by the substrate. The two most common parameters, V_{max} and K_m, are usually defined in terms of the Michealis-Menten rate equation (single substrate, single intermediate enzyme kinetics). For multiple substrate reactions, additional K_m must be defined. For many enzymes, there are regulatory sites and additional information on the binding affinity of the regulatory ligand need to be known. Therefore it is not uncommon to describe an enzyme by a set of six or more parameters [1].

Methods and Algorithms: We use an approximate method that we have earlier developed for the study of reaction kinetics [2]. In brief, we express the substrate concentration in units of the K_m, therefore reducing the substrate concentration to a dimensionless number. In all enzyme catalyzed reactions, we observe a limiting rate (rate reaching a maximum value) with increase in the substrate concentration and we therefore replace [S] with [S]/(1+[S]) in the kinetic equations. Other allowances are made to conform known biochemical properties. Regulators (inhibitory) are modeled as 1/(1+[I]) where I the inhibitor concentration expressed in terms of K_I where K_I is the dissociation constant for the enzyme-inhibitor complex. Finally the set of first order differential equations were set up and solved using Octave [3].

Results: The results obtained from the software are presented graphically. As an example, we have simulated the common symport and antiport (ion transporters) widely found on membrane surfaces that preserve charge balance.

Conclusion: The results would have been more interesting if we could include the membrane potential also in the overall fluxes.

References

1. Enzyme Database – BRENDA https://www.brenda-enzymes.org/; see also: Schomburg I., Chang A., Schomburg D. (2002) BRENDA, enzyme data and metabolic information. Nucleic Acids Research. 30(1):47-49.
2. Korla K., Mitra C.K. (2014) Kinetic modelling of coupled transport across biological membranes. Indian Journal Biochem Biophys. 51(2):93-99.
3. GNU Octave: A scientific programming language. https://www.gnu.org/software/octave/; version used in the current work is 4.4.0.

DOI 10.18699/SBB-2018-25

Association of *MDR1* gene C3435T (rs1045642) polymorphism with colorectal cancer in the population of Central Russia

A.S. Moskalev, E.M. Barysheva, O.Yu. Bushueva*
Kursk State Medical University, Kursk, Russia
e-mail: olga.bushueva@inbox.ru

Key words: colorectal cancer, *MDR1* gene, SNP

Motivation and Aim: Colorectal cancer (CRC) is one of the most common cancers worldwide. Currently, the incidence of CRC tends to increase. Genetic factors together with environmental factors might increase the risk of colorectal cancer [1]. It is well known that genes encoding drugs metabolism enzymes are involved in the pathogenesis of colorectal cancer. The multidrug resistance gene 1 (MDR1, ATP binding cassette subfamily B member 1, ABCB1) encodes a transmembrane glycoprotein P, involved in the transformation of drugs. Recent studies indicated MDR1 gene seemed to play an important role in tumor progression, especially in the colorectal carcinogenesis [2].
Methods and Algorithms: A total 379 unrelated Russian subjects including 244 CRC patients and 135 age- and sex-matched controls were recruited for this study. Genomic DNA was isolated from peripheral blood samples using a standard phenol/chloroform procedure. Genotyping of the 3435C>T polymorphism (rs1045642) of the *MDR1* gene was done using TaqMan-based assay on the CFX96 real-time PCR Detection System. Hardy-Weinberg equilibrium was tested to compare the observed and expected genotype frequencies among cases and controls using chi-square test. The association between the polymorphism and CRC risk was estimated by odds ratio (OR) with 95 % confidence interval (CI). The statistical significance was established at $P<0,05$. Statistical calculations were performed with Statistica for Windows 8.0.
Results: The MDR1 genotype frequencies were in agreement with Hardy-Weinberg equilibrium in CRC and control groups ($P >0,05$), which indicates uniformity of the sample. The frequency of homozygous genotype 3435CC was 22,1 % in patients and 33,3 % in healthy controls. The frequency of heterozygous genotype 3435CT was 50,4 and 47,4 % in patients and controls, respectively. Homozygous genotype 3435TT was observed in 27,5 % of cancer patients and in 19,3 % of healthy individuals. Thus, we have established that homozygous genotype 3435CC of the *MDR1* was associated with protective effect against the risk of CRC (OR = 0,57; 95%CI = 0,36–0,91; P = 0,02).
Conclusion: In conclusion, our study showed that MDR1 C3435T polymorphism might be significantly associated with risk of CRC in the population of Central Russia. Further studies, especially the gene–gene and gene–environment interactions are required.

References

1. Grady W.M., Carethers J.M. (2008). Genomic and epigenetic instability in colorectal cancer pathogenesis. Gastroenterol. 135(4):1079-1099.
2. Chin J.E., Soffir R.E.N.A., Noonan K.E., Choi K. Y. U. N. G. H. E. E., Roninson I.B. (1989). Structure and expression of the human MDR (P-glycoprotein) gene family. Mol. Cell. Biol. 9(9):3808-3820.

DOI 10.18699/SBB-2018-26

Antibacterial effect of nanocomposites on the basis of various humic a veshchestvna the activator of ring rot of potatoes *Clavibacter michiganensis* ssp. *sepedonicus*

O.A. Nozhkina[1]*, A.I. Perfilyeva[1], I.A. Graskova[1], B.G. Sukhov[2]
[1] *Siberian Institute of physiology and biochemistry of plants of the SB RAS, Irkutsk, Russia*
[2] *Irkutsk Institute of chemistry of A.E. Favorskii of the SB RAS, Irkutsk, Russia*
** e-mail: alla.light@mail.ru, smallolga@mail.ru*

Key words: silver nanocomposites, ring rot, humic substances, activity of peroxidase, bacteriostatic effect

Motivation and Task: The bacterial disease ring decay of potatoes caused by a bacterium of *Clavibacter michiganensis* SSP. *Sepedonicus* (*Cms*), has no ecologically safe measures of fight today. Bacteria extend on potatoes xylem vessels, form a biofilm, break water transport in the top part of a stalk, chlorosis leaves. It leads to a zavyadeniye – a typical symptom of a disease. Bacteria have ability to undergo winter conditions of the environment that gives them the chance to be transferred from generation to generation. All known means of protection come down to disinfecting by toxic substances. Need of development of ecologically safe means of fight against ring decay led to use of nanocomposites of silver based on the humic substances (HS) of various nature. HS have no toxicity, are steady, stimulate growth of plants and a pochvoobrazuyushchy biota, dyes, corrosion inhibitor, etc.
Methods and Algorithms: in a research were used As-1405 strain bacteria (it is received from the All-Russian collection of microorganisms, IBFM RAS), three types of the NC of silver packed into humid substances and their predecessors (NCHS – dirt/$AgNO_3$, NCHS – coals/$AgNO_3$, NCHS – slates/$AgNO_3$, HS – dirt, HS – coals, HS – slates and nitrate of silver) and potatoes plants which are grown up by grade *in vitro* Lukyanovskii. All substances were synthesized at the Irkutsk institute of chemistry by it A.E. Favorskii, they are readily soluble in water and their aqueous solutions are convenient to use. Studying of bactericidal effect of the NC and biofilm formation carried out by measurement of optical density of bacterial suspension. Plants incudated the NC and its predecessors, measured biometric indicators – a gain of plants, quantity of leaves, mass of roots and a land part, activity of peroxidase.
Results and Discussion: the bacteriostatic effect was shown by the predecessor of $AgNO_3$ NC which considerably suppressed growth of bacteria from the very beginning of the experiment. GV differently influenced the studied indicator. So, HS – cl and HS – sl and their NC inhibited a gain of bacterial suspension and ability of a biofilm-formation in comparison with control, and HS – dt and its NC, on the contrary, stimulated reproduction of bacteria, and here reduced formation of biofilms. HS – cl stimulated biofilm formation of a bacterium of *Cms*. After processing of plants of the NC and their predecessors, analyzed change of activity of protective enzyme – peroxidase in potatoes fabrics and a gain of plants. It was revealed that $AgNO_3$, HS – dt, HS – sl, NCHS – dt/$AgNO_3$, NCHS – sl/$AgNO_3$ more than twice reduced activity of peroxidase. HS – cl and NCHS – cl/$AgNO_3$ rendered on activity of peroxidase of potatoes, as well as on biofilm formation of bacteria of *Cms* the stimulating influence. The received results demonstrate presence at some of the studied substances (NCHS – dt/$AgNO_3$, NCHS – cl/$AgNO_3$) the bacteriostatic of properties and effect. Positive influence on plants characterized HS - sl and NCHS – sl/$AgNO_3$. Also, did not render the expressed negative effect on potatoes NCHS – dt/$AgNO_3$. The submitted data confirm a possibility of use of the NC with silver nanoparticles for processing of cultural plants against bacterial diseases.
Conclusion: studying of influence of the NC based on humic substances of various nature on the activator of ring decay of plants of *Clavibacter michiganensis* ssp *sepedonicus* potatoes. Showed, some NC and their predecessors can be the agents having the greatest antibacterial activity and not influencing potatoes plants that gives the grounds to judge their applicability as improving means against an activator strain without doing at the same time harm. All substances are available and easily applicable in use. The received result about influence of the NC on biofilm formation of *Cms* gives to us the grounds to use the NC as means of fight with phytopathogenic.
Acknowledgements: work is supported by a grant of the Russian Federal Property Fund and the Government of the Irkutsk region (project No. 17-416-380001).

DOI 10.18699/SBB-2018-27

Physicochemical properties of *vlhA* promoters in *Mycoplasma gallisepticum* and their possible regulatory role

M. Orlov[1]*, I. Garanina[2], A. Ryasik[1], E. Zykova[1], G. Fisunov[2], A. Sorokin[1]
[1] *Institute of Cell Biophysics SB RAS, Pushchino, Russia*
[2] *Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia*
** e-mail: orlovmikhailanat@gmail.com*

Key words: mycoplasma, promoter, SIDD

Motivation and Aim: *Mycoplasma gallisepticum* is the pathogen affecting respiratory tract in poultry, thus causing significant economic losses to the agricultural industry. The intracellular parasite has a minimalistic genome with a reduced repertoire of transcription factors. Consequently, it is considered unable to regulate transcription by conventional mechanisms yet capable of adjusting gene expression in response to external stimuli in a highly dynamic manner with multiple regulators involved. Of particular interest here are phase variation antigens encoded by members of the variable lipoprotein and hemagglutinin (*vlhA*) gene family that to be involved in bacteria-host interaction and pathogenesis [1]. Here we report research on *M. gallisepticum* promoter regions physics considering its possible role in regulatory mechanisms.
Methods and Algorithms: For promoters from various *M. gallisepticum* strains (543 *vlhA* promoters and 370 others) primary structure as well as several physicochemical and structural properties were assessed. This includes sequence logos and sliding window GC-content analysis as well as electrostatic potential, bendability, Stress-Induced Duplex Destabilization (SIDD), and open states activation energy profiles evaluation.
Results: Sequence analysis of *vlhA* promoters has confirmed the presence of well-established GAA-repeats downstream of transcription start site and overall highly similar sequence context spanning for approximately 200 nt around transcription start site. Among physical characteristics calculated for the promoters electrostatic potential and bendability profiles demonstrated flat regions associated with GAA-repeats, and open state activation energy profiles showed a pronounced slope downstream of transcription start site. The former might be involved in specific DNA-protein interactions, particularly, with a putative regulatory protein [2], while the latter could facilitate transcription bubble formation.
Conclusion: Reported physicochemical and structural properties of *vlhA* promoters in *M. gallisepticum* suggests their possible connection with the unclear regulatory mechanisms underlying phase variation of surface antigens. Here we highlight certain characteristic features of *vlhA* promoters physical properties. As expression regulation in the bacteria differs significantly from others, its alternative mechanisms including various DNA physicochemical properties involvement are to be furtherly considered.

References

1. Mazin P.V. et al. (2014) Transcriptome analysis reveals novel regulatory mechanisms in a genome-reduced bacterium. Nucleic Acids Res. 42:13254-13268. DOI 10.1093/nar/gku976.
2. Liu L., Dybvig K., Panangala V.S., van Santen V.L., French C.T. (2000) GAA trinucleotide repeat region regulates M9/pMGA gene expression in *Mycoplasma gallisepticum*. Infect Immun. 68:871-876. DOI 10.1128/IAI.68.2.871-876.2000.

DOI 10.18699/SBB-2018-28

Meta-analysis genomics and interactomics data of relationship between the host and *Bacillus anthracis*

G.A. Pechkosvkii*, A.V. Abramovich
Stavropol Research Institute for Plague Control, Stavropol, Russia
** e-mail: grigorii.pechkovskii@gmail.com*

Key words: host-pathogen interaction, *Bacillus anthracis*, data analysis

Motivation and Aim: *B. anthracis* passes in the host organism through the stages of development from spores into a vegetative cell. During this time there is a significant change in the genes expression of *B. anthracis* genes [1]. At present, a considerable amount of data has been collected. But there are still many mysteries about the life cycle of anthrax in the host. We have made an attempt to create pipeline for multiomics data analysis and applied it on genomic and interactomics *B. anthracis* data.
Methods and Algorithms: The pipeline was organized using custom python script. On the input of the pipeline, sequences are taken in the assemblies (from omics data) and the table to which groups this assembly belongs. The first stage in algorithm uses a BLAST for the selection of sequences with a certain degree of identity and places them in separate initial fasta files. Then groups them according to the specified cluster, makes a multiple alignment initial and grouped fasta files used MUSLE. The analysis used interactom data (IMEx IM-13779)[2], genomic data 36 complete RefSeq genome *Bacillus cereus* group, 12 genome well and 7 weakly susceptible to anthrax animal. Sequence was obtained from interaction data with the Uniprot id. And formed of a proteins assembly for *B. anthracis* and *H. sapiens*.
Results: Calculated the multiple alignment parameters (entropy, distance matrix), obtaining a set of intersecting clusters sequence. Common proteins in the *B. cereus* group 450 out of 936 with mean entropy 37,16 (SD 50,7), common proteins in *B. anthracis* cluster 854 out of 935 with mean entropy 4,25 (SD 25,5). Intersecting set proteins between well and weakly susceptible animal 925 out of 1638 with mean entropy 353,9 (SD 404) in well and 261,3 (SD 311,5) in weakly susceptible animal.
Conclusion: Thus, the obtained data suggest that quite a large number proteins *B. cereus* group is homologues to protein involved in the interaction between the host and *B. antracis*. But it has significantly more substitutions compared to the strain of *B. antracis*. Perhaps these proteins have undergone evolutionary changes to maintain pathogenicity. Also the values between well and weakly susceptible animal indicate a strong variety in protein. Probably, such data will be useful to understand inherited susceptibility to infection. Therefore, the evidence can help in the search for proteins that affect the course of anthrax disease. The data and script are available via email request.

References

1. Bergman N.H. et al. (2007). Transcriptional profiling of Bacillus anthracis during infection of host macrophages. Infection Immunity. 75(7):3434-3444.
2. Dyer M.D. et al. (2010). The human-bacterial pathogen protein interaction networks of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*. PloS One. 5(8):e12089.

DOI 10.18699/SBB-2018-29

Transcriptomic comparative analysis of hippocampal tissue and primary cultures after hyaluronidase treatment

V. Pershin[1, 2]*, A. Balashova[2], E. Guryev[1, 2], L. Kurbatov[3], M. Gainullin[1, 4], I. Mukhina[1, 2]

[1] *Privolzhsky Research Medical University, Nizhny Novgorod, Russia*
[2] *Lobachevsky State University, Nizhny Novgorod, Russia*
[3] *Orekhovich Institute of Biomedical chemistry, Russian Academy of Medical Sciences, Moscow, Russia*
[4] *Oslo University Hospital, Oslo, Norway*
** e-mail: bp1995@yandex.ru*

Key words: Extracellular matrix, epilepsy, hippocampus, gene expression, transcriptomics.

Motivation and Aim: It is well-known that traumatic injury or enzymatic digestion of brain extracellular matrix can cause seizure-like activity in primary hippocampal culture [1]. However, underlying molecular mechanism and similarity between *in vitro* and *in vivo* models of epilepsy is poorly investigated. The aim of this study was to investigate changes of gene expression and perform comparative transcriptomic analysis while modeling of epileptogenesis.

Methods and Algorithms: C57BL/6J mice were used to prepare hippocampal cell culture and intra-hippocampal hyaluronidase injection. Hyaluronic acid, which is the basis of extracellular matrix, has been removed by 75U/ml hyaluronidase. Transcriptomic analysis was performed by mouse full-genome 2-colour Agilent Microarrays. Normalization, processing and detection of differentially-expressed genes were performed by online free software – Babelomics 5.0 [2]. Functional annotation clustering was made by David bioinformatics resource [3].

Results: Hyaluronan digestion resulted in significant changes of expression of 3 gene clusters in neuronal hippocampal cultures – synaptic gene cluster, ribosomal gene cluster and mitochondrial gene cluster, while in hippocampus hyaluronidase injection resulted in changes of such functional gene clusters: synaptic gene cluster, neurogenesis gene cluster, actin-binding gene cluster, circadian rhythm cluster and ubiquitinylation gene cluster. When comparing control groups of *in vivo* and *in vitro* samples there were found 3 significant gene clusters: calcium ion-binding cluster, cluster of genes expressing EGF-domains and cortical cytoskeleton gene cluster.

Conclusion: The results obtained in this study suggest that hyaluronic acid digestion mainly affect synaptic gene cluster that indicate to similarity of *in vivo* and *in vitro* models. A wide variety of gene clusters from isolated hippocampus point to high complexity of biological processes in mice brain. Transcriptomic comparative analysis revealed strong difference between cell culture and hippocampus that should be considered when interpreting *in vitro* data.

Acknowledgements: Supported by the RFBR (16-04-01763).

References

1. Vedunova M.V. et al. (2013) Seizure-like activity in hyaluronidase-treated dissociated hippocampal cultures. Frontiers in cellular Neuroscience. 7(149):81-89. DOI 10.3389/fncel.2013.00149.
2. Alonso R. et al. (2015) Babelomics 5.0: functional interpretation for new generations of genomic data. Nucleic Acids Res. 43(1):117-121.
3. Huang D.W. et al. (2009) Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 4(1):44-57.

DOI 10.18699/SBB-2018-30

Using molecular docking to find the inhibitor of bacterial cellulose synthesis

I. Petrushin
Irkutsk State University, Irkutsk, Russia
e-mail: ivan.kiel@gmail.com

Key words: molecular docking, biofilm, cellulose synthase

Motivation and Aim: Most of bacterium live into the biofilms. This film protects bacterial colonies against antibiotics and other external impact. It's important to destroy biofilm when treating bacterial infections [1]. Biofilm consist of cellulose, which is resistant to external impact. That's why we chose to block its synthesis by interfering normal work of specific protein – cellulose synthase. Cellulose production begins when small molecule, ligand (c-di-GMP) binds with cellulose synthase. We study how to block active site of cellulose synthase by different ligand.
Methods and Algorithms: We can estimate the energy of interaction between protein and ligand using molecular docking. In this work we use well known AutoDock Vina package to calculate the energy and AutoDock Tools to prepare input files and visualize the results. To narrow the field of search we decided to study group of natural plant metabolites (flavonoids) with protein cellulose synthase which structure and mechanism described in [2]. Flavonoids are commercially available and have antiseptic properties.
Results: Virtual screening of hundreds of compounds similar to original ligand (c-di-GMP) was studied. We discovered several flavonoids with binding energy closer to energy for original ligand – c-di-GMP.
Conclusion: It's necessary to estimate ligand–protein interaction with molecular dynamics simulation and study ligand impact to biofilm *in vitro*.

References

1. Monds R.D., O'Toole G.A. (2009) The developmental model of microbial biofilms: ten years of a paradigm up for review. Trends Microbiology. 17:73-87.
2. Mazur O., Zimmer J. (2011) Apo- and cellopentaose-bound structures of the bacterial cellulose synthase subunit BcsZ. J. Biol. Chem. 286:17601-17606.

DOI 10.18699/SBB-2018-31

Heterogeneity of linkage disequilibrium across human population and its influence on statistical properties of the conditional and joint models for genetic association testing

A.V. Severinov[1, 3]*, S.A. Slavskii[1, 3, 5]**, T.I. Shashkova[1, 2, 3], D.D. Gorev[1, 2, 3], D.G. Alexeev[1, 2],G.A. Bazykin[5, 6], Y.S. Aulchenko[2, 4]

[1] *LLC 'Knomics', Moscow, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Moscow Institute of Physics and Technology, Moscow, Russia*
[4] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[5] *Skolkovo Institute of Science and Technology, Skolkovo, Russia*
[6] *Institute for Information Transmission Problems* (*Kharkevich Institute*) *of the RAS, Moscow, Russia*
** e-mail: aleksandr.severinov@phystech.edu; ** slavsky@phystech.edu*

Key words: genome-wide association study, linkage disequilibrium, conditional and joint association testing

Motivation and Aim: Conditional and joint (COJO) tests of association between multiple single-nucleotide polymorphisms (SNPs) on the basis of summary statistics reported in single-SNP genome-wide association scans (GWAS) require knowledge of linkage disequilibrium (LD) in the region undergoing the testing. Because of different reasons, in most situations the LD structure from the sample(s) where the original GWAS was performed is not available. Therefore methods based on analysis of GWAS results often have to use LD computed from some other, reference sample. When doing so, it is (implicitly) assumed that the LD structure in the reference sample approximates the LD in original GWAS sample well. It is not quite clear how possible difference in the LD structure between the GWAS and reference populations influence the statistical properties of methods, such as COJO, that utilize summary statistics. In practice, it is also not known how large could be the LD difference between a typical GWAS and a reference populations. In this work, our aim was to estimate the distribution of difference in the LD structure between European populations and to relate the error in estimation of LD due to use of reference population to the statistical properties of multi-SNP conditional and joint (COJO) analysis of genetic associations.
Methods and Algorithms: For estimation of the variance of LD between different populations we developed and applied a model of genetic drift. Using effective population size and number of generations since divergence as input parameters along with allele frequencies we considered a problem of the genetic drift of haplotypes across generations. The theoretical results were compared to the distributions observed in real whole genome data from different European populations. To investigate the effects of the variance in LD between reference and GWAS population, we have developed and implemented an algorithm for simulating genotypes for two SNPs with certain LD coefficient and a quantitative phenotype. After obtaining simulated data we ran COJO analysis on it varying input parameters (particularly LD-coefficient between SNPs). Then, we estimated the rate of false positives and false negatives for both conditional and joint tests as a function of deviation of the LD-coefficient from its true value.
Results and Conclusions: Using *in silico* analysis we estimated possible differences in LD between European populations and confirm our results by analysis of real data. On the other hand, we estimated the effect of use of biased LD estimates onto the statistical properties of the COJO method. Lastly, we demonstrated what may be the practical consequences of use of LD computed in a reference population instead of true unknown LD in the context of COJO analysis.

DOI 10.18699/SBB-2018-32

Identification of biomarkers and intervention targets for coronary artery disease based on results of genome-wide association scans

T.I. Shashkova[1, 2, 3]*, A. Torgasheva[4], D.D. Gorev[1, 2, 3], E. Pakhomov[2], L. Klaric[5, 6], A.V. Severinov[1, 3], S. Sharapov[2, 4], Y.A. Tsepilov[2, 4], D.G. Alexeev[1, 2], J.F. Wilson[5], P. Joshi[5], Y.S. Aulchenko[2, 4]

[1] *LLC 'Knomics', Moscow, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Moscow Institute of Physics and Technology, Moscow, Russia*
[4] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[5] *University of Edinburgh, Edinburgh, UK*
[6] *Genos Glycoscience Laboratory, Zagreb, Croatia*
** e-mail: shashkova@phystech.edu*

Key words: genome-wide association studies, coronary artery disease, biomarker, intervention target

Motivation and Aim: Results from tens of thousands of GWAS that have been performed over the last decade are publicly available. The results are typically presented in the form of genome-wide summary statistics (for each SNP the allelic frequencies, estimates of the coefficients of regression and their standard errors are usually reported). This information can be used for multiple purposes – from research in fundamental biology and genetics, to biomarker and target discovery for therapeutic intervention. The aim of this work is to demonstrate that mining large databases of GWAS results allows identification of biomarkers and intervention targets. We focussed on coronary artery disease (CAD), one of the most economically and socially significant and one of the most studied complex common human diseases.

Methods and Algorithms: We developed a system, 'GWAS-MAP', that allows for a platform for storage, quality control and analysis of GWAS summary statistics. Our platform embeds LDsr and MRbase libraries, facilitating genetic correlations and mendelian randomization (MR) analyses, respectively. The analysis of pleiotropy is possible via our own implementation of summary-level mendelian randomization (SMR)/heterogeneity in dependent instruments (HEIDI) testing. We populated our database with about 220 GWASes. These included GWASes of lipid levels, GWASes for 128 metabolites, and 82 proteins from OLINK panel. We have also used eQTL data from a range of sources.

Results: Observed genetic correlations were consistent with previous studies. We selected 51 loci that were associated with CAD at genome-wide significant level in published GWASes. To define loci profiles in "omics" space we considered SMR results for CAD in metabolomic and proteomic space, which allowed us to cluster loci in several biologically meaningful groups. To understand the biological bases of the locus action, and to potentially provide a drug target, within each locus, we have prioritised the genes using the SMR/HEIDI test between CAD and gene expression. Our results confirmed existing knowledge of CAD mechanisms and suggested several CAD biomarkers and intervention targets.

Conclusion: The analysis of CAD using GWAS-MAP allowed for identification of biomarkers and potential intervention targets. Some of these biomarkers and targets are already well known and are used in clinical practice, validating the approach, whilst some are new – showing our approach to biomarker and target discovery is effective.

DOI 10.18699/SBB-2018-33

Identifying parasite resistance genes in *Solanum tuberosum* by RNA-seq

N. Shmakov[1]*, A. Egorova[1], A. Glagoleva[1, 2], A. Kochetov[1, 2], D. Afonnikov[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: shmakov@bionet.nsc.ru*

Key words: resistance genes, NBS-LRR, RNA-seq, differential expression, transcriptome assembly

Motivation and Aim: Enhancing resistance of cultivated plants, namely potato *Solanum tuberosum*, to parasitic organisms is an important but complicated task. It was revealed that nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes compose the largest pathogen resistance gene family [1]. This genes family includes genes potentially responsible for resistance to golden potato cyst nematode (GPCN) *Globodera rostochiensis*. GPCN is one of the most economically important potato pathogens [2]. Thus, discovery of genes controlling potato resistance to GPCN is an important task. To achieve this goal, we performed RNA-seq analysis of two potato cultivars varying in their GPCN resistance.

Methods and Algorithms: *S. tuberosum* plants of genotype i-0144786, susceptible to GPCN, and genotype i-0144787, resistant to GPCN, were inoculated with GPCN cysts or treated with water as control group, and total RNA was extracted from plant roots on different time stages. Libraries of paired-end short reads were sequenced using Illumina HiSeq 2500 system. Libraries were filtered from low-quality, low-length reads, singletons, ambiguous reads, and adapter sequences were removed. Libraries were mapped to the reference *S. tuberosum* genome, and search for differential expression was performed. In addition, *de novo* transcriptome reconstruction was carried out, and, after excluding lowly-presented transcripts, search for differentially expressed transcripts was conducted. Additionally, transcripts were functionally annotated. To achieve this, domain structures were predicted in assembled contigs. Verification of differential expression through qRT-PCR was performed for a selected list of genes.

Results: A number of differentially expressed transcripts were detected both with library mapping analysis and transcriptome *de novo* assembly. Based on motif homology, NBS-LRR genes were predicted, including a number of *de novo* assembled transcripts presented in i-0144787 resistant genotype but not in i-0144786 susceptible genotype.

Conclusion: In this study we revealed several NBS-LRR genes associated with GPCN resistance

Acknowledgements: Supported by the RSF (16-16-04073). Plants of genotypes i-0144786 and i-0144787 were presented by VIR.

References

1. Shao Z.Q., Wang B., Chen J.Q. (2016) Tracking ancestral lineages and recent expansions of NBS-LRR genes in angiosperms. Plant Signal Behav. 11(7):e1197470.
2. Kochetov A. et al. (2017) Differential expression of NBS-LRR encoding genes in the root transcriptomes of two *Solanum phureja* genotypes with contrasting resistance to *Globodera rostochiensis*. BMC Plant Biology. 17(Suppl. 2):251. DOI 10.1186/s12870-017-1193-1.

DOI 10.18699/SBB-2018-34

The interplay of Piwi and heterochromatin proteins in transposable element silencing in the germline of *Drosophila melanogaster*

A. Stolyarenko, T. Leinsoo, Y. Abramov, M. Klenov
Institute of Molecular Genetics RAS, Moscow, Russia
** e-mail: stol@img.ras.ru*

Key words: Piwi, Argonaute, HP1a, heterochromatin, transcriptional silencing, RNAseq

Motivation and Aim: Transposable element (TE) silencing is critical for germline genome stability and fertility of animals, including Drosophila melanogaster. The piRNA-binding protein Piwi of the Argonaute family was shown to induce transcriptional repression of TEs in *Drosophila* ovaries [1]. However, the interplay of Piwi and the main heterochromatin protein HP1a and other heterochromatin proteins in this process is still not fully understood, and Piwi is usually considered to team with HP1a to carry out TE repression [2].
Methods and Algorithms: We took advantage of a genetic approach of combining in one genotype germline knockdowns (KD) of Piwi and HP1a and comparing them with single Piwi and HP1a KDs in the same genetic background via RNA-seq. We divided TEs responding to the double knockdown according to their response to Piwi and HP1a KDs into four groups.
Results: A group of TEs are equally controlled by Piwi or HP1a which points to the participation of these two proteins in a single pathway to silence the elements of this group. Another group of TEs are controlled by Piwi alone, but not HP1a. A group of elements is controlled by HP1a and Piwi independently, but in combination. These elements do not respond to the knockdown of Eggless H3K9 methyltransferase creating binding sites for HP1a. This implies independent double control by systems based on Piwi and HP1a. Interestingly, according to our previously published ChIP-seq data, the chromatin of these TEs is especially highly enriched in the repressive H3K9me3 mark, which is indicative of their localization in heterochromatin. Finally, there is a group of Piwi-independent elements controlled by HP1a alone. Interestingly, we found that these elements coincided with the ones known to be derepressed upon mutation of endosiRNA-binding protein Ago2 [3]. We propose that the repression of these elements by Ago2 is assisted by HP1a in the germline.
Conclusion: Our results indicate that Piwi-induced silencing and Piwi-independent repression by heterochromatic proteins may represent two distinct systems that silence transcription of TEs in the ovarian germline. These two systems can act autonomously or perhaps in association with each other to repress an element.
Acknowledgements: Supported by the RFBR (16-04-01524).

References

1. Klenov M. et al. (2007) Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. Nucleic Acids Res. 35(16):5430-5438.
2. Sienski G. et al. (2015) Silencio/CG9754 connects the Piwi–piRNA complex to the cellular heterochromatin machinery. Genes Dev. 29(21):2258-2271.
3. Czech B. et al. (2008) An endogenous small interfering RNA pathway in *Drosophila*. Nature. 453(7196): 798-802.

DOI 10.18699/SBB-2018-35

Bioinformatics study of genes expression in rat brain areas by RT-PCR and their role in behavior

K.A. Tabanyuhov[1, 2]*, I.V. Chadaeva[1, 3], A.O. Bragin[1], Y.L. Orlov[1, 3]*
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State Agrarian University, Novosibirsk, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: tabanyuhov93@mail.ru;orlov@bionet.nsc.ru*

Key words: transcriptomics, laboratory rats, aggressive behaviour, gene expression

Motivation and Aim: We studied mechanisms of hereditary-mediated aggressive behavior on laboratory animal models based on transcriptome profiling. Set of experiments in different model animals were used to compile the data set of genes related to behavior and estimate their conservation across species in a database prototype.
Methods and Algorithms: We used unique experimental model of grey rats (*Rattus norvegicus*) selected by aggressive behavior toward human. We estimated the gene expression in rat brain areas [1]. RNA-seq sequencing of rat brain areas samples was done using Illumina HiSeq. Next we used RT-PCR.
Results: We tested several genes presumably associated with the manifestation of aggressive behavior. We have studied the following genes expression in rat hypothalamus: Egr1, Gabrd, Zic2, Shank3. All the genes were differentially expressed in a hypothalamus of tame and aggressive rats. Egr1 had lower expression in tame rat hypothalamus than in aggressive rats. 3 other genes – Gabrd, Zic2 and Shank3 had significantly higher expression in tame rat hypothalamus than in aggressive rats (P-value less than5E-5). GABRD codes the gamma-aminobutyric acid type A (GABA-A) receptor delta subunit. GABA-A receptors are ligand-gated chloride channels. Alternatively spliced transcript variants have been described for this gene, but their biological validity has not been determined. The protein encoded by EGR1 gene belongs to the EGR (early growth response) family of C2H2-type zinc-finger proteins. Studies suggest this is a cancer suppressor gene.Zic2. The protein encoded by this gene is a Zic family member 2 and acts as a transcriptional repressor and may regulate tissue – specific expression of dopamine receptor D1. Dopaminergic transmission is associated with motivation, learning and cognition. The set of computer tools and data processing pipelines helped to find genes and gene regulation patterns applied to behavior models [1].
Conclusion: Many synapse-related genes have statistically significant deviation in splicing depending on brain region and aggressive/tame status in rat. It is a novel phenomenon of the transcriptome data related to aggressive behavior
Acknowledgements: The research has been supported by RFBR with support by ICG SB RAS budget project (0324-2018-0017).

References

1. Babenko V.N. et al. (2016) Analysis of differential gene expression by RNA-seq data in brain areas of laboratory animals. Journal Integrative Bioinformatics. 13(4):292.

DOI 10.18699/SBB-2018-36

Overexpression of *whiA* in *Mycoplasma gallisepticum*

E. Tsoy[1, 2], D. Evsyutina[1, 3]*, G. Fisunov[1]
[1] *Federal Research and Clinical Centre of Physical-Chemical Medicine, Moscow, Russia*
[2] *Pirogov Russian National Research Medical University, Moscow, Russia*
[3] *Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia*
** e-mail: dar-evsyutina2@yandex.ru*

Key words: Mollicutes, transcription, WhiA, transcription factor, regulation of gene expression

Motivation and Aim: Mycoplasma species are bacteria that lack cell wall and have reduced genome. There is a scanty knowledge about proteinaceous regulators of transcription in Mycoplasmas [1]. WhiA is a conserved protein in Gram-positive bacteria, and also is present in the Mycoplasmas. This protein has DNA-binding domain. In *Streptomyces coelicolor* WhiA regulates sporulation [2], but in *Bacillus subtilis* it influences in chromosome segregation [3] and is involved in cell division [4]. The function of this protein in the Mycoplasmas remains unclear. In this work we searched for targets of WhiA in an avian pathogen *Mycoplasma gallisepticum*.
Methods and Algorithms: Transposon-based vector pTn4001opt_whiA were constructed for overexpression *whiA* gene in *Mycoplasma gallisepticum S6.* Transformation of *M. gallisepticum* was done by electroporation. All bacteria strains were cultured in liquid medium for exponential phase and cDNA samples were prepared as previously described [5].
Results and Conclusions: The growth rate of trasformants and wild-type bacteria was the same. We selected two independent clones for all experiments. RT-PCR analysis was done for genes involved in all main biological processes and metabolic pathways. The expression of about 80 genes (10 % of all ORFs) was checked. Expression level of *nei* (*mutM*), *fba* and *glpF* was changed under *whiA* overexpression condition. All these proteins can take part in control of redox homeostasis. Fba is a central enzyme of glycolysis but it also has a moonlight function as transcription regulator of catalase and RNA polymerase subunit [6]. Nei (MutM) is a base excision repair enzyme that identifies and eliminates a large variety of oxidized purines from DNA. GlpF is a transporter of glycerol in a pathway for production peroxide. The exact role of WhiA is still not unclear but our results show a further direction of the study.
Acknowledgements: This work was funded by the Russian Science Foundation grant 14-24-00159 “Systems research of minimal cell on a *Mycoplasma gallisepticum* model”.

References

1. Fisunov G.Y. et al. (2016) Reconstruction of transcription control networks in Mollicutes by high-throughput identification of promoters. Frontiers Microbiology. 7:1977.
2. Kaiser B.K., Stoddard B.L. (2011) DNA recognition and transcriptional regulation by the WhiA sporulation factor. Scientific Reports. 1:156.
3. Bohorquez L.C. et al. (2018) The conserved DNA binding protein WhiA influences chromosome segregation in Bacillus subtilis. Journal Bacteriology. JB.00633-17.
4. Surdova K. et al. (2013) The conserved DNA-binding protein WhiA is involved in cell division in *Bacillus subtilis*. Journal Bacteriology. JB.00507-13.
5. Mazin P.V. et al. (2014) Transcriptome analysis reveals novel regulatory mechanisms in a genome-reduced bacterium. Nucleic Acids Research. 1;42(21):13254-68.
6. Ziveri J. et al. (2017) The metabolic enzyme fructose-1,6-bisphosphate aldolase acts as a transcriptional regulator in pathogenic Francisella. Nature Communications. 8:853.

DOI 10.18699/SBB-2018-37

Bioinformatics tools for 3D chromosome contacts analysis

A.V. Tsukanov[1], A.O. Bragin[1], O. Thierry[2, 3], Y.L. Orlov[1, 3]*
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *University of Bordeaux, Paris, France*
[3] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: ya.cukanton@yandex.ru*

Key words: genomics, 3D genome structure, gene expression regulation, transcription factor binding sites, Hi-C, ChIA-PET

Motivation and Aim: Transcription regulation in eukaryotes is a complex process, in which chromatin interactions play a critical role for gene expression regulation as well as to further influence other cellular activities. Series of post-genome technologies have been developed to study the binding of transcription factors (TF) for transcription regulation, such as chromatin immunoprecipitation arrays (ChIP-chip), ChIP-Seq [1]. Another challenge is to define whether such binding sites distal from gene regions are functional, i. e. physically contact target gene promoters via chromosome loops or attracting RNA polymerase II complex for gene transcription. Identification of genome-wide distal chromatin interactions that lead the regulatory elements to their target genes may provide novel insights into the study of transcription regulation. Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) method for such analysis requires development of specialized software that will be reviewed in our presentation [2].
Methods and Algorithms: The aim of the work was to review existing tools for 3D genome structure develop a computer program 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. Gene annotation was obtained from UCSC Genome Browser (http://genome.ucsc.edu).
Results: The result of the program is a distribution of CTCF transcription factor binding sites on domains on the human chromosomes. 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: With the rapidly increasing resolution of Hi-C datasets, the size of the chromatin contact map will soon exceed the memory capacity of general computers. The same problem related to ChIA-PET and subsequent data integration to be solved by our software development.
Acknowledgements: The authors are grateful to I.V. Chadaeva for science discussion. The research has been supported by RFBR (grant 18-34-00496) and Computing was done at Siberian Supercomputer center SB RAS (SSCC) with support by ICG SB RAS budget project (0324-2018-0017).

References

1. Li G. et al. (2014) Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. BMC Genomics. 15(Suppl. 12):S11.
2. Orlov Y.L. et al. (2017) Computer methods of analysis of chromosome contacts in the cell nucleus based on sequencing technology data. Biomed Khim. 63(5):418-422 (in Russian).

DOI 10.18699/SBB-2018-38

Meta-analysis of whole-transcriptome data suggests new mechanisms of auxin-induced ethylene biosynthesis and signaling in *Arabidopsis thaliana*

E. Ubogoeva
Novosibirsk State University, Novosibirsk, Russia
e-mail: ubogoeva@gmail.com

Key words: auxin-ethylene crosstalk, RNA-seq, microarray

Motivation and Aim: Plant hormones auxin and ethylene are the key regulators of plant growth and development, which are widely used in agriculture. The crosstalk between these hormones is essential for regulation of many physiological processes. Ethylene often achieves its morphogenetic effects by modulating auxin biosynthesis and transport and thereby changing the distribution of auxin – the major regulator of plant morphogenesis [1]. This crosstalk is reciprocal as auxin induces ethylene biosynthesis. However, the mechanisms of auxin effects on ethylene metabolic and signaling pathways is still not clear. Here we perform meta-analysis of auxin-induced whole-transcriptome data to unveil the molecular mechanisms of auxin-induced ethylene biosynthesis and signaling in *Arabidopsis thaliana* root.
Methods and Algorithms: We used our own RNA-seq data and publicly available microarray data [2] on auxin-induced transcriptomes of *A. thaliana* root for meta-analysis. Differentially expressed genes were filtered based on Gene Ontology (GO) term inherence. We used the binomial trail estimate to separate the robust predictions. We used hierarchical clustering to find co-expression patterns.
Results: We identified 128 auxin-responsive genes, which were associated with ethylene biosynthesis, signaling or ethylene response. Auxin sensitivity of these genes was selectively verified by qRT-PCR. The auxin sensitivity of some genes was shown for the first time. We also predicted several primary ARF targets. We further determined specific auxin-induced expression profiles for ethylene biosynthesis genes. Based on the data obtained, a scheme of auxin-dependent regulation of ethylene biosynthesis and transcriptional response was proposed.
Conclusion: Due to increasing number of publicly available high-throughput genomic data meta-analysis is a powerful approach for the genome-wide screening.
Acknowledgements: Supported by Russian Foundation for Basic Research and the government of Novosibirsk region, grant No. 18-44-540039.

References

1. Muday G.K., Rahman A., Binder B.M. (2012) Auxin and ethylene: Collaborators or competitors? Trends Plant Sci. 17(4):181-195.
2. Lewis D.R., Olex A.L., Lundy S.R., Turkett W.H., Fetrow J.S., Muday G.K. (2013) A kinetic analysis of the auxin transcriptome reveals cell wall remodeling proteins that modulate lateral root development in Arabidopsis. Plant Cell. 25(9):3329-3346.

DOI 10.18699/SBB-2018-39

Variability in Gibbs energy of tRNA molecules in mitochondrial genomes of Chordates: neutral selection or evolution towards optimization of translation?

K. Ushakova[1]*, A.A. Mikhaylova[1], A.G. Mikhaylova[1], D. Knorre[3], I. Mazunin[1], A. Reymond[5], K. Gunbin[1, 4], K. Popadin[1, 5]**

[1] *School of Life Sciences, Immanuel Kant Baltic Federal University, Kaliningrad, Russia*
[2] *Institute of Biodiversity Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK*
[3] *A.N. Belozersky Institute Of Physico-Chemical Biology, MSU, Moscow, Russia*
[4] *Institute of Cytology and Genetics of the SB RAS, Novosibirsk, Russia*
[5] *Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland*
** e-mail: kristina.ushakova@outlook.ru; ** konstantinpopadin@gmail.com*

Key words: mtDNA, Gibbs energy, stability, mitochondrial tRNA, ancestral state reconstruction

Motivation and Aim: It is known that translation of frequent codons in prokaryotes and some eukaryotes is optimized by increasing the copy number of corresponding tRNA gene. However, highly streamlined mitochondrial genomes of Chordata mostly hold only one tRNA gene for each amino acid. So how is mitochondrial translation optimized? We hypothesized that stability of tRNA molecules might be an important variable, correlating with codon usage (CU). It is known that translation of frequent codons in prokaryotes and some eukaryotes is optimized by increasing the copy number of corresponding tRNA gene. However, highly streamlined mitochondrial genomes of Chordata mostly hold only one tRNA gene for each amino acid. So how is mitochondrial translation optimized? We hypothesized that stability of tRNA molecules might be an important variable, correlating with codon usage (CU).
Methods and Algorithms: To test this hypothesis we reconstructed secondary structures and Gibbs energy of each tRNA from almost 4000 Chordata mito-genomes, as well as deriving various genomic features for every species. Ecological data was downloaded from the AnAge database. We also conducted reconstruction of ancestral tRNA states, using the CAT evolutionary model, at each internal node of phylogenetic tree to observe the evolutionary trend in stability.
Results: We observed that (i) In different classes of Chordata tRNA stabilities are highly variable: tend to be more stable in Aves versus Mammalia and in Actinopterygii versus Amphibia and Reptilia. GC content of the whole mitochondrial genome demonstrates the same relationship, suggesting that tRNA stability, might be just a neutral consequence of the whole genome GC content. However, comparing tRNA GC content with whole genome – we observed that warm-blooded opposed to cold-blooded Chordata have increased tRNA GC content versus background – it is possibility that tRNA stability might be under stronger selection in species with high basal metabolic rate. (ii) Comparing different species within each class, we observed positive correlations between tRNA stability and whole genome GC content. (iii) Comparing different tRNA molecules within the same genome of each species, we observed a positive correlation between tRNA stability and CU, especially in warm-blooded species. We concluded that tRNA stabilities of warm-blooded Chordata species, tend to be under stronger selection constraints of translation efficiency than those of cold-blooded Chordates.
Acknowledgements: The study was supported by the 5 Top 100 Russian Academic Excellence Project at the Immanuel Kant Baltic Federal University.

References

1. Lehmann G. et al. (2008) Do mitochondrial DNA and metabolic rate complement each other in determination of the mammalian maximum longevity? Rejuvenation Research. 11(2):409-417.
2. Satoh T.P. et al. (2010) Transfer RNA gene arrangement and codon usage in vertebrate mitochondrial genomes: a new insight into gene order conservation. BMC genomics. 11(1):479.

DOI 10.18699/SBB-2018-40

Heat shock response elements in the promoters of heat stress activated LTR retrotransposons in *Macrostomum lignano*

V. Vavilova[1]*, E. Berezikov[1, 2], K. Ustyantsev[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *European Research Institute for the Biology of Ageing, Groningen, The Netherlands*
** e-mail: valeriya-vavilova@bionet.nsc.ru*

Key words: *Macrostomum lignano*, heat shock, LTR retrotransposons

Motivation and Aim: Retrotransposon activity is generally repressed by host cell due to their impact on the genome stability, although some retrotransposons can become active again under the influence of various external stressors. However, the molecular mechanisms underlying the heat shock (HS) induced activation of retrotransposons in animals are poorly understood. *Macrostomum lignano* is a free-living flatworm that is increasingly used as model organism to study stem cell biology and regeneration. We have observed that LTR retrotransposons (LTR-RTs)-encoded transcripts were among the first top ten genes with increased expression after HS in this animal, opening up possibilities for the detailed investigation of the underlying molecular mechanism.
Methods and Algorithms: The HS elevated transcripts of the *M. lignano* LTR-RTs were mapped back to the genome assembly (http://gb.macgenome.org/). The corresponding full-length DNA sequences spanning both long terminal repeats were extracted and their internal structure was identified as in [1]. The core promoter sequences were mined inside long terminal repeats (http://www.fruitfly.org/seq_tools/promoter.html) and HS elements (HSEs) [2] were identified using regular expressions. HSEs were also screened in promoters of *M. lignano* HS protein genes and those of the known LTR-RTs available at GyDB (http://gydb.org). The phylogenetic analysis was performed using IQ-tree webserver (http://www.iqtree.org/).
Results: *M. lignano* HS-activated LTR-RTs belong to two phylogenetically distinct clusters of *Ty3/gypsy* group: *Mag* and *CsRN1*, with copy numbers of 49 and 10 in the Mlig_3_7 genome assembly. Elements of both clusters have multiple canonical HSEs (nGAAnnTTCnnGAAn) upstream of the core promoter regions, which are also present in the promoters of *M. lignano* HS protein genes. Interestingly, previously the canonical HSEs were found only for the *Ty1/copia* LTR-RTs of *Arabidopsis thaliana*, which is also activated upon HS [3]. HSEs were absent in long terminal repeats of other known LTR-RTs available at GyDB, for which there is no data for the HS activation.
Conclusion: *M. lignano* LTR-RTs acquired multiple canonical HSEs upstream of their core promoters similar to the ONSEN element in *A. thaliana*, suggesting similar activation mechanisms and, thereby, possible convergent evolution of these elements. This is the first example of HSEs in animal LTR-RTs.
Acknowledgements: The project was supported by the RFBR (Grant No. 18-34-00288).

References

1. Ustyantsev K. et al. (2017) Convergence of retrotransposons in oomycetes and plants. Mobile DNA. 8(4).
2. Mittal D. et al. (2011) Binding affinities and interactions among different heat shock element types and heat shock factors in rice (*Oryza sativa* L.). FEBS Journal. 278(17):3076-3085.
3. Cavrak V.V. (2014) How a retrotransposon exploits the plant's heat stress response for its activation. PLoS Genetics. 10(1):e1004115.

DOI 10.18699/SBB-2018-41

Research of molecular mechanisms of pathogenesis of depression: bioinformatical analysis of transcriptomic data

I.N. Vlasov*, E.A. Bondarenko, M.I. Shadrina, P.A. Slominskiy
Institute of Molecular Genetics, RAS, Moscow, Russia
** e-mail: invlasov@mail.ru*

Key words: depression, RNA-seq, animal models

Motivation and Aim: According to the data of World Health Organisation (WHO), depression is one of the most widespread mental disorders, affectimng more than 300 million people of all ages worldwide. Depression is also a leading cause of disability worldwide. Although there are known, effective treatments for depression, fewer than half of those affected in the world (in many countries, fewer than 10 %) receive such treatments.

At this point, a significant amount of studies on the pathogenesis of depression have been done, and a several hypotheses have been formulated. Such hypotheses include monoaminic, stress induced, neuro-plastic and several others. Despite that, no comprehensive understanding of ethiopathogenesis of depression exists. Because of that search for molecular mechanisms, involved in pathogenesis of depression has to go on.

Methods and Algorithms: One of the problems in the research of depression is the impossibility of obtaining brain tissue from a living patient with depression. Due to that, one of the possible approaches to studying depression is to study pathogenesis of depression on animal models. Since no perfect animal model of depression exists, several different models have to be studied. Simultaneous analysis of various models of depression allows to identify the most important processes, which could be involved in the development of this disorder.

Results: At this moment, we are conducting the bioinformatical analysis of transcriptomic data, derived from various models of depression. We study both models, obtained in our own laboratory – model of acute stress (forced swimming test, Porsolt test), model of stimulation of immune system (lipopolysaccharide injections), and ones available in Gene Expression Omnibus (GEO). Also we are analysing the transcriptomic data, derived post-mortem from the patients with major depressive disorder from GEO.

Acknowledgements: This research has been conducted with the support of Russian Science Foundation (grant 16-15-10199).

DOI 10.18699/SBB-2018-42

Analysis of the molecular mechanisms of *Pinus sylvestris* and *Hordeum vulgare* adaptation to the action of low doses of ionizing radiation

P. Volkova[1], G. Duarte[2], E. Kazakova[1]*
[1] *Russian Institute of Radiology and Agroecology, Obninsk, Russia*
[2] *Institute Jean-Pierre Bourgin, Versailles, France*
** e-mail: alvaly@mail.ru*

Key words: adaptation, irradiation, low doses, gene expression, *Pinus sylvestris*, *Hordeum vulgare*

Motivation and Aim: The revealing of molecular mechanisms for the formation of adaptive plant responses to low-intensity stress factors (including low doses of ionizing radiation) is an important problem in modern biology. In order to clarify the mechanisms of adaptation of plants to the action of low-dose radiation we are conducting two experiments: (1) the study of chronically irradiated populations of *Pinus sylvestris* from the exclusion zone of accident at the Chernobyl nuclear power plant, and (2) on *Hordeum vulgare* plants growing from gamma-irradiated seeds. The results obtained can be applied in the field of radiation protection of biota and genetic engineering of crops.
Methods and Algorithms: For the identification of the genes involved in the formation of the response to chronic irradiation (annual dose up to 130 mGy) in populations of *P. sylvestris*, a transcriptome analysis using high-throughput RNA sequencing was performed. Analysis of expression of genes involved in the response of seedlings of *H. vulgare* on irradiation of seeds in doses of 0, 20, 100 Gy is performed using DNA-microarrays (Agilent). The validation of the results in both experiments will be done by analyzing the expression of selected candidate genes by RT-PCR.
Results: After analysis of transcripts in chronically irradiated populations of *P. sylvestris* we identified the main groups of genes responding to chronic exposure. These genes code the transcription factors, components of the antioxidant system, membrane proteins, histones, genes associated with the biosynthesis of ABA, genes maintaining work of chloroplasts and activity of transposons. There was no differential expression of genes associated with repair processes of DNA damage [1]. Candidate genes for the further validation were chosen. The analysis of the transcriptome of *H. vulgare* is currently under way.
Conclusion: The populations of *P. sylvestris* respond to chronic radiation exposure by the differential expression of hundreds of genes, the analysis of which has helped to approach the understanding of the basis of plant reactions to chronic low-dose irradiation under natural conditions. It is also assumed that *H. vulgare* transcriptome analysis will identify the genes involved in the adaptive response of plants to low-dose irradiation in the laboratory conditions (in particular, the formation of the effect of radiation stimulation).
Acknowledgements: Supported by RSF (14-14-00666), by RSF (17-74-10028).

References

1. Duarte G.T. et al. (2018) Radioactive contamination in the Chernobyl zone promotes changes in gene expression in gymnosperms. Nature Ecology Evolution (under review).

DOI 10.18699/SBB-2018-43

Detection of pathogenic mutations in patients with non-coronary heart diseases by targeted NGS

D. Yermakovich[1]*, L. Sivitskaya[1], T. Vaikhanskaya[2], A. Liaudanski[1], N. Danilenko[1]
[1] *Institute of Genetics and Cytology NAS, Minsk, Belarus*
[2] *Republican Scientific and Practical Center of Cardiology, Minsk, Belarus*
** e-mail: cytoplasmic@mail.ru*

Key words: targeted NGS, non-coronary heart diseases, bioinformatics processing, pathogenic variants

Motivation and Aim: Various heart diseases, the etiology of which is not related to the change of coronary vessels, belong to the group of non-coronary. The most common of them are the autosomal dominant forms of cardiomyopathies – dilated (DCM), arrhythmogenic (AC), hypertrophic (HCM). They can develop due to a mutation in more than 100 genes. Mutations in sarcomere protein genes are the main cause of HCM and DCM [1]. Variants in genes encoding desmosomal proteins are usually detected in patients with AC [2].

Methods and Algorithms: The targeted NGS was carried out for identification of pathogenic mutations in the 24 unrelated patients with non-coronary heart diseases using the TrueSight Cardio Sequencing Kit. Additionally, investigation of the LMNA gene sequence in the 18 patients with DCM was conducted using the Nextera XT kit. The processing of raw FASTQ files was carried out using software (Trimmomatic, Bowtie2, Samtools, Bcftools, VCFlib) on the GNU/Linux platform. Using the ANNOVAR program, the primary annotation of the discovered variants was obtained. The bash script has been created to automate the processing and annotation of sequencing results. The functional impact of the variants with uncertain significance has been tested *in silico* by the next services: Human Splicing Finder, Jpred4, Align-GVGD, Condel.

Results: In 8 unrelated patients, 11 nucleotide variants in 8 different genes have been detected, that have a high probability of being the cause of the disease. Family genotyping showed a segregation of mutations with cardiac phenotype. *In silico* analysis indicated the pathogenicity of them. According to population databases, the variants have an extra low frequency (MAF < 0.01). Four of them are described in the literature as pathogenic: rs397517853 (NEXN, nonframeshift deletion), rs267607626, rs121912496 (LMNA, missense), rs794728589 (LMNA, splicing). Two missense – rs766910280 (SCN1B), rs377473560 (MYH6) – haven't been registered as a clinical variant before. The others: stop-gain and splicing mutations in TTN and 3 missenses in MYH7, ACTC1, DES – were identified for the first time.

Conclusion: The targeted sequencing is one of the most convenient and productive method of the searching for pathogenic variants in cohort of patients with polygenic diseases. In our research, the method allowed us to find 11 mutations in different genes. They have been classified as likely pathogenic and pathogenic.

References

1. Bezzina C.R., Lahrouchi N., Priori S.G. (2015) Genetics of sudden cardiac death. Current Cardiology Reports. 17(53):1-9.
2. Richard P. et al. (2003) Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. Circulation. 107:2227-2232.

DOI 10.18699/SBB-2018-44

Hepatoprotective effect of various antioxidants in the pathogenesis of opisthorchiasis

O. Zaparina[1]*, A. Kovner[1], A. Rakhmetova[2], M. Pakharukova[1, 2], V. Mordvinov[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: zp.oksana.93@gmail.com*

Key words: opisthorchiasis, inflammation, reactive oxygen species, antioxidants, liver flukes

Motivation and Aim: *Opisthorchis felineus* (liver fluke) is endemic to the territories of Europe and Asia the trematode of the family Opisthorchiidae, which infests the hepatobiliary system of fish-eating mammals including humans, causing opisthorchiasis. Pathogenesis of opisthorchiasis is accompanied by structural and functional disorders of the liver, including the formation of foci of chronic inflammation, cholestasis, cholecystitis and precancerous changes in the epithelium of the bile ducts. The mechanisms of development of this disease have not been studied; however, increased production of reactive oxygen species, which is accompanied by chronic inflammation, probably plays an important role in the pathogenesis of opisthorchiasis. The aim of the research was to study the effect of antioxidants of different mechanisms of action (natural antioxidant resveratrol and mitochondrial antioxidant SKQ1 (10- (6'-Plastoquinonyl) decyltriphenylphosphonium)) on the structural and functional state of the liver and to assess the possible role of oxidative stress in the pathogenesis of opisthorchiasis.
Methods and Algorithms: Experimental opisthorchiasis *in vivo* on golden hamsters *M. auratus* (1 and 3 months); semi-quantitative analysis of histological slides (with hematoxylin-eosin and Van-Gieson staining); Western blot analysis; blood biochemistry.
Results: Pathological changes in the liver, including inflammation, dysplasia, metaplasia, proliferation of the bile duct epithelium and periductal fibrosis were quantified. Moreover, the content of markers of proliferation and dysplasia of cholangiocytes, as well as markers of inflammation were assessed. In addition, the activity of liver transaminases, cholesterol and bilirubin was evaluated. The hepatoprotective effect of antioxidants was demonstrated irrespective of the mechanism of their action. An improvement in the biochemical parameters of the liver, a decrease in dysplasia of the bile duct epithelium and a decrease in markers of inflammation on the background of experimental opisthorchiasis was demonstrated.
Conclusion: Oxidative stress plays an important role in the pathogenesis of opisthorchiasis. Antioxidants, regardless of the mechanism of their action, have a hepatoprotective effect on the liver by reducing inflammation and dysplasia, and, probably, by reducing oxidative stress.
Acknowledgements: Supported by the RFBR (No. 16-04-00356A, 18-04-00417A).

DOI 10.18699/SBB-2018-45

Systems analysis of the genes responding to chilling stress in *Arabidopsis thaliana* L.

A. Zlobin[1, 2], N. Omelyanchuk[1], V. Mironova[1, 2]*
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: victoria.v.mironova@gmail.com*

Key words: cold acclimation, chilling resistance, cold stress

Motivation and Aim: The chilling resistance of plants, unlike frost resistance, consists in experiencing low positive temperatures, at which ice does not form in the cell, but which affect the reduction of both yield and resistance to other unfavorable factors including pathogens, nutrient deficiencies, high light intensity, etc. Thus, the response to cold affects many processes in the plant, despite the fact that it is triggered by one factor. The objective of this project is the search for and systems analysis of genes that provide a response to low positive temperatures and resistance to them in Arabidopsis. The focus of the project is the mechanisms of chilling resistance in plant roots.
Methods and Algorithms: In our analysis we used more than 20 different datasets on study of chilling stress in *Arabidopsis thaliana* taken from public access. We preprocessed the datasets uniformly, applying free software environment R. For each dataset we used Benjamini-Hochberg correction, FDR-adjusted p-value <0.05 and received list of differentially expressed genes (DEGs). Datasets with less than 50 DEGs were excluded from the analysis.
Results: As a result of meta-analysis of the datasets, we composed a summary table for genes responding to chilling stress in Arabidopsis. More than ¾ of Arabidopsis genome significantly changed the level of transcription in response to chilling stress at least in one experiment. We identified the genes which systematically changed the level of transcription (in five or more datasets). Six genes we found at the gene network core, among them only SUS1 gene was discovered earlier as cold sensitive in Arabidopsis.
Conclusion: We performed systems analysis of whole-genome experiments on the study of chilling induced changes in transcriptomes. Meta-analysis of these data allowed us reconstructing the gene network of the plant response to chilling stress, description of its functional modules and identification of the feedbacks.
Acknowledgements: The work was supported by RSF 18-14-00249.

DOI 10.18699/SBB-2018-46

New method for estimation of number of transcription factor binding sites using results of processing of ChIP-seq data by different peak callers

S. Kolmykov[1, 2], Yu. Kondrakhin[1, 3], F. Kolpakov[1, 3]*
[1] *Biosoft.Ru Ltd., Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3] *Institute of Computational Technologies SB RAS, Novosibirsk, Russia*
** e-mail: fkolpakov@gmail.com*

Key words: ChIP-seq, quality control, peak caller, transcription factor binding sites

Motivation and Aim: Identification of transcription factor binding sites in genomes has been one of the most important tasks of modern biology. The accumulation of a large number of ChIP-seq data sets worldwide has led to the establishment of dedicated databases, in particular, ENCODE, GTRD [1] and ReMap [2]. Obviously, it is necessary to perform quality analysis of collected data sets because the quality of ChIP-seq experiments can vary significantly. The common practice for assessment of ChIP-seq data sets quality is to use well known quality criteria developed within ENCODE project. For example, NRF, PBC1, PBC2 is extensively used to control the quality of alignments while IDR and FRiP are exploited for assessment of peak callers outputs quality [3]. In addition to these quality metrics we have proposed two novel quality metrics FNCM (False Negative Control Metric) and FPCM (False Positive Control Metric) to control false negatives and false positives rates, respectively.
Methods and Algorithms: Both developed metrics are based on assessment of transcription factor binding sites number with the help of population size estimation approach. In particular, for creation of these metrics we used Chao's method, Lanumteang-Bohling method, Zelterman's method and maximum likelihood method.
Results: We applied proposed metrics to obtain refined data sets of transcription factor binding sites and for peak callers comparison of ChIP-seq data sets from GTRD database. In particular, the refined data sets allow to perform more reliable comparative analysis of position weight matrix methods for binding sites prediction. In the case of peak caller comparison, we revealed the following ranking of peak callers, when ChIP-Seq input controls are available: MACS > GEM > SISSRS > PICS. On the other hand, the altered ranking MACS > SISSRS > PICS > GEM was obtained when ChIP-Seq input controls were absent.
Conclusion: The proposed metrics are appeared to be fruitful for obtaining refined data sets of transcription factor binding sites and for peak callers comparison.
Acknowledgements: Supported by the RFBR (No. 17-00-00296 KOMFI).

References

1. Yevshin I. et al. (2017) GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments. Nucleic Acids Research. 45(Database issue):D61-D67.
2. Chèneby J. et al. (2018) ReMap 2018: an updated atlas of regulatory regions from an integrative analysis of DNA-binding ChIP-seq experiments. Nucleic Acids Research. 46(Database issue):D267-D275.
3. Stephen G.L. et al. (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research. 22(9):1813-1831.

DOI 10.18699/SBB-2018-47

The prevalence of microRNA binding sites around the *CGA*, *FSHB*, *LHB* and *TSHB* genes in mammals with a different number of dominant follicles

O.V. Lyangasova, M.A. Amelina, T.P. Shkurat*
Southern Federal University, Rostov-on-Don, Russia
** e-mail: tshkurat@yandex.ru*

Key words: microrna, *CGA*, *FSHB*, *LHB*, *TSHB*, dominant follicle

Motivation and Aim: Folliculogenesis is a constant process of the hierarchy of follicles, in which the growth and maturation of some follicles and atresia of others occurs simultaneously. Both the number of simultaneously maturing dominant follicles and the number of young are an evolutionarily fixed feature of the species.The genes that regulate the time and number of ripening follicles are practically the same in all mammals. The system of regulation of these genes can impact on the onset of ovulation in ontogeny and on the number of simultaneously maturing follicles. The aim of the study was to perform a comparative bioinformation analysis of the localization of microRNA binding sites in introns and intergenic space of *CGA* (gonadotrophic hormone alpha chain), *FSHB* (follicle stimulating hormone beta chain), *LHB* (luteinizing hormone beta chain) and *TSHB* (thyroid-stimulating hormone beta chain) genes in mammals with one and more simultaneously maturing dominant follicles.
Methods and Algorithms: The genomes of animals with a different number of dominant follicles have been used. The animals with only one dominant follicle: *Homo sapiens*, *Ovis aries*, *Bos taurus*, *Gorilla gorilla*, *Macaca mulatta*, *Pan troglodytes*, *Pongo abelii*, *Carlito syrichta*, *Equus caballus*, *Vicugna pacos*, *Otolemur garnettii*, *Loxodonta africana*, *Nomascus leucogenys*, *Myotis lucifugus*. The animals with several dominant follicles: *Canis lupus*, *Musmus culus*, *Oryctolagus cuniculus*, *Sus scrofa*, *Rattus norvegicus*, *Cavia porcellus*, *Mustela putorius furo*, *Felis catus*, *Tupaia belangeri*, *Echinopstel fairi*, *Sorex araneus*, *Erinaceus europaeus*, *Ochotona princeps*. The microRNA sequences for homologous search were taken from the miRBase database (http://mirbase.org/). The bioinformatic analysis was carried out with the Mscanner software developed by us (Shkurat et al., 2015).
Results: A high correlation of the number of microRNA binding sites with some studied indicators of the reproductive system was found (the age of maturation in females, the number of broods per year, the duration of ovulation, the length of pregnancy, the time interval between births, the weight of the birth-weight, the weight of the female). The number of microRNA binding sites around the *CGA* gene in all animals positively correlated with the duration of pregnancy r = 0.89; the birth weight r = 0.86; the interval between genera r = 0.79 and with the lifetime r = 0.7. A negative correlation was found with the duration of ovulation (estrus) r = –0.83; the number of young in the brood r = –0.82; the number of broods per year r = –0.74. The number of microRNA binding sites around the *FSHB* gene positively correlated with the age at onset of puberty in females r = 0.77, the interval between genera r = 0.87, the number of broods per year r = 0.74, the lifespan r = 0.79, the cycle duration r = 0.68. More than 4 million pairs of nucleotides were analyzed, 5,967 coincidences with the mature microRNA sequences recorded in the miRBase database were found. 167 to 262 mature microRNA sequences are located in the cis-regulatory regions of the *CGA* gene in animals with the only one dominant follicle, which is almost twice as large compared to the group of animals with several simultaneously maturing follicles (n = 68–130). 267 to 367 sequences of mature microRNAs are located in the *FSHB* gene in animals with the only one dominant follicle, which significantly exceeds these values in the group of animals with several dominant follicles (n = 99–217). There are 1 to 9 microRNA binding sites around the *LHB* gene in all animals, and 11 to 219 around the *TSHB* gene, without correlating with the number of dominant follicles.
Acknowledgements: The study was performed in the framework of the basic part of the Ministry of education and science of the Russian Federation state task, project No. 6.6762.2017/BT.

DOI 10.18699/SBB-2018-48

Kinks of plasmid PBR322

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

Key words: plasmid pBR322, nonlinear dynamics, kinks, McLaughlin-Scott equation

Motivation and Aim: Plasmid pBR322 is a small, circular DNA that is widely used in genetic research, and its components are applied at creation of new instrumental plasmids [1, 2]. This plasmid has been never investigated theoretically. We suggest, however, that pBR322 is a convenient object for mathematical modeling the internal mobility of the DNA molecule and, in particular, the movement of local conformational distortions – bubbles or kinks. In the present paper, the problem of modeling the motion of kinks in the plasmid pBR322, taking into account information on the features of its sequence containing two coding regions: CDS-1 and CDS-2, is posed and solved.
Methods and Algorithms: The movement of kinks in the plasmid pBR322 is investigated by the methods of mathematical modeling. To find the time dependence of the kink velocity and coordinate, we use the method of McLaughlin-Scott [3] complemented by the block method [4] which allows approximately to take into account the dependence of the coefficients of the equation of motion on the sequence of nucleotides. To present the results in a compact and convenient for further analysis form, the 3D trajectory method has been developed and applied.
Results: The energy profile of plasmid pBR322 has been constructed. It was shown that the kink movement in the plasmid can be interpreted as the movement of a quasi particle in the potential field with the profile having two barriers: CDS-1 and CDS-2. The time dependences of the kink velocity and coordinate were calculated for three values of the initial velocity: 150, 1650 and 1879 m/s. The 3D trajectories of the kinks as well as the projections of these trajectories in the phase plane were constructed.

References

1. GenBank: https://www.ncbi.nlm.nih.gov/nuccore/J01749.1
2. Watson N. (1988) A new revision of the sequence of plasmid pBR322. Gene. 70(2):399-403.
3. McLaughlin D.W., Scott A.C. (1978) Perturbation analysis of fluxon dynamics. Phys. Rev. A. 18:1652.
4. Yakushevich L.V., Krasnobaeva L.V. (2018) Analytical approaches for exploration of the dynamics of genes having one coding region. Biophysics. 63(1):41-53.

DOI 10.18699/SBB-2018-49

Gene networks of human hearing impairments: reconstruction and analysis

V.I. Zamyatin[1, 2]*, Z.S. Mustafin[1, 2], E.V. Ignatieva[1, 2], Yu.G. Matushkin[1, 2], O.L. Posukh[1, 2], S.A. Lashin[1, 2]

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

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

** e-mail: zamyatin@bionet.nsc.ru*

Key words: hearing impairments, gene networks, SNP, transcription factors

Motivation and Aim: Hearing loss is very genetically heterogeneous disorder since more than 100 mapped loci and about one hundred genes are now known to be implicated in hearing impairment (HI). There are nonsyndromic (isolated) forms of hereditary HI and syndromic HI (combined with other traits/disorders). Despite the fact that a lot of genes are proven to be associated with HIs, there is still no systematic network reconstructions of the molecular and signalling mechanisms of HIs. We tried to reconstruct the gene networks of human HIs considering several layers of biological organization: networks of protein-protein interactions and co-regulation networks.

Methods and Algorithms: Sets of genes associated with HIs were extracted from the Hereditary Hearing Loss resource (hereditaryhearingloss.org) as well as from [1]. Transcription factors (TF) from these sets were identified using AnimalTFDB database. Search for the binding sites for all revealed TFs by using Hocomoco MoLoTool (molotool.autosome.ru) with default settings was carried out in the upstream regions (–2000;0) of all genes from these sets. Protein-protein interactions were extracted via GeneMANIA (genemania.org) and String (string-db.org). Network reconstruction was performed in Cytoscape. Evolutionary analysis was made using Orthoscape application [2]. The data on SNPs were extracted from the 1000 Genomes Project and their effects were estimated via Ensembl VEP.

Results: We have reconstructed the gene networks for potential mechanisms of human HIs including gene regulatory and protein-protein interactions networks. The separate regulatory circuits probably corresponding to different HI mechanisms were found in these networks. Moreover, the network analysis revealed several genes encoding TFs which were not previously known to be assosiated with HIs. Evolutionary analysis of the HI gene networks has revealed several "evolutionary young" (*TPRN*, *EDN3*, *CD164*), and also predominant conservation of TFs involved in HIs. Interesting features of the *ACTG1* gene were found during the evolutionary analysis and the SNPs functional annotation in human populations: according to Ka/Ks ratio, the evolutionary conservation of this gene sequence does not correspond to its SNP enrichness. These findings probably suggest that the influence of stabilizing selection on this gene in human populations has been weakened some time ago.

Conclusion: The gene networks including the protein-protein interactions and co-regulation networks were reconstructed for human hearing impairments. The subsequent detailed analysis of these networks will help to expand the set of HI involved genes and understanding of the mechanisms of hearing and its impairments.

Acknowledgements: Partially supported by the RFBR No. 17-29-06016-ofi_m.

References

1. Stamatiou G., Stankovic K.A. (2013). Comprehensive Network and Pathway Analysis of Human Deafness Genes. Otology Neurotology. 34:961-970.
2. Mustafin Z.S. et al. (2017) Orthoscape: a Cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles. BMC Bioinformatics. 18(1):S1.

Author index

Научное издание

SYSTEMS BIOLOGY
AND BIOINFORMATICS (SBB-2018)

The Tenth International Young Scientists School

Abstracts

Printed without editing

СИСТЕМНАЯ БИОЛОГИЯ
И БИОИНФОРМАТИКА (SBB-2018)

Десятая школа молодых ученых

Тезисы докладов

Публикуется в авторской редакции

Выпуск подготовлен информационно-издательским отделом ИЦиГ СО РАН

Подписано к печати 01.08.2018. Формат 70 × 108 $^{1}/_{16}$. Усл. печ. л. 4,73.
Тираж 150 экз. Заказ № 191

Федеральный исследовательский центр
«Институт цитологии и генетики Сибирского отделения Российской академии наук»
630090, Новосибирск, проспект Академика Лаврентьева, 10

Отпечатано в типографии ФГУП «Издательство СО РАН»
630090, Новосибирск, Морской проспект, 2