SBB-2019

Systems Biology and Bioinformatics

The 11th International Young Scientists School

NOVOSIBIRSK, RUSSIA
JUNE 24–28, 2019

ORGANIZERS

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

SPONSORS

Russian Foundation
for Basic Research
Grant No. 19-04-20036

NSU
(Project 5-100)

Open access journal, PeerJ, is pleased to announce the "PeerJ Award for Best Paper" at SBB 2019. The award aims to support students and early career researchers and bring continued awareness to the benefits that open access and open communication have in furthering scientific progress. PeerJ publishes rapid open access, peer-reviewed research in life and environmental sciences. We are pleased to support SBB 2019 and rigorous scientific effort.
URL: https://peerj.com

CONTACTS

The Federal Research Center Institute of Cytology and Genetics The Siberian Branch of the Russian Academy of Sciences Russian Federation, 630090 Novosibirsk, Prospekt Lavrentyeva, 10,
phone: +7(383) 363-49-80, fax: +7(383) 333-12-78, conf.bionet.nsc.ru/sbb2019/
e-mail: sbb2019@bionet.ru, popik-olga@bionet.nsc.ru

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

SYSTEMS BIOLOGY AND BIOINFORMATICS (SBB-2019)

The Eleventh International Young Scientists School

Abstracts

June 24–28, 2019
Novosibirsk, Russia

Novosibirsk
ICG SB RAS
2019

УДК 575
S98

Systems Biology and Bioinformatics (SBB-2019) : The Eleventh International Young Scientists School (June 24–28, 2019, Novosibirsk, Russia); Abstracts / Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences. – Novosibirsk : ICG SB RAS, 2019. – 55 pp. – ISBN 978-5-91291-046-3.

International program committee

Local organizing committee

Volunteers

ISBN 978-5-91291-046-3

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

Director: Corr. Member of the RAS Alexey V. Kochetov
Academic Director: Professor, Full Member of the RAS Nikolay A. Kolchanov
Academic Secretary: Candidate of Science (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. It is situated in the Novosibirsk Akademgorodok. 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 affiliated branches: 1) Siberian Research Institute for Plant Industry and Breeding (SRIPIB). The institute is located in Krasnoobsk Village and the Novosibirsk rural area. It conducts academic, prospective, and applied studies including the collection, examination, preservation, and utilization of plant genetic resources for obtaining new biological knowledge; expansion and improvement of crop gene pools; 2) Research Institute of Clinical and Experimental Lymphology (RICEL); 3) Research Institute of Internal and Preventive Medicine (RIIPM). RICEL and RIIPM are situated in the Sovetskiy and Oktyabr'skiy districts of Novosibirsk. They conduct academic, prospective, and applied studies in molecular medicine and human genetics. They also provide medical care.

Tasks of ICG SB RAS: Solution of top-priority problems in the development of the Russian science and technology sector in 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, 2019, ICG included 89 scientific units, which employed 1354 members; of them 487 researchers, 2 RAS Advisors, 8 Full Members of the RAS, 4 Corresponding Members of the RAS, 93 Doctors of Science, and 342 Candidates of Science.

Postgraduate education and residency training: As on January 1, 2019, ICG trains 97 postgraduates.
Publications: The Institute ranks among acknowledged leaders in Russian biology. Numerous works of its researchers are published in Russian and foreign academic journals. In 2018, the overall number of publications in peer-reviewed periodicals exceeded 530. The overall number of ICG publications indexed in the Web of Science Core Collection was 402. In 2018, papers of the researchers of the Institute received over 4412 citations accounted by WoS.

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/).
Institute of Cytology and Genetics is the founder of the following journals (mass media): *Vavilov Journal of Genetics and Breeding*, *Letters to Vavilov Journal of Genetics and Breeding*, *Atherosclerosis*, *Siberian Scientific Medical Journal*, and *Live Science*.

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

Address: Prospekt Lavrentyeva 10, Novosibirsk, 630090 Russia; *phone:* +7(383) 363-49-80;
fax: +7(383) 333-12-78; *URL:* www.bionet.nsc.ru; *e-mail:* icg-adm@bionet.nsc.ru

*NSU IN FACTS & FIGURES

7200 STUDENTS

1400 FOREIGN STUDENTS from 57 countries

6 departments

3 institutes

2500 UNIVERSITY STAFF MEMBERS

880 ASSOCIATE PROFESSORS

570 FULL PROFESSORS with doctoral degrees

73 MEMBERS OF THE RAS

MORE THAN **80** COUNTRIES REPRESENTED BY MOOCs STUDENTS

152 PARTNER UNIVERSITIES in 29 countries

JOINT PROGRAMS WITH COMPANIES:

Baker Huges, OSCiAI, Intel, Yandex, 2GIS, Academpark, Biotechnopark, Sberbank Technologies

35 PARTNER INSTITUTES in Academgorodok

139 LEADING RESEARCHERS

H>20

RESEARCH GROUPS ARE INVOLVED IN NSU **38** INTERNATIONAL COLLABORATIONS

140 LABORATORIES

6 MIRROR LABORATORIES

5-100 LEADER OF TOP 5-100 PROJECT –
the project oriented at making leading Russian universities competitive on world market of education services

TOP-100 IN PHYSICS
QS University Rankings

TOP-500 ACADEMIC RANKING OF WORLD UNIVERSITIES
3rd position among Russian universities

12th POSITION
QS UNIVERSITY RANKINGS BRICS

2nd POSITION
QS UNIVERSITY RANKINGS EMERGING EUROPE AND CENTRAL ASIA

68th POSITION
QS UNIVERSITY RANKINGS NATURAL SCIENCES

190th POSITION
GOOGLE SCHOLAR CITATIONS

244th POSITION
QS UNIVERSITY RANKINGS
among 20 000 world universities

N*

Contents

DOI 10.18699/SBB-2019-01

High-throughput sequencing techniques to flax genetics and breeding: a systematic review

Akhmetshina A.O.[1]*, Khlestkina E.K.[1, 2]
[1] *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
** e-mail: a.akhmetshina@vir.nw.ru*

Key words: databases, genes, genome, markers, RNA-seq, microRNA-seq, sequencing, transcriptome, *Línum usitatíssimum,* flax

Motivation and Aim: in recent years, DNA markers for flax (*Línum usitatíssimum*) breeding have been actively used. Marker-assisted selection (MAS), as well as genomic selection and new generation sequencing (NGS), require extensive knowledge of flax genome. The aim of this analysis was to carry out systematic review on the application of NGS methods in flax genetics and breeding.
Methods and Algorithms: the search was performed in Scopus database by the different combinations of the following keywords (flax, linseed, marker, QTL, high throughput sequencing. genomic selection, GBS, GWAS, RNA-seq, miRNA-seq, resistance, fiber content, yield, fatty acid composition, tolerance) in article titles, abstracts and key-words.
Results: according to Scopus database, 304 articles described different flax breeding directions as fiber flax and linseed were revealed. The articles using NGS in flax (Linum) research increased from 2 in 2012 to 6 in 20186. A large number of SNPs were identified thereby and approbation of method "Genotyping by sequencing" on flax became possible. The obtained data was the basis for the identification of new economically valuable genes in flax by genome-wide association studies (GWAS) and for the genomic selection. The use of DNA markers in flax genetics and breeding is more advanced for linseed than fiber flax. For both crops technological properties (fatty acid composition for linseed and fiber quality for fiber flax) are among most frequently studied with DNA markers (45 and 16 articles respectively). Then, DNA markers for resistance genes are more frequently developed and used (25 and 11 articles respectively). The transcriptome and miRNA are mostly related with abiotic stress tolerance of flax (drought, salinity and imbalance of nutrients). However, the flax resistance to biotic factors was not widely presented in omics studies.
Conclusion: The fiber flax remains under-investigated in comparison with linseed in development and use of DNA markers diagnostic for economically valuable traits. Studies on flax quality and disease resistance are mainly focused on finding and tagging genomic loci determining these traits, while investigation of abiotic stress tolerance is focused on identification of metabolic networks and regulatory factors involved in tolerance, based on omics studies. The systematized results can be useful for flax geneticists and breeders.

DOI 10.18699/SBB-2019-02

Molecular dynamics modeling of multicolor FRET-experiment

Andryushchenko V.A.
Institute of Thermophysics, SB RAS, Novosibirsk, Russia
** e-mail: vladimir.andryushchenko@gmail.com*

Key words: protein folding, FRET spectroscopy, molecular dynamics, free energy surfaces

Motivation and Aim: The Förster resonance energy transfer (FRET) method has become a powerful tool for studying protein dynamics. Currently, the donor and acceptor used in experiments usually attached to the ends of the protein chain [1]. Results commonly presented as a one-dimensional (1D) free energy profile. However, since protein dynamics are very complex, such 1D profiles do not provide a sufficiently detailed description of the styling process. At the same time, the donor and acceptor can be attached not only to the ends of the protein but to the inner parts of the protein chain (a multicolor experiment) [2]. In this case, two-dimensional free energy surfaces (FES) can be constructed that provide incomparably richer information about the folding process than 1D profiles [3]. Naturally, the question arises as to the optimal arrangement of the minimum amount of fluorophores in order to obtain the best possible information on protein folding.
Methods and Algorithms: Since the main goal was to understand the situation in general, a coarse-grained protein representation was used, i.e., a bead placed at the position of the Cα-atom represented each protein residue. The simulations performed with molecular dynamics methods. Selected residues (beads) represented the donor and acceptor. Two fluorophores attached to the ends of the protein and another inside the protein chain. The site of attachment of the last fluorophore was changing.
Results: The BBL domain protein folding has been modeling for different locations of the fluorophores along the protein chain. For all cases, two-dimensional free-energy surfaces and radial distribution functions, characterizing the relative position of the fluorophores relative to each other and the protein chain were constructed. On the base of the obtained data, the optimal location of fluorophores was determined, which made possible to reproduce the characteristic states of a protein in the space of distances between fluorophores.
Conclusion: The data available in the experiments i.e. distances between fluorophores (with a suitable arrangement of the three fluorophores) is enough to recover information on protein states similar to the information available in the modeling of protein folding in "usual variables", such as the radius of gyration and RMSD.
Acknowledgements: Supported by the RFBR (18-04-00013).

References

1. Neuweiler H., Johnson C.M., Fersht A.R. Direct observation of ultrafast folding and denatured state dynamics in single protein molecules. *PNAS*. 2009;106(44):18569-18574.
2. Lerner E. et al. Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. S*cience*. 2018;359(6373):eaan1133.
3. Andryushchenko V.A., Chekmarev S.F. Modeling of multicolor single-molecule forster resonance energy-transfer experiments on protein folding. *Journal Physical Chemistry B*. 2018;122(47):10678-10685.

DOI 10.18699/SBB-2019-03

Estimation of 12S marker fragment effectiveness for ancient phylogeny reconstruction

Bolbat A.[1]*, Sorokovikova N.[1], Kaygorodova I.[1, 2]
[1] *Limnological Institute, SB RAS, Irkutsk, Russia*
[2] *Irkutsk State University, Irkutsk, Russia*
** e-mail: bolbatav@lin.irk.ru*

Key words: phylogenetics, 12S, mitochondrial DNA, ribosomal RNA, Maximum Likelihood, Bayesian Inference, Annelida

Motivation and Aim: For many years different biopolymer sequences were used to infer the relations of various taxa. Yet few studies were devoted to the estimation of these sequences' effectiveness in phylogeny reconstruction at different taxonomic levels. One of widely used marker fragments is mitochondrial 12S rRNA gene [1]. In present study we tried to determine at which taxonomic levels the use of this gene is justified using Annelida worms as a model taxon.
Methods and Algorithms: We used a set of Annelida 12S gene sequences representing phylogenetic relationships of at various taxonomic levels – from species to class. The sequences were aligned with Clustal Omega globally and with MUSCLE locally. The phylogeny was reconstructed by Maximum Likelihood method implemented in MEGA7 [2] and Bayesian Inference method implemented in BEAST2 [3].
Results: Two different methods of phylogeny reconstruction – Maximum Likelihood and Bayesian Inference – yielded overall similar but slightly different topologies. The bootstrap node supports produced by Maximum Likelihood reconstruction were generally lower than Bayesian posterior supports which may be due to a loss of information during bootstrap resampling [4]. However both methods showed a significant decrease of node support values from tips to root of the trees.
Conclusion: The results suggest the high value of 12S rRNA gene sequences in reconstructing phylogeny at species and genus levels. However, these sequences perform poorly for reconstructing phylogeny at higher levels and can only be used in addition to other gene fragments.
Acknowledgements: The authors are thankful to Viktor Tarakanov for assistance in sample collection. The sample collection was funded by state assignment 0345-2016-0002 (AAAA-A16-116122110066-1) "Molecular ecology and evolution of living systems of Central Asia in the conditions of global ecological changes" Sample analysis was supported by RFBR grant No. 17-29-00097.

References

1. Patwardhan A., Ray S., Roy A. Molecular markers in phylogenetic studies – A review. *J. Phylogen. Evol. Biol*. 2014;2(2):131.
2. Kumar S., Stecher G., Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol*. 2016;33(7):1870-1874.
3. Bouckaert R. et al. BEAST 2: A software platform for bayesian evolutionary analysis. *PLoS Comput. Biol*. 2014;10(4):e1003537.
4. García-Sandoval R. Why some clades have low bootstrap frequencies and high Bayesian posterior probabilities. *Israel Journal Ecol. Evol*. 2014;60(1):41-44.

DOI 10.18699/SBB-2019-04

DNA barcode-based delimitation of the *Glossiphonia* species

Bolbat N.[1]*, Bolbat A.[2], Kaygorodova I.[1, 2]
[1] *Irkutsk State University, Irkutsk, Russia*
[2] *Limnological Institute, SB RAS, Irkutsk, Russia*
** e-mail: mandzyak@lin.irk.ru*

Key words: species delimitation, molecular taxonomy, *Glossiphonia,* Eastern Siberia

Motivation and Aim: Our knowledge about global hirudofauna is getting continuously enriched, however the taxonomic uncertainty of the *Glossiphonia* genus still persists. To resolve the issues associated with the determination of species status, molecular and bioinformational methods must be integrated into the practice of taxonomic research along with classical analysis of morphological features. In present study we utilized a combination of different approaches to solve the problem of leech species delimitation within the *Glossiphonia* genus.
Methods and Algorithms: Morphological analysis was carried out in accordance with the existent classification [1, 2]. Molecular analysis was performed by the standard phenol-free extraction and the subsequent amplification with universal COI primers. Phylogenetic analysis was performed with the use of Maximum Likelihood (ML) method implemented in MEGA 7.0 [3], to determine the threshold value of intraspecific genetic variations we used GMYC method based on the analysts of ultrametric tree obtained from Bayesian inference (BI) in BEAST v.1.8.4 [4]. Additionally, ABGD program was used to determine the barcode gap [5].
Results: The results of morphological, molecular and 4 bioinformational methods (ML, BI, GMYC и ABGD) indicate the independent taxonomic status of at least eight species within the *Glossiphonia* genus, including two Eastern Siberian species – *G. verrucata* and *Glossiphonia* sp.
Conclusion: The combination of results of present study points out the increased objectivity of biodiversity estimation within the *Glossiphonia* leech genus in case of implementing classical methods in conjunction with molecular, phylogenetic and bioinformational methods of analysis.
Acknowledgements: The study was performed in frame of the state assignment No. 0345-2016-0002 (AAAA-A16-116122110066-1), and in part of the Baikal fauna investigation was financially supported by the Russian Foundation for Basic Research (RFBR), project No. 17-2900097.

References

1. Lukin E.I. Leech of fresh and brackish waters. Fauna of the USSR. Leningrad: Science Publ., 1976.
2. Nesemann H., Neubert E. Clitellata, Branchiobdellada, Acanthobdellada, Hirudinea. J. Schwoebel and P. Zwig (eds.). *Susswasser fauna von Mitteleuropa*. 1999;6(2):1-178.
3. Kumar S., Stecher G., Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol*. 2016;33(7):1870-1874.
4. Drummond A.J., Suchard M.A., Xie D., Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol*. 2012;29:1969-1973.
5. Puillandre N, Lambert A, Brouillet S, Achaz G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol*. 2012;21:1864-1877.

DOI 10.18699/SBB-2019-05

Transcriptome assembling of non-model organisms on example holothuria Eupentacta fraudatrix

Boyko A.*, Dolmatov I.
National Scientific Center of Marine Biology of FEB RAS, Vladivostok, Russia
** e-mail: alteroldis@gmail.com*

Key words: transcriptome assembling, non-model orginasms, cardiac mechanics

Motivation and Aim: Over the past 5 years, more and more work on high-throughput sequencing of transcriptomes of non-model organisms has appeared. In such studies, there are many problems, for example a huge number of sequences in the assembly and complexity of predicted protein annotations. In connection with the study of the mechanisms of regeneration and development of echinoderms in our laboratory, we also encountered this problem.
Methods and Algorithms: Over the 400 millions clean reads were assembled using SPAdes 3.13[1] with 2 iterations for read correction step and with k-mer length of 25, 33 and 49. Of all the obtained contigs the Coding Sequences (CDSs) were extracted using TransDecoder 5.5.0[2]. The code of TransDecoder was modified in such a way that stop-codon or the beginning of the sequence, but not "ATG" (Met), was taken for the beginning of CDS. All CDSs were verified using BLAST-search by the SwissProt and Echinobase. Then all the obtained sequences were clustered into CD-HIT 4.7 [5] with three iterations. After each iteration, sequences in the clusters were assembled with an identity threshold of 80 %, using the own Python script, defined by us as HomoloCAP3. This script is a software add-on to CAP3[6] that makes it possible to use the data of pre-clustering of sequences and automatically select CAP3 parameters such as overlap and gap lengths and clipping range.
Results: The first stage of assembling in SPAdes resulted in a total of 703,169 contigs. This was unsatisfactory, as the level of fragmentation, the percentage of redundancy of almost identical contigs was high. Apparently, this situation arose due to the variability came from 5` and 3` untranslated regions. For this reason it was decided to use only CDSs for further assembling. As a result of the clustering and assembling with HomoloCAP3, filtering of the contaminant sequences and subsequent clustering with the aim to identify isoforms, we obtained a total of 85,805 contigs and 72,204 genes.
Conclusion: This approach to finalizing the assembly has been used for the first time and can significantly reduce the number of sequences with simultaneous increase in the number of full-length transcripts.

References

1. Bankevich A. et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol*. 2012;19:455-477.
2. Haas B.J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*. 2013;8(8):1494-512.
3. Fu L. et al. CD-HIT : accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150-3152.
4. Huang X., Madan A. CAP3: A DNA Sequence Assembly Program. *Gen. Res*. 1999;9:868-877.

DOI 10.18699/SBB-2019-06

Development of Tdp1 inhibitors based on natural biologically active compounds as prototypes of antitumor drugs

Chepanova A.A.[1]*, Zakharenko A.L.[1], Filimonov A.S.[2], Khomenko T.M.[2], Mozhaitsev E.S.[2], Ponomarev K.Yu.[2], Munkuev A.A.[2], Suslov E.V.[2], Luzina O.A.[2], Zakharova O.D.[1], Volcho K.P.[2], Salakhutdinov N.F.[2], Lavrik O.I.[1]
[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*
[2] *Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia*
** e-mail: arinachepanova@mail.ru*

Key words: tyrosyl-DNAphosphodiesterase I, inhibitor of tyrosyl-DNA phosphodiesterase I

Motivation and Aim: Developing inhibitors for DNA repair enzymes is a promising approach to improve anticancer therapy, in particular for drug-resistant tumors. Tdp1 plays a key role in the repair of Top1-DNA covalent complexes formed by topoisomerase-1 (Top1) poisons such as drugs of the camptothecin group, which makes it a promising target in the treatment of cancer [1]. The development of new inhibitors of DNA repair enzymes based on natural compounds and their derivatives is particularly relevant, since such compounds often have complementarity to targets of biological origin and possess broad range of biological activities.
Methods and Algorithms: Screening of compounds – potential inhibitors is carried out using a real-time fluorescence measurement method, which makes it possible to determine the initial reaction rate with high accuracy [2]. The effect of the selected compounds on the proliferation of transplanted tumor cell lines and the evaluation of cell death are studied using the MTT test. We used derivatives of usnic acid, coumarin and adamantane for Tdp1 inhibitors screening.
Results: All studied classes of compounds have a pronounced inhibitory effect. The dependence of the inhibitory activity on the structure of the compounds was revealed. Among the inhibitors found, there are both moderately toxic and non-toxic compounds. Usnic acid, coumarin, and adamantan derivatives proved to be sensitizers of the tumor cell lines to Top1 poison topotecan, with twofold enhancement of cytotoxicity of topotecan [3–6].
Conclusion: The compounds studied can be used to develop on their basis effective sensitizers of malignant cells to Top1 poisons. The therapeutic effect of such substances can be a selective increase in the activity of Top1 poisons in tumors. Non-toxic inhibitors of Tdp1 are of particular interest, since will avoid additional side effects.
Acknowledgements: Supported by RFBR (grant 18-44-540023) and by RSF (grant 19-13-00040).

References

1. Yang S.W. et al. *Proc. Natl. Acad. Sci. USA*. 1996;93:11534-11539.
2. Zakharenko A.L. et al. *Bioorg. Med. Chem*. 2015;9:2044-2052.
3. Khomenko T.M. et al. *Bioorg. Med. Chem*. 2016;24:5573-5581.
4. Ponomarev K.Yu., et al. *Bioorg. Med. Chem*. 2018;76:392-399.
5. Zakharova O. et al. *Bioorg. Med. Chem*. 2018;26(15):4470-4480.
6. Zakharenko A.L. et al. *Eur. J. Med. Chem*. 2019;161:581-593.

DOI 10.18699/SBB-2019-07

Computer tools for spatial chromosome contacts analysis by ChIA-PET and Hi-C data

Dergilev A.I.[1]*, Luzin A.N.[1], Kovalev S.S.[2], Babenko R.O.[1], Orlov Y.L.[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
** e-mail: arturd1993@yandex.ru*

Key words: genome, transcription factor, medical informatics, ChIP-seq

Motivation and Aim: 3D contacts and interactions in interphase nuclei of eukaryotic cell play critical role for gene expression regulation. Series of post-genome technologies have been developed to study the binding of transcription factors for transcription regulation, such as chromatin immunoprecipitation arrays (ChIP-Seq) [1]. Correspondingly, set of software tool for processing of such data has been developed. 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 [2]. The aim of the work was to review existing tools for 3D genome structure and spatial topological domains analysis.
Methods and Algorithms: The data have been obtained via available data sources containing experimental information from ChIP-seq, Hi-C, ChIA-PET tests using different sequencing platforms. Gene annotation was obtained from UCSC Genome Browser. We reviewed existing software and created a database prototype of bioinformatics tools for 3D genome structure analysis.
Results: We tested program for analysis of ChIA-PET experimental data. 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. ChIA-PET and Hi-C technologies provide huge volume of data demanding development of new computer tools in different applications.
Acknowledgements: The work was supported by the RFBR and ICG SB RAS budget project (0324-0019-0040).

References

1. Li G. et al. Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. *BMC Genomics*. 2014;15(Suppl. 12):S11.
2. Fullwood M.J. et al. An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature.* 2009;462(7269):58-64.

DOI 10.18699/SBB-2019-08

Statistical analysis, clusterization and visualization of genome distribution of transcription factor binding sites

Dergilev A.I.[1]*, Tsukanov A.V.[1, 2], Luzin A.N.[1], Babenko R.O.[1], Orlov Y.L.[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
** e-mail: arturd1993@yandex.ru*

Key words: genome, transcription factor, medical informatics, ChIP-seq

Motivation and Aim: The analysis of gene transcription regulation based on the data of modern technologies of high-performance sequencing is an actual task of bioinformatics [1]. It requires the development of new computer tools including supercomputer applications. We consider the problems of processing of genome ChIP-seq profiles for detections of transcription factors binding site in a genome, determining the peaks of such profiles and search the binding sites in the nucleotide sequences of the peaks.
Methods and Algorithms: The computer programs have been developed to analyze the location of the binding sites in the genome relative to gene regions, to calculate clusters of such sites and visualize their positions in the genome. Clusters of binding sites of transcription factors in the human genome have been calculated using the Cistrome database.
Results: We have calculated matrices of the joint occurrence of pairs of binding sites of different transcription factors in the genome for various types of tissues and cells. A computational experiment on the computer generation of random clusters in the genome was carried out, as well as an assessment of the occurrence of large clusters for experimentally obtained binding sites of transcription factors in the human genome. The patterns of occurrence of binding sites of pluripotency factors in embryonic stem cells were described. The developed software is available on request to the authors.
Conclusion: Problem of analysis of genome distribution of transcription factor binding sites in human genome is of importance for personalized medicine and genomics studies.
Acknowledgements: The work was supported by the RFBR and ICG SB RAS budget project (0324-0019-0040).

References

1. Spitsina A.M., Bragin A.O., Dergilev A.I., Chadaeva I.V., Tverdokhleb N.N., Galiyeva E.R., Tabikhanova L.E., Orlov Y.L. Computer tools for analysis of transcriptomics data: program complex ExpGene. *Program Systems: Theory Applications*. 2017;8:2(33):45–68. DOI 10.25209/2079-3316-2017-8-2-45-68. (in Russian)

DOI 10.18699/SBB-2019-09

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

Devyatkin V.[1, 2]*, Redina O.[1], Muraleva N.[1], Kolosova N.[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, SNPs

Motivation and Aim: Aging is the largest risk factor for many diseases. However, little is known about a genetic overlap between age-related diseases. The senescence-accelerated OXYS rats selected in the ICG SB RAS (Novosibirsk) are a unique model to identify the pathways that modulate the onset and progression of multiple age-related diseases as these rats develop a phenotype similar to human geriatric disorders including cataract, cardiomyopathy, hypertension, retinopathy and neurodegenerative pathology of the brain with features of Alzheimer's disease. The aim of our work was to investigate the transcriptome of OXYS rats and to identify the mutations (SNPs) in genes, which can potentially contribute to the development of accelerated aging.
Methods and Algorithms: The RNA-Seq data were obtained from sequencing of prefrontal cortex, retina and hippocampus of senescence-accelerated OXYS 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. The mutation was considered as reliable SNP if it was detected in at least 3 OXYS rats in homozygous state and coverage (DP) was ≥ 10 in at least one animal. The effect of an amino acid substitution on protein function was predicted by the Variant Effect Predictor Web service (https://www.ensembl.org/vep); the consequence type, SIFT score and prediction were obtained for each variant. The SNPs found in OXYS rats were compared with the data of genomic sequencing of 42 other rat strains, 11 of which simulate different forms of hypertension, 10 strains are used as a normotensive control, and the rest are used as experimental or control strains to study diseases not related to hypertension or aging.
Results: In the genome of OXYS rats 42478 SNPs overlapping with 9903 genes were detected. Of these, 725 SNPs were found both in OXYS rats and in one or several other hypertensive rat strains, and 1380 SNPs were specific to OXYS rats. These 2105 SNPs can contribute to the development of the accelerated aging phenotype; of these, 9 SNPs can lead to significant structural rearrangements of transcripts, and 33 SNPs probably have a significant effect on the structure or function of the protein products. Several of these SNPs are mapped to genes associated with mitochondrial and neurodegenerative diseases, mental and cardiovascular disorders, which is consistent with the complex manifestation of the senile phenotype in OXYS rats.
Conclusion: The results of the study may serve as a background for further verification of SNPs contribution to the development of complex age-related diseases in OXYS and other rat strains as well as in conducting GWAS in humans.
Acknowledgements: Supported by the RFBR (project No. 18-015-00336, No. 18-315-00216).

DOI 10.18699/SBB-2019-10

Molecular processes in *Solanum phureja* roots in response to *Globodera rostochiensis* infection

Egorova A.A.[1, 2]*, Shmakov N.A.[1], Gerasimova S.V.[1], Vasilyev G.V.[1], Shatskaya N.V.[1], Strygina K.V.[1], Afonnikov D.A.[1], Kochetov A.V.[1, 2]

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

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

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

Key words: potato, RNA-seq, *Solanum phureja*, resistance, *Globodera rostochiensis*

Motivation and Aim: Golden potato cyst nematode (GPCN) is an important pathogen of potatoes, tomatoes, and other plants in the family *Solanaceae*. GPCN juveniles damages the roots of susceptible plants. Nowadays GPCN is found worldwide and it appears to be the one of the most deleterious pathogens for potato. Protection against GPCN is complicated because the eggs of nematodes can remain viable in the soil for more than 30 years. In addition, most chemical nematicides are not efficient, thus the protection is mostly obtained through the introduction of the resistance cultivars.

In this study, we analyzed the genetic mechanisms of resistance of diploid potato *Solanum phureja* in response to GPCN through comparison of root transcriptomes of resistant and susceptible genotypes.

Methods and Algorithms: For RNA-seq, total RNA was extracted from root samples collected in time points 0, 24 and 72 hours after inoculation with GPCN. Sample preparation was carried out by colleagues from Vavilov Institute of Plant Genetic Resources and All Russian Research Institute for Plant Protection (St. Petersburg, Russia) according to [1]. Sequencing was performed on Illumina NextSeq 500 platform. FastQC and Prinseq tools were used to assess sequences quality and filter the libraries. STAR and TopHat were used to map the filtered libraries to the reference genome. Search for the differentially expressed genes was performed using Cufflinks pipeline and EdgeR package for R. Lists of differentially expressed genes (DEGs) were further analyzed with Biomart and the databases AgriGO, KEGG, and PlantCyc.

Results: Analysis of *S. phureja* transcriptomic data revealed differential expression of a number of genes in response to nematode invasion. In resistant genotype, we observed increased expression of genes associated with ROS production, photosynthesis, PR proteins, peptidase inhibitors activity and suppression of genes, associated with mevalonate pathway. In susceptible genotype, we observed increased expression of genes associated with auxin regulation, cell division, and mevalonate pathway.

Conclusion: We revealed candidate genes, associated with complex resistant mechanism and formation of syncytium (feeding cell) by transcriptome analysis of two *S. phureja* genotypes contrasting in GPCN resistance.

Acknowledgements: Supported by grant from RSF No. 16-16-04073.

References

1. Kochetov A.V. at al. Differential expression of NBS-LRR-encoding genes in the root transcriptomes of two Solanum phureja genotypes with contrasting resistance to *Globodera rostochiensis*. *BMC Plant Biol*. 2017;17(Suppl. 2):251.

DOI 10.18699/SBB-2019-11

MiRNA-directed gene activation during tuberculosis infection

Elizarova A.[1, 2]*, Medvedeva Y.A.[1, 2]
[1] *Institute of Bioengineering, Research Center of Biotechnology RAS, Moscow, Russia*
[2] *Moscow Institute of Physics and Technology (State University), Moscow, Russia*
** e-mail: elizarova@phystech.edu*

Key words: miRNAs, macrophages, tuberculosis, enhancers

Motivation and Aim: Tuberculosis pathogen, Mycobacterium tuberculosis (Mtb), remains successful due to the ability to get through adverse macrophage environment. MiRNAs play a role in the response to Mtb infection. Yet only regulation via mRNA repression has been investigated so far. Recent studies show that miRNAs act in an unconventional way – by activating enhancers with miRNA seed [1]. We wonder if such regulation is possible during Mtb infection.
Methods and Algorithms: We sought for enhancer-miRNA-gene trios assuming that a miRNA up-regulates a gene by targeting its enhancer. We used a time-course expression dataset from [2] and enhancer-gene interactome from [3]. Since the exact mechanism of enhancer regulation by miRNA remains unclear, we used various strategies to determine enhancers possibly targeted by differentially expressed (DE) miRNAs: we examined if miRNA could form either duplex or triplex with an enhancer sequence. To look for miRNA:enhancer duplexes we either (1) selected active enhancers containing seed sequence of any DE miRNA and high identity of the rest miRNA to surrounding DNA (Needle); or (2) sought for miRNA targets with MiRanda. Additionally, we predicted RNA:DNA triplex formation (Triplexator). Then, Spearman correlation coefficient was calculated for each DE miRNA-gene pair for genes with enhancer targeted by a miRNA.
Results: Among highly correlated trios we found a small regulatory network of Klf6 and BC016423, regulated by the same enhancer via duplex formation with miR-22 and miR-221. Klf6 could also be regulated by miRNA miR-221 via RNA:DNA triplex formation with another enhancer. KLF6 is a transcription factor essential for macrophage motility and plays an important role in the regulation of macrophage polarization promoting M1 phenotype.
Conclusion: We conclude that miRNA-directed activation of enhancers and their target genes is possible in Mtb infection at least for several DE miRNAs. Yet experimental validation is crucial (such as enhancer activity assays and chemiluminescent EMSA for validation of miRNA-DNA interaction).
Acknowledgements: This work is supported by the RFBR (17-54-80033).

References

1. Xiao M., et al. MicroRNAs activate gene transcription epigenetically as an enhancer trigger. *RNA Biol*. 2017;14:1326-1334.2.
2. Roy S. et al. Transcriptional landscape of Mycobacterium tuberculosis infection in macrophages. *Sci Rep*. 2018;8:6758.3.
3. Denisenko E., et al. Transcriptionally induced enhancers in the macrophage immune response to *Mycobacterium tuberculosis* infection. *BMC Genomics*. 2019;20:71.

DOI 10.18699/SBB-2019-12

Expression and role of hox genes in the regeneration of internal organs in the sea cucumber *Eupentacta fraudatrix*

Garipova V.A.[1, 2]*, Dolmatov I.Y.[1, 2]
[1] *A.V. Zhirmunsky Institute of Marine Biology, NSCMB, FEB RAS, Vladivostok, Russia*
[2] *Far Eastern Federal University, Vladivostok, Russia*
** e-mail: garipulia@mail.ru*

Key words: Hox genes, Echinodermata, regeneration, gene expression, *in situ* hybridization

Motivation and Aim: One of the most important adaptations for the survival of the organism is such a biological phenomenon as the ability to regenerate. Regeneration is widespread among animals and can be found in almost every large taxon. However, the potential for regeneration is different for everyone. Representatives of the Echinodermata type are distinguished by pronounced regenerative possibilities [1]. The sea cucumber *Eupentacta fraudatrix* was selected as the object of study because it is an excellent model for studying the mechanisms of regeneration. Many developmental genes play an important role in regeneration. Thus, the *hox* genes play a fundamental role in the correct formation of organs along the anteroposterior axis of multicellular bodies, such as worms, insects, vertebrates and echinoderms [2]. The participation of *hox* genes in the Holothuroidea class during regeneration is poorly understood. In this regard, the purpose of this work is to study the expression of *hox* genes in the process of regeneration of the internal organs of the sea cucumber *E. fraudatrix*.
Methods and Algorithms: A gene expression was estimated with qPCR, using *ef1α* and *tubulin* as reference genes. Gene identification was carried out using the construction of a phylogenetic tree. In order to construct RNA probes for *in situ* hybridization, the necessary sequences were produced. Hybridization was performed using antisense and sense (as control) RNA probes. AK and intestine, as well as AK rudiments, intestines, and body wall were subjected to hybridization at different stages of *E. fraudatrix* regeneration.
Results: We determined of 8 complete *hox* gene transcripts. Of them, two are anterior genes(hox1, hox3), three medial (*hox5*, *hox7*, *hox8*) and three back genes (*hox9/10*, *hox11/13a*, *hox11/13c*). With the regeneration of internal organs, the greatest change in activity is observed for the *hox5*, *hox7*, *hox9/10* and *hox11/13a* genes. The expression of the *hox5* and *hox9/10* genes was maintained above the level of the norm at all investigated stages of regeneration, *hox11/13a* had the highest level of activity on the 5th and 7th day after evisceration. The activity of the *hox7* and *hox9/10* genes is most pronounced in the structures of the ambulacral AK system. At the same time, these genes show spatial collinearity of expression. Observed expression of the studied genes of the Hox family in the sea cucumber *E. fraudatrix* shows their participation in the regulation of regenerative processes.

References

1. Dolmatov I.Y., Mashanov V.S. Regeneration in holothurians. – Vladivostok: Dalnauka. 2007;212.
2. Gaunt S.J. Hox cluster genes and collinearities throughout the tree of animal life. *Int. J. Dev. Biol*. 2018;62(11-12):673-683.

DOI 10.18699/SBB-2019-13

Gene networks in a post-genomic era

Ignatieva E.V.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
e-mail: eignat@bionet.nsc.ru

Key words: gene networks, feedback, transcription regulatory network, PPIs, SNPs, human diseases

Gene networks are molecular genetic systems that determine phenotypic characteristics of organisms (molecular, biochemical, structural, morphological, behavioral, etc.) using information encoded in the genomes of these organisms [1]. The essential types of structural and functional components of gene networks are: (1) genes; (2) RNA and proteins encoded by genes; (3) low molecular weight components (metabolites, energy transfer molecules, steroid hormones, cations, anions, etc.); (4) negative and positive feedbacks stabilizing the parameters of the gene network at a certain level, or, on the contrary, deviating them from the initial value, providing a transition to a new functional state [2]. Reconstruction of gene networks and deciphering their connectivity structure provide methodological basis for modern systems biology [3]. This review (1) introduces the concepts of *Gene networks* and related terms used in the post-genomic era (*Gene regulatory networks* (GRNs), *transcription regulatory networks* (TRNs), *protein-protein interaction* (PPI) *networks*, *associative gene networks*, etc.); (2) gives an examples of the hierarchical organization of the gene networks (the levels cell, tissue, the whole body, communities of organisms); (3) defines the role of the feedback mechanisms operating at different levels; (4) lists several computer and/or experimental approaches used for gene networks reconstruction; (5) characterizes a number of Internet accessible information resources oriented to humans and animals and containing data on gene networks and their functional modules; (6) refers to computer systems that allow the user to construct molecular-genetic networks using the data automatically extracted from the texts of scientific publications or data on experimentally defined binary interactions between the objects in the network. Applications of network and pathway-based methodology to explore pathogenetic mechanisms underlying human diseases at a systems biology level are also provided.

Acknowledgements: Supported by the State Budgeted Project No. 0324-2019-0040.

References

1. Ananko E.A., Aksenovich T.I., Gunbin K.V. et al. Introduction to informational biology and bioinformatics: study guide: eds N.A. Kolchanov, O.V. Vishnevsky, D.P. Furman. V. 3: Chapter 3: Gene Network Theory; Chapter 4: Mapping of genes controlling complex human traits. – Novosibirsk: 2015 (in Russian).
2. Kolchanov N.A. et al. Gene Networks. *Vavilov Journal Genetics Breeding*. 2013;17(4/2):833-850.
3. Ignatieva E.V. et al. Online resources on gene networks containing human and animal data. *Vavilov Journal Genetics Breeding*. 2017;21(8):895-902.

DOI 10.18699/SBB-2019-14

Reconstruction of the gene networks of human neurotransmitter systems

Ivanov R.A.[1, 2]*, Klimenko A.I.[1, 2], Savostyanov A.N.[1, 2, 3], Lashin S.A.[1, 2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Physiology and Basic Medicine, Novosibirsk, Russia*
** e-mail: ivanovromanart@bionet.nsc.ru*

Key words: gene networks, SNP, social behavior, depression, neurotransmitters

Motivation and Aim: The main interest of a modern neuroscience is aimed at studying the biological bases of individual differences in human behavior. The medical relevance of this topic is due to the fact that development of a number of neurological and psychiatric diseases, as well as the characteristics of their occurrence, are directly related to the psychological characteristics of individuals [1]. Many papers on the mechanisms of regulation of individual behavior indicate the important role of neurotransmitter systems in the mechanisms of its regulation [2]. Today gene network approaches are powerful tools for studying genes associated with human traits.
Methods and Algorithms: In our research we focused on studying the genetics factors of anxiety behavior and depression. Selection of gene sets was formed based on an analysis of literature data, as well as a number of databases. Gene networks were reconstructed using Cytoscape based on interaction data extracted from GeneMANIA and StringDB.
Results: Using the methods mentioned earlier, we compiled a set of genes associated with anxiety behavior and set of genes associated with mental disorders for which the difference in phenotype manifestation among different human populations was shown. Based on this sets of genes, the gene networks were reconstructed.
Conclusion: Analysis of the gene networks is an effective tool for a comprehensive theoretical study of mechanisms of regulation of human behavior and it can be used to search for and prioritize genes associated with diseases.
Acknowledgements: The study was supported by the RFBR grant No. 18-29-13027 MK.

References

1. Knyazev G.G., Bocharov A.V., Levin E.A., Savostyanov A.N., Slobodskoj-Plusnin J.Y. Anxiety and oscillatory responses to emotional facial expressions. *Brain Res*. 2008;1227:174-188.
2. Ebstein R.P. The molecular genetic architecture of human personality: beyond self-report questionnaires. *Mol. Psychiatry*. 2006;11(5):427-445.

DOI 10.18699/SBB-2019-15

Isolation and identification of soil bacteria of the genus *Bacillus*

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

Key words: soil bacilli, isolation, identification, mass spectrometry

Motivation and Aim: Necessity to carry out indication of *Bacillus anthracis* not only in biomaterial, but also in environmental objects containing a great number of bacteria of this genus requires testing the specificity of applied methods, test systems and diagnostic preparations on a representative collection of strains of closely related saprophytes.
The aim of the work was to isolate soil representatives of the genus *Bacillus*, to carry out their identification and to store them in the form of a working collection for experimental and production purposes.
Methods and Algorithms: We studied 68 samples of soil from steppe areas, pastures, settlements, cattle-breeding farms of various areas of Stavropol Territory and neighboring republics. Liquid suspensions of soil were inoculated on Crome™ Bacillus Agar (HIMEDIA) and twenty-four hours later colonies differing from each other in their morphology and in the character of their change of medium were selected from the surface of the medium and subcultured on LB agar. Samples were prepared by lysis of 18-hour vegetative cultures in 80 % TFU with subsequent ultramicrocentrifuge filtration. Identification of cultures it was carried out by the method of MALDI TOF MS using the device Microflex LRF Bruker. For identification we used the commercial database of Bruker company (version Bruker Taxonomy V 7). Spores were grown on KG Agar (HIMEDIA) and stored in cryoprotector (Deltalab) at –20 ºC.
Results: 243 cultures of 20 species of the genus *Bacillus* have been isolated and identified: *B. antrophaeus*, *B. asahii*, *B. cereus*, *B. firmus*, *B. halotolerans*, *B. horneckiae*, *B. idriensis*, *B. licheniformis*, *B. marisflavi*, *B. megaterium*, *B. pseudomycoides*, *B. pumilis*, *B. siamensis*, *B. simplex*, *Bacillus sp.* (cluster of strains of *B.cereus* group the species of which were not identified), *B. sporothermodurans*, *B. subtilis*, *Lysinobacillus boronitolerans*, *Lysinobacillus fusiformis*, *Lysinobacillus sphaericus*, *Viridibacillus neidei*.
Conclusion: The isolated cultures may be used for testing the specificity of various test systems and preparations for the diagnosis of anthrax, and the database of mass spectra – for specific identification of representatives of the genus *Bacillus*.

References

1. Vilain S., Luo Y., Hildreth M.B., Brözel V.S. Analysis of the life cycle of the soil saprophyte *Bacillus cereus* in liquid soil extract and in soil. *Applied Environmental Microbiol*. 2006.
2. Rahi P., Prakash O., Shouche Y.S. Matrix-assisted laser desorption/ionization time-of-flight mass-spectrometry (MALDI-TOF MS) based microbial identifications: challenges and scopes for microbial ecologists front. *Microbiol*. 2016.

DOI 10.18699/SBB-2019-16

Early life stress leads to enhanced susceptibility to social stress in adult male C57BL/6 mice

Kisaretova P.E.[1, 2]*, Reshetnikov V.V.[1], Ershov N.I.[1], Bondar N.P.[1, 2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: kisaretova@bionet.nsc.ru*

Key words: early life stress, maternal separation, social defeat stress, H3K4me3, prefrontal cortex

Motivation and Aim: To this date human society collected a lot of empirical evidence on the capacity of early life surroundings to drive changes in cognition that persist through whole life of individual. In case of adverse experience, we see emotional and behavioral dysregulation. Failure to find appropriate treatment suggests our lack of understanding of processes responsible for these effects. Epigenetic mechanisms nowadays are the most promising thing to look into for the answers. We investigate the effects of adverse early life experience on susceptibility to social stress in adulthood on methylome of prefrontal cortex in male C57BL/6 mice.
Methods and Algorithms: Animals were separated into three groups: subjected to chronic social defeat stress (SS); prolonged maternal separation during early life period in combination with social stress in adulthood (MS + SS); normal rearing condition (NC). Statistical significance of behavioral tests was examined by one-way ANOVA. Prefrontal cortex tissue was used for mRNA and DNA extraction. Active promoter enrichment of DNA was accomplished using Native ChIP assay with H3K4me3 antibodies. ChIP-seq and RNA-seq were carried out on Illumina HiSeq4000 platform, using 4 animals per group. Differential expression analysis was performed by DESeq2 package on R. Results with p-value < 0.01 and fold change >1.25 were considered significant.
Results: Anxiety levels were increased in SS + MS group compared to NC control group ($p<0.05$). Social behavior was significantly reduced in SS ($p<0.01$) and MS + SS ($p<0.001$) groups compared to NC group. Using RNAseq analysis18 genes were found to be differentially upregulated simultaneously in SS and SS + MS groups. The most promising of them were chosen for further analysis with qPCR using extended animal groups. Homer1a, Bdnf, Npas2 confirmed to be differentially expressed. ChIPseq analysis resulted in identification of 135 differentially methylated regions in both SS and SS+MS groups compared to NC.
Conclusion: Early postnatal stress leads to enhanced outcomes of social stress in male C57BL/6 mice compared to group of animals subjected only to social stress. Genes found to be differentially expressed in experimental groups compared to control are associated with immune processes such as antigen presentation and T-cell mediated cytotoxicity. More detailed investigation of epigenetic profile is necessary for understanding molecular processes underlying susceptible or resilient phenotype enouncement.
Acknowledgements: Supported by the RFBR (18-34-00603).

DOI 10.18699/SBB-2019-17

Database on alternative gene splicing in glioma cell cultures

Kovalev S.S.[1]*, Gubanova N.V.[1], Lieberfarb E.Y.[1, 2], Galieva A.G.[3], Orlov Y.L.[1, 3]*
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State Medical University, Novosibirsk, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: sergey.kovalev.1994@list.ru;orlov@bionet.nsc.ru*

Key words: bioinformatics, transcriptomics, medical informatics, glioblastoma, alternative splicing, databases

Motivation and Aim: Fundamental biomedical research in oncology, the search for new markers of tumor development, modern post-genomic studies of gene expression on cell cultures need transcriptome profiling and analysis of gene expression and isoforms. Such experiments, in turn, require development of new computer tools and database for analysis of bulk sequencing data. The aim of our study is a computer search for genes and gene isoforms, the difference of their expression is associated with the development of glioblastoma [1].
Methods and Algorithms: The work is based on modern high-throughput sequencing technologies and international biomedical data banks analysis. RNA-seq data on glioma cell cultures were obtained at ICG SB RAS. The search for candidate genes in tumors for therapeutic treatment, including individual gene isoforms, is very relevant in healthcare and modern high-tech medicine.
Results: This work presents the bioinformatics problems related to the development of computer pipelines for the processing of transcriptomic data, the revealing of the differentially expressed genes, the analysis of alternative splicing, and the description of the gene ontologies categories for the genes sets found. The tasks of automatic search and description of gene functions in connection with cancer diseases, visualization of results and development of biomedical databases are considered. A prototype database of differential alternative splicing of genes is presented, "Differential Alternative Splicing of Human Genes in Secondary Glioblastoma (DASGG)", with the ability to work through a website, to search for expression levels of individual isoforms in tumor cells.
Conclusion: Problem of databases development for cancer studies are of great importance for personalized medicine. The results could be applied for fundamental research on glioma stem cells, search for new diagnostic markers.
Acknowledgements: Supported by the RFBR and ICG SB RAS budget project (0324-0019-0040).

References

1. Kovalev S.S., Lieberfarb E.Y., Gubanova N.V., Bragin A.O., Galieva A.G., Tsukanov A.V., Babenko V.N., Orlov Y.L. Computer analysis of gene alternative splicing in glioma cell cultures by RNA-seq data. *Vestnik NSU. Series: Information Technologies*. 2018;16(3):22-36. DOI 10.25205/1818-7900-2018-16-3-22-36 (in Russian).

DOI 10.18699/SBB-2019-18

Computer reconstruction of the ecological structure of the synthetic microbial community, which models core human gut microbiota

Kropochev A.[1, 2]*, Klimenko A.I.[1], Lashin S.A.[1, 2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: kropochev@bionet.nsc.ru*

Key words: trait-based ecology, ecological networks, synthetic microbial communities

Motivation and Aim: At present, environmental modeling experiences a shift from taxonomy-based analysis to functioning, since the trait-based approaches allow for a deeper mechanistic understanding of environmental processes. However, it should be noted that such approaches require careful study of the methodology and taking into account a large amount of heterogeneous information. Human gut microbiota is a complex microbial community that plays a significant role in preserving human's health. Despite human gut microbiota's being extensively studied, there is no structured source of information on ecological interactions of its key players. Therefore, at the first stage, our goal is the reconstruction of the ecological structure of the microbial community, which models core human gut microbiota. This will be the basis for the further construction of qualitative and quantitative ecological models using trait-based approaches.
Methods and Algorithms: For our study, we chose a synthetic microbial community from the large intestine comprised of the species that collectively possess its core metabolic capacities [1].
Primary reconstruction of the ecological structure includes:
1) Using @MInter text mining system [2] to identify microbial interactions in the community;
2) Expert analysis of the literature obtained by using text-mining algorithms and author search;
3) Using bioinformatics analysis to fill the gaps in knowledge obtained from the literature.
Results: We have qualitatively reconstructed the network of metabolic interactions between members of the microbial community, described the metabolism of glycans, short-chain fatty acids, amino acids, sulfates and other metabolites.
Conclusion: The reconstruction we obtained summarizes the accumulated knowledge on the metabolism of bacteria in our community and can be used as a basis for further model building.
Acknowledgement: The study was supported by the Budget Project 0324-2019-0040.

References

1. Desai M.S. et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell*. 2016;167(5):1339-1353.
2. Lim K.M.K. et al. @MInter: automated text-mining of microbial interactions. *Bioinformatics*. 2016;32(19):2981-2987.

DOI 10.18699/SBB-2019-19

Genetic mechanisms associated with determinate growth habit in cowpea (*Vigna unguiculata* (L.) Walp.)

Krylova E.A.[1]*, Burlyaeva M.O.[1], Khlestkina E.K.[1, 2]
[1] *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
** e-mail: e.krylova@vir.nw.ru*

Key words: determinant growth habit, cowpea, *Vigna unguiculata*, terminal flower

Motivation and Aim: Cowpea (*Vigna unguiculata* (L.) Walp.) is an important grain legume crop. It is known two types of growth habit that are classified as determinate and indeterminate. In wild species the growth habit is indeterminate and terminal shoot meristem of these plants remains in a vegetative state throughout the production of vegetative and reproductive structures. In plants with determinant growth habit, terminal shoot meristem switches from vegetative to reproductive state. It is known that *TFL1* is gene controlling determinate growth habit in many legume species. The aim of the present study was (1) to identify and characterize *TFL1* homologs in *Vigna* genome sequence and (2) to reveal allelic diversity of *TFL1* gene in *V. unguiculata* accessions with type of growth habit, sensitive and insensitive for different (dry and wet) environment conditions.
Methods and Algorithms: Based on the allelic differences of the *TFL1* gene described in the literature, DNA markers for distinguishing genotypes with different growth habit were developed. The homologous sequences of *TFL1* gene have been identified in *Vigna* Genome Server ('VigGS', http://viggs.dna.affrc.go.jp) using BLAST search. *In silico* analysis was performed with MEGA, FGENESH+ and PLACE software.
Results: Additional copies of the *TFL1* gene were found in *Vigna* genome. Phylogenetic analysis allowed to establish that the duplications occur in the common ancestor of legumes. 12 combinations of DNA markers were developed based on the known mutations of the *TFL1* gene and used for re-sequencing this gene in *V. unguiculata* accessions, sensitive and insensitive to changing type of growth habit in different environment conditions.
Conclusion: Comparison of *V. unguiculata* accessions with type of growth habit, sensitive and insensitive for different (dry and wet) environment conditions, suggest more complicate regulation of this trait than mechanism based on just *TFL1* allelic differences. Further comparative transcriptomic studies are needed.

DOI 10.18699/SBB-2019-20

The investigation of Poly(ADP-ribose)polymerase activity in the context of nucleosome

Kurgina T.[1, 2]*, Kutusov M.[1, 2], Belousova E.[2], Naumenko K.[2], Lavrik O.[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Yekaterinburg, Russia*
** e-mail: t.a.kurgina@gmail.com*

Key words: PARP1, DNA repair, nucleosome

Poly(ADP-ribose)polymerase 1 (PARP1) – is a key enzyme that regulates the activity of DNA-repair machine. It makes PARP1 inhibitors promising anticancer drugs [1]. But some problems occur during therapy by these agents, for example cancer resistance or recurrence of disease. These problems require not only novel inhibitor developing, but also fundamental studies of PARP1 and its interaction with other proteins. To address this need we have developed a test-system, which allows analyzing PARP 1 activity in real time [2]. In this work we modified this test-system to observe PARP 1 activity on nucleosomes *in vitro*. Nucleosome is a DNA-protein complex that consist 147 bp DNA is wrapped around histone octamer. This model close to in vivo conditions more than the short model DNA, used previously. We showed the ability of our method shed light on PARP1 interaction with nucleosome, nucleosome with DNA-lesion and different proteins on nucleosome. PARP1 DNA-binding capacity after poly(ADP-ribose)lation and degradation of PAR by several enzymes (for example PARG) was studied. It is shown that automodified PARP1 can effectively bind DNA and DNA in nucleosome context after degradation of PAR. However, PARylation activity of this enzyme is dramatically decreased after modification and modification removing. In addition, an interaction of PARP1 with several nuclei proteins on nucleosome was studied. Some of these proteins can modulate PARP1 activity and cause decreasing of PARP1 inhibitors efficacy. Results of these studies show that the expression level of these proteins in patients can influence the choice of anticancer therapy strategy.

Acknowledgements: The work was supported by the grant from RSF 17-74-20075.

References

1. Jain P.G., Patel B.D. Medicinal chemistry approaches of poly ADP-Ribose polymerase 1 (PARP1) inhibitors as anticancer agents. *European Journal Med. Chemistry*. 2019;1(165):198-215.
2. Kurgina T.A., Anarbaev R.O., Sukhanova M.V., Lavrik O.I. A rapid fluorescent method for the real-time measurement of poly(ADP-ribose) polymerase 1 activity. *Analytical Biochemistry*. 2018;15(545):91-97.

DOI 10.18699/SBB-2019-21

The mathematical modeling of oscillations in enzymatic systems

Lakhova T.N.[1]*, Kazantsev F.V.[1, 2], Lashin S.A.[1, 2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: tlakhova@bionet.nsc.ru*

Key words: mathematical model, feedback loop, cycle, oscillatory behavior, microbial enzymatic system

Motivation and Aim: The cyclic processes play essential role in the functioning of biological systems. A lot of such processes are controlled by enzymatic reactions, rates of which may be dramatically influenced by various inhibitors, activators, cofactors, allosteric regulators, etc. It is well-known that the oscillatory behavior in biological systems frequently emerges via influence of the feedback loops. Mathematical modeling is one of the main approaches for the investigation of such complicated systems. Now there are dozens open access databases that store information about biological pathways and Gene Networks that could be the source of data for models of enzymatic reactions. The aim of our research is to develop the technology that can help find potentially oscillating microbial enzymatic subsystems and to explore their behavior.
Methods and Algorithms: The Mammoth database [1] was used as an enzymatic reaction source. Data processing and analysis of structural models were performed with Python. Visual analysis of the structure models was carried out with Cytoscape (https://cytoscape.org/). The mathematical models were reconstructed in terms of ordinary differential equations and computer simulations were processed with Copasi (http://copasi.org).
Results: We have analyzed gene network, which consists of 300 *E. coli* reactions. We have found 10 potential oscillatory cycles. In particular, one in the interest cycles was considered and its oscillatory behavior was proved. This cycle is the part of tryptophan pathway.

References

1. Kazantsev F. et al. MAMMOTh: a new database for curated MAthematical Models of bioMOlecular system. *J. Bioinform. Computational Biol*. 2018;16(1):1740010.

DOI 10.18699/SBB-2019-22

Organization and evolution of the polyphenol oxidase gene family in barley

Levanova N.M.[1, 2] *, Glagoleva A.Y.[1], Khlestkina E.K.[1, 2, 3], Shoeva O.Y.[1]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia*
** e-mail: n.levanova@g.nsu.ru*

Key words: *Hordeum vulgare*, *Ppo*, polyphenol oxidase, salt stress, pigmentation

Motivation and Aim: Tyrosinases are ubiquitous enzymes among all living organisms. In plants, tyrosinases called polyphenol oxidases (PPO) are involved in the response to abiotic and biotic stresses. In cereals, *Ppo* is predominantly represented by a gene family. Two *Ppo* genes were previously identified in barley: *Ppo1* and *Ppo2*, but additional copies were expected. The aim of the current study was to characterize the *Ppo* gene family in barley, to analyze the functional activity of the *Ppo* genes and their evolutionary relationships with orthologs of other cereals.
Methods and Algorithms: We used the NCBI database, IPK Barley BLAST Server and EnsemblPlants for sequences identification. For sequence alignment and the evolutionary tree construction, MULTALIN and MEGA 7.0 software were used. *Cis*-regulatory elements were analyzed using PLACE database. We developed specific primers for each copy using PrimerQuestTool. We used RT-PCR and qPCR to analyze expression of the *Ppo* genes.
Results: We identified two additional genes *Ppo3* and *Ppo4* based on the previously reported *Ppo1* and *Ppo2* and located them on chromosomes 3H and 4H, respectively. Copies contain the conservative tyrosinase domain, however, have a different intron-exon structure, as well as predicted promoter structure. Expression analysis in various organs (root, coleoptile, leaf, stem, developing spike, pericarp, and hulls) and also under salt stress showed that the *Ppo* genes possess different expression patterns. At the constructed evolutionary tree *Ppo1* and *Ppo2* clustered separately from copies of *Ppo3* and *Ppo4*. Presumably, the formation of the gene cluster comprising the *Ppo1* and *Ppo2* genes occurred as a result of segment duplication in the common ancestor of the Triticeae tribe. Orthologs of the *Ppo3* and *Ppo4* genes of barley were detected in other cereal species.
Conclusion: The *Ppo* gene family in barley contains at least four genes that maintain their functional activity during evolution.
Acknowledgements: Supported by the Russian Science Foundation (No. 16-14-00086).

DOI 10.18699/SBB-2019-23

Context complexity of sites containing single nucleotide polymorphisms in human genome

Luzin A.N.[1]*, Dergilev A.I.[1], Tabikhanova Z.E.[2], Safronova N.S.[1], Orlov Y.L.[1, 3]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Novosibirsk State Pedagogical University, Novosibirsk, Russia*
[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
** e-mail: a.luzin@g.nsu.ru*

Key words: bioinformatics, sequence complexity, medical informatics, polymorphism, SNP, databases

Motivation and Aim: Analysis of sequence text complexity is an approach for study genome structure based on sequencing data. We have analyzed sequence text complexity in flanking regions for the set of known single nucleotide polymorphisms (SNP) from the "1000 genomes" project. The aim was to find general and specific features for SNP sites associated with the diseases.
Methods and Algorithms: We used previously developed [1] and applied novel statistical computational methods to analyze genetic text based on its complexity. A complexity profiling in sliding window is applied to the sites, containing single nucleotide polymorphisms within a human genome. The complexity estimates were computed using previously developed program tool. This tool allows for both (i) complexity estimation of phased samples, and (ii) fast and effective defining the frequency spectrum of oligonucleotides with fixed lengths, and making a frequency comparison of oligonucleotides in different samples.
Results: A local decrease in text complexity level in SNP-containing sites is shown. Complexity profiles for SNP-containing sites shows that flanking monomer repeats define lower context complexity of sites containing SNPs within a human genome. An effect of local decrease in text complexity in SNP-containing sites is confirmed by analysis of polymorphisms in available model genomes (mice and rat).
Conclusion: Problem SNP sites analysis is of importance for personalized medicine and genomics studies. The changes in point mutation frequency were shown earlier for microsatellite containing sequences. Using extended data sets this work shows enrichment of polytracks and simple sequence repeats in local genome surroundings of SNP containing sites. We have found high frequent oligonucleotides within genome regions containing SNPs. Such oligonucleotides are related to nucleotide poly-tracks. The presence of poly-A tracks might be associated with an increased probability of double helix DNA breaks around mutable loci and following fixation of nucleotide changes.
Acknowledgements: Supported by the RFBR and ICG SB RAS budget project (0324-0019-0040).

References

1. Orlov Y.L., Te Boekhorst R., Abnizova I.I. Statistical measures of the structure of genomic sequences: entropy, complexity, and position information. *J. Bioinform. Comput. Biol*. 2006;4:523-36.

DOI 10.18699/SBB-2019-24

Inhibitors of tyrosyl-DNA phosphodiesterase 1 (Tdp1) leelamine and berberine derivatives as prototypes of antitumor drugs

Mamontova E.[1, 2]*, Kovaleva K.[2, 3], Shirokova E.[2, 3], Zakharenko A.[1], Yarovaya O.[3], Luzina O.[3], Zakharova O.[1], Salakhutdinov N.[2, 3], Lavrik O.[1, 2]

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

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

[3] *Institute of Organic Chemistry, SB RAS, Novosibirsk, Russia*

** e-mail: evgeniya.mm.94@gmail.com*

Key words: DNA repair, chemotherapy, tyrosyl-DNA phosphodiesterase 1, SCAN1, leelamine, berberine

Motivation and Aim: Tyrosyl-DNA phosphodiesterase 1 (Tdp1) is a promising target for cancer and neurodegenerative therapy. Tdp1 plays an important role in the removal of damage caused by topoisomerase 1 (Top1) inhibitors camptothecin and its derivatives topotecan and irinotecan. Last two compounds are the antitumor drugs widely used in clinical practice [1]. It makes this enzyme a promising target to enhance anticancer treatment in combination with DNA damaging therapies such as Top1 inhibitors. Moreover there is a natural mutant of Tdp1 which is responsible for the development of a neurodegenerative disease spinocerebellar ataxia with axonal neuropathy (SCAN1) [2]. The suppression of SCAN1 mutant activity can improve the SCAN1 patients' condition and prevent the progression of the disease. The natural compounds leelamine and berberine derivatives were chosen as potential Tdp1 and SCAN1 inhibitors based on their multiple biological activities.

Methods and Algorithms: To determine the inhibitory properties of compounds, fluorescent DNA biosensor was used [3]. IC_{50} values for each compound were calculated. To determine the own cytotoxicity of compounds and their cytotoxicity in combination with topotecan a standard MTT test was used.

Results: During the screening of 42 compounds 20 effective inhibitors of Tdp1 with IC_{50} values less than 1 μM were found. The screening of 16 compounds against SCAN1 showed 10 inhibitors of SCAN1. Most of compounds appeared to be nontoxic for lung cancer A549 cell line. Two leelamine derivatives enhanced the cytotoxicity of topotecan.

Conclusion: Both leelamine and berberine derivatives showed themselves as promising compounds for further development of anticancer drugs and drugs for treatment of the neurodegenerative disease SCAN1.

Acknowledgements: Supported by RFBR (grants 17-04-01071 and 18-33-00297) and by RSF (grant 19-13-00040).

References

1. Pommier Y. Topoisomerase I inhibitors: camptothecins and beyond. *Nature Reviews*. *Cancer*. 2006;6(10):789-802.
2. Takashima H. et al. Mutation of TDP1, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. *Nature Genetics*. 2002;32(2):267-272.
3. Zakharenko A. et al. Synthesis and biological evaluation of novel tyrosyl-DNA phosphodiesterase 1 inhibitors with a benzopentathiepine moiety. *Bioorganic Med. Chemistry*. 2015;23(9):2044-2052.

DOI 10.18699/SBB-2019-25

Cell-type deconvolution from DNA methylation in carotid atherosclerotic plaques using bioinformatics approach

Markov A.[1]*, Zarubin A.[1], Sharysh D.[2], Golubenko M.[1, 3], Babushkina N.[1], Valiakhmetov N.[1], Kazantsev A.[3], Barbarash O.[3], Puzyrev V.[1, 2], Nazarenko M.[1, 2, 3]

[1] *Research Institute of Medical Genetics of Tomsk NRMC, Tomsk, Russia*

[2] *Siberian State Medical University, Tomsk, Russia*

[3] *Research Institute for Complex Issues of Cardiovascular Diseases, Kemerovo, Russia*

** e-mail: anton.markov@medgenetics.ru*

Key words: atherosclerosis, DNA methylation, DNA microarrays, cell-type deconvolution

Motivation and Aim: A range of aberrantly methylated genomic loci was identified in arteries during atherosclerosis, but the top lists of loci from different studies demonstrate a little overlap, possibly, due to diverse cellular composition [1]. We assessed the bioinformatic framework "EpiDISH" [2] to account for cell type heterogeneity in carotid atherosclerotic plaques (CAP).

Methods and Algorithms: We used DNA methylation (DNAme) data from Gene Expression Omnibus to build the EpiDISH reference for cell types constituting arterial tissues. Experimental specimens of CAP and blood leukocytes (PBL) were collected from 43 patients (aged 63±7 years). Immunohistochemical analysis and genome-wide DNAme profiling (using Infinium MethylationEPIC BeadChip) were performed in 16 CAP. Methylation level of separate CpG-sites was validated in 34 patient-matched CAP and PBL samples by NGS-based targeted bisulfite sequencing.

Results: Unsupervised cluster analysis of CAP methylomes (based on 775836 CpG-sites) showed two main sample groups which could not be explained by clinical characteristics or type of atherosclerotic plaque. About 15 % of CpG-sites (including CpG-site cg06330621 within *ZCCHC14* gene) were shown to be differently methylated between these groups (the difference between average methylation levels $|\Delta\beta|>0.1$, FDR-adjusted p-value<0.05). EpiDISH allowed to get approximate cell-type content, and the proportion of immune cells explained well the sample clustering. We showed a proportion of CD68+ macrophages corresponded with EpiDISH "monocyte" fraction (estimation error varied from 3 to 11 %). Validation assay of DNAme within *ZCCHC14* locus showed distinct tissue-specific DNAme patterns in paired CAP and PBL samples.

Conclusion: Genome-wide DNAme in carotid atherosclerotic plaques is strongly associated with underlying immune cell infiltration. An anticipated cell-type content in the whole specimen of the artery can be obtained from DNAme data.

Acknowledgements: Supported by the RSF (17-75-10146).

References

1. Aavik E., Babu M., Ylä-Herttuala S. DNA methylation processes in atherosclerotic plaque. *Atherosclerosis*. 2019;281:168-179.
2. Teschendorff A.E., Breeze C.E., Zheng S.C., Beck S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in epigenome-wide association studies. *BMC Bioinformatics*. 2017;18(1):105.

DOI 10.18699/SBB-2019-26

Reconstruction and analysis of the network of interactions between genes that regulate human body weight

Matrosova E.A.[1]*, Ignatieva E.V.[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
** e-mail: e.a.matrosovaa@yandex.ru*

Key words: obesity, gene network, PPI, SNPs, transcription factor binding sites

Motivation and Aim: Obesity is a multifactorial chronic disease manifested in an excessive increase in adipose tissue mass and is a risk factor for many diseases. The prevalence of this pathology in the world is increasing, which makes this problem particularly relevant. In order to obtain additional information on the genetic and molecular basis of obesity, we built and analyzed a network of interactions between proteins encoded by genes involved in the regulation of body weight.
Methods and Algorithms: For the list of genes, which was formed earlier [1], information on protein-protein interactions (PPIs) from the GeneMANIA, STRING and BioGRID databases was searched and extracted. The reconstruction and analysis of networks were carried out using Cytoscape. For the reconstruction of networks specific for adipocytes and brain cells, expression information from publications was used [2] and [3]. Functional annotation of genes (GO analysis) was performed using DAVID tool. For finding potential transcription factor binding sites disturbed by GWAS SNPs collected in [1], UCSC Variant Annotation Integrator, dbSNP and PERFECTOS-APE were used.
Results: We obtained PPI networks involving proteins controlling body weight and/or feeding behavior (1) encoded by all genes from [1] and (2) (3) specific for adipocytes and brain cells. GO analysis of lists of genes involved in networks has shown that they are enriched with genes controlling transcriptional and hormonal regulation. 24 potential binding sites for transcription factors that are involved in the regulation of body weight [1], which may be affected by SNPs, were identified. We found that transcriptional regulators reliably more often corresponded to vertices with increased degree and high values of centrality and radiality, for transmembrane receptors, the opposite pattern was revealed. For proteins INO80E and DMXL2, protein-protein interactions have been identified that allow the construction of hypothetical mechanisms for their participation in the regulation of body weight.
Acknowledgements: Supported by the State Budgeted Project No. 0324-2019-0040.

References

1. Ignatieva E.V. et al. A compendium of human genes regulating feeding behavior and body weight, its functional characterization and identification of GWAS genes involved in brain-specific PPI network. *BMC Genet*. 2016;17(Suppl. 3):158.
2. Henry F.E. et al. Cell type-specific transcriptomics of hypothalamic energy-sensing neuron responses to weight-loss. *Elife*. 2015;4.
3. Wells A. et al. The anatomical distribution of genetic associations. *Nucleic Acids Res*. 2015;43(22): 10804-20.

DOI 10.18699/SBB-2019-27

Functional study of snoRNAs activity using transcriptomic analysis of CRISPR/Cas9-modified human cells

Matveeva A.M.[1, 2]*, Filippova J.A.[1], Zhuravlev E.S.[1], Semenov D.V.[1], Stepanov G.A.[1, 2]
[1] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: anastasiya.maatveeva@gmail.com*

Key words: small nucleolar RNAs, genome editing, CRISPR/Cas9, gene expression profiling

Motivation and Aim: Small nucleolar RNAs (snoRNAs) represent a special class of small regulatory RNAs, participating in post-transcriptional modification of ribosomal RNA in eukaryotes. Box C/D snoRNAs present one of the subclasses and with the help of conserved elements in their structure provide targeted 2'-O-methylation of rRNA nucleotides [1]. SnoRNAs possess multiple non-canonical functions as well, such as modulation of pre-mRNA alternative splicing and an ability to be processed into short-derived RNAs resembling miRNAs [2].
Methods and Algorithms: We obtained plasmid constructs expressing CRISPR/Cas9 components guiding double-strand breaks into GAS5 (Growth Arrest Specific 5) gene regions encoding target snoRNAs. 293FT-derived single cell clones were characterized for mutations and target gene expression. The 2'-O-methylation level of corresponding rRNA nucleotides was estimated with a specially designed RT-PCR analysis. RNA-Seq method allowed carrying out a differential gene expression profiling.
Results: RNA-Seq analysis revealed both upregulated and downregulated genes for cell lines with either U75 or U77 snoRNAs being modified using CRISPR/Cas9. These genes were then categorized into groups using Enrichr online tool [3]. One of the most highly presented groups were found to be controlled with transcription factors SUZ12 and EZH2, which are involved in the processes of the chromatin remodeling. Common upregulated genes groups for both cell lines were distinguished, the group of high interest being associated with the innate immune response processes. Interestingly, this group was more diverse for the cell line with mutant U75 snoRNA.
Conclusion: The results obtained suggest that immune response might be activated due to the inability of mutant snoRNAs to form stable canonical ribonucleoprotein complexes with nucleolar proteins. Induced snoRNA structure alterations could result in binding with novel protein partners and acquisition of new functions.
Acknowledgements: This work was supported by the RFBR grant 18-29-07073 and partially (in method development) by State Budget Program (0245-2019-0001).

References

1. Jorjani H., Kehr S., Jedlinski D.J., Gumienny R., Hertel J., Stadler P.F., Zavolan M., Gruber A.R. An updated human snoRNAome. *Nucleic Acids Res*. 2016;44:5068-5082.
2. Falaleeva M., Welden J.R., Duncan M.J., Stamm S. C/D-box snoRNAs form methylating and non-methylating ribonucleoprotein complexes: Old dogs show new tricks. *Bio Essays*. 2017.
3. Chen E.Y., Tan C.M., Kou Y., Duan Q., Wang Z., Meirelles G. V., Clark N.R., Ma'ayan A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013.

DOI 10.18699/SBB-2019-28

The genes determining synthesis of pigments in cotton

Mikhailova A.S.[1, 2]*, Strygina K.V.[2, 3], Khlestkina E.K.[2, 3]
[1] *St. Petersburg State University, St. Petersburg, Russia*
[2] *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia*
[3] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
** e-mail: a.mikhailova@vir.nw.ru*

Key words: anthocyanins, bHLH, flavonoids, MBW complex, Myb, proanthocyanidins

Motivation and Aim: Flavonoid compounds proanthocyanidins (PAs) and anthocyanins in cotton fibres and stems provide the brown and red colour, respectively [1]. Naturally coloured cotton fibres are hypoallergenic and do not lead to environmental pollution in the textile production. Stems contained flavonoid compounds are insensitive to various environmental stress factors as ultraviolet radiation, anomaly temperature changes and fungal infections [2]. The anthocyanins and PAs biosynthesis is under control of R2R3-Myb, bHLH-Myc and WD40 transcription factors (TF) forming MBW regulatory complex [3]. The aim of the present study was the identification and characterization of homeologous and paralogous R2R3-Myb and bHLH-Myc genes copies in genomes of diploid and allotetraploid cotton species (*G. arboretum* A_2, *G. raimondii* D_5, *G. hirsutum* AD_1 and *G. barbadense* AD_2).
Methods and Algorithms: The search of homologous sequences was made in databases for not annotated cotton sequences using BLAST. The cluster analysis using the MEGA software was based on NJ algorithm. The number of non-synonymous substitutions per non-synonymous sites (*Ka*), the number of synonymous substitutions per synonymous sites (*Ks*) and the *Ka*/*Ks* ratio for *Myb* and *bHLH* genes were calculated. The obtained value was used for the calculation of divergence time of duplicated copies.
Results: The information about known *R2R3-Myb* (orthologs of *AtTT2*) and *bHLH-Myc* (orthologs of *AtTT8*) genes was used to identify duplicated genes by BLAST search in genome of cotton. Due to this analysis, we identified three groups of *Myb* and one group of *Myc* genes. The results of investigation of phylogenetic relationship and the analysis of structural organization of duplicated regulatory genes are represented in current report. The calculation of the divergence time revealed that the duplication of *Myb* genes occurs on early stages of cotton genome development. Genotype-specific features are revealed for some duplicated genes.
Conclusion: The duplications of the *R2R3-Myb* genes occurred several times in the course of cotton evolution. All identified sequences have conserved R2R3 and bHLH domains. These findings can be used further for the development of diagnostic PCR- and CAPS-markers in marker-assisted selection of cottons with naturally coloured fibres.

References

1. Feng H., Li Y., Wang,S., Zhang L., Liu Y., Xue F. Molecular analysis of proanthocyanidins related to pigmentation in brown cotton fibre (*Gossypium hirsutum* L.). *J. Experimental Botany*. 2014;65(20):5759-5769.
2. Khlestkina E.K. Genes determining the coloration of different organs in wheat. *Russian J. Genetics: Applied Research*. 2013;3(1):54-65.
3. Gong W., Du X., Jia Y., Pan Z. Color cotton and its utilization in China. In cotton fiber: physics, chemistry and biology. – Cham: Springer International Publishing. 2018;117-132.

DOI 10.18699/SBB-2019-29

Developing of web-service WebMCOT for finding cooperative site-binding TF DNA-motifs

Mukhin A.M.[1, 2]*, Levitsky V.G.[1, 2], Lashin S.A.[1, 2]
[1] *Novosibirsk State University, Novosibirsk*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
** e-mail: mukhin@bionet.nsc.ru*

Key words: transcription factors, transcription factor binding sites, composite element

Motivation and Aim: There are transcription and regulatory regions in eukaryotes genes. Transcription factors (TF) are proteins which regulate DNA transcription, and transcription factor binding sites (TFBS) are DNA sites for TF. TF action is cooperative. DNA sites are modelled by DNA motifs – position weight matrices where columns designate nucleotides, rows designate positions in a motif. Combinatorial binding of multiple TFs to a regulatory region provides fine-tuning of the gene expression; two TFBS in a composite element (CE) [1] are needed for starting transcription. Motifs can be overlapped or be separated with a spacer. The aim of this work is to develop a web-service WebMCOT (Web Motifs Co-Occurrence Tool [2]) for prediction of CEs in ChIP-seq data.
Materials and Methods: We used the following programming languages for each module: JavaScript and HTML for making web site, Python for web-service building, C++ for computational core. Also we used additional libraries: Vue.JS for making single page application, Flask for building REST API service, Celery for distribute task to another servers and cache-database Redis for optimal working with Celery. For checking C++ program and optimization memory, we used Jemalloc allocator with Jeprof heap profiling.
Results: We have developed a web-service WebMCot for prediction of CEs in ChIP-seq data. There are 3 modules of this web-service: SPA, REST API web-service and statistical module. The SPA and the REST API web-service sending messages via HTTP-packets. The statistical module has been installed into the Institute of Cytology and Genetics HPC cluster. The web-service and the statistical module are connected via Redis database by the Celery library.
Conclusions: The GUI-based web-service has been implemented that can predict novel CEs.
Acknowledgements: The work was supported by Russian Foundation for Basic Research (project No. 18-29-13040).

References

1. Kel O.V., Romaschenko A.G., Kel A.E., Wingender E., Kolchanov N.A. A compilation of composite regulatory elements affecting gene transcription in vertebrates. *Nucleic Acids Res*. 1995;23:4097-4103.
2. Levitsky V.G., Oshchepkov D.Y., Zemlyanskaya E.V, Mironova V.V., Ignatieva E.V., Podkolodnaya O.A., Merkulova T.I. The overlapped motifs co-occurrence in ChIP-seq data. *Proceedings of the 11th International Conference on Bioinformatics of Genome Regulation and Structure\Systems Biology* (*BGRS\SB-2018*). 2018;58.

DOI 10.18699/SBB-2019-30

Application of Cre/LoxP system for induction of multiple chromosomal aberrations in the human genome

Mungalov R.[1, 2]*, Khabarova A.[1], Fishman V.[1]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: mungalov.roman@yandex.ru*

Key words: Cre/LoxP system, genome architecture, gene expression

Motivation and Aim: In this research we have a goal to study an interplay between disturbances of three-dimensional human genome organization [1], induced by chromosomal aberrations, and gene expression and epigenetic modifications. We suggest an alternative way of the induced mutagenesis, namely Cre/LoxP system. It will allow us to create a big collection of chromosomal aberrations (up to 10 events per cell), in spite of unique events using CRISPR/Cas9 [2].The analysis of just a few number of subclones, each of them will have about 10 aberrations, will allow us to get enough information about gene expression changes in different parts of the human genome.
Methods and Algorithms: We use Cre/LoxP system [3] to induce chromosomal aberrations in near-haploid human cell line HAP-1. We apply lentiviruses as vectors for transporting LoxP-sites in random parts of human genome and then perform exogenic expression of Cre-recombinase, inducing different recombinations between LoxP-sites. Number of LoxP-site integrations and recombinations we estimate using qPCR. To localize LoxP-site's integrations in the genome we are planning to create inverse-PCR library of subclones, carrying different numbers of LoxP-sites. Then this library will be sequenced. Several subclones will be further analyzed by capture HiC and bioinformatics modeling of the genome 3D-landscape.
Results: We have obtained about 50 subclones, carrying various number of LoxP-site integrations, and sequenced ones with the largest number of integrations to localize LoxP-insertions. To improve a system of transporting LoxP-sites to the genome we have constructed new vectors, based on Sleeping Beauty transposons.
Conclusion: Using Cre/LoxP system we have achieved an integration of about 25 LoxP-sites (maximum now) in the human near-haploid cell line HAP-1. In addition, we have sequenced subclones with the maximum number of LoxP-site integrations. Further recombination and modeling of 3D-landscape will show the interplay between the genome architecture and gene expression.

References

1. Rowley M.J., Corces V.G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet*. 2018;19:1.
2. Lupiáñez D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 2015;161(5):1012-1025.
3. Sauer B. et al. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc. Natl. Acad. Sci*. 1988;85(14):5166-5170.

DOI 10.18699/SBB-2019-31

T7-like promoter function: DNA physics implication

Orlov M.A.*, Kondratiev M.S., Sorokin A.A.
Institute of Cell Biophysics RAS, Pushchino, Russia
** e-mail: orlovmikhailanat@gmail.com*

Key words: T7, DNA physics, RNA-polymerase, promoter

Motivation and Aim: RNA polymerase – promoter interactions has been the subject to extensive studies for decades. However, there are still certain relevant mechanistic details to be established. Recent advances in the field was enabled to more attention paid to DNA physicochemical properties in addition nucleotide sequence analysis. This biomolecules properties are primarily responsible for their interaction, particularly at initial stages of RNA-polymerase--promoter interaction. T7 bacteriophage and related genomes can serve as suitable study objects in the field due to their small genomes being transcribed by two different RNA-polymerase including host-specific and native one encoded by phage DNA. The latter enzyme is characterized by small size, high processivity, and capability of discriminating promoters that have very similar primary structure but differ in biochemical and physical properties as well as activation time during lifecycle [1]. The DNA physicochemical properties of the promoter are likely to affect their differential recognition as well as implication in replication, etc. [2].
Methods and Algorithms: The extent to which the T7-DNA promoter are susceptible to melt under superhelicity stress was assessed using Stress-Induced Duplex Destabilization (SIDD) model [3]. Molecular dynamics (MD) simulations were used to determine contribution of various promoter DNA and T7-RNA-polymerase to root mean squared deviation of the molecules in the structurally resolved complex as well as in free form.
Results: Native promoters representing class III are substantially more destabilized as compared to class II ones according to their SIDD profiles. Promoters serving as secondary replication origin demonstrates higher SIDD maxima. MD experiments have yielded RMSD dynamics for a highly destabilized strong T7-promoter (class III consensus sequence) and weaker class II promoter sequence.
Conclusions: Highly similar or identical in their nucleotide sequences promoters of T7 bacteriophage were shown to differ in their physicochemical properties. This both directly or by modulating promoter strength may affect their specific recognition by native RNA-polymerase and genes expression regulation.

References

1. Kamzolova S.G., Beskaravainy R.M., Osypov A.A. et al. Electrostatic map of T7 DNA: Comparative analysis of functional and electrostatic properties of T7 RNA polymerase-specific promoters. *Journal Biomolecular Structure Dynamics*. 2014;32:1184-1192.
2. Orlov M.A., Ryasik A.A., Sorokin A.A. Destabilization of the DNA Duplex of Actively Replicating Promoters of T7-Like Bacteriophages. *Mol. Biol*. 2018;52(5):686-692.
3. Wang H., Benham C.J. Promoter prediction and annotation of microbial genomes based on DNA sequence and structural responses to superhelical stress. *BMC Bioinformatics*. 2006;7:248.

DOI 10.18699/SBB-2019-32

Characteristic of the spinal muscular atrophy cell model

Ovechkina V.[1, 2]*, Valetdinova K.[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Chemical Biology and Fundamental Medicine, SB RAS, Novosibirsk, Russia*
** e-mail: Vs_ovechkina@mail.ru*

Key words: spinal muscular atrophy, induced pluripotent stem cells, cell model

Motivation and Aim: After the induced pluripotent stem cells were obtained by the Yamanako group, experiments on the development of cellular models of disease have become widespread in medical biology. Here we performed a comparative description of a previously obtained cellular model of spinal muscular atrophy (SMA) at different stages of differentiation. We compared the expression of well-known and newly founded by genome-wide linkage analysis genes, involved in molecular pathogenesis of SMA.
Methods and Algorithms: We utilized induced pluripotent stem cells, obtained from conditionally healthy patient and patients with SMA I and II types [1]. Using the open Internet resource BLAST we selected primer pairs to estimate by Real-time PCR the number of *SMN1, SMN2, PLS3, SLC23A2, NCALD, RPL6, CDKA2P1* genes [2–4], which products are involved in the development of SMA. Statistical processing of the results were carried out using the non-parametric Kruskal-Wallis criterion for independent groups using the Statistica 64. Differences were considered significant when the probability parameter value was $p<0.05$.
Results: We shown a decrease in the expression level of the full-sized *SMN* transcript at all stages of neural differentiation in cell models with the SMA phenotype of types I and II in comparison with control. An increase in the level of the incomplete *SMN* transcript was also detected. We shown a change in the transcription of modifier genes: in case of SMA type I, the number of *RPL6* transcripts increases in mature motor neurons, and the expression level of *PLS3, NCLD, CDK2AP1* decreases.
Conclusion: The earlier obtained cell model of SMA complies well-known literary characteristics of mature motor neurons with SMA phenotype and can be used for further molecular applications. Newly founded by genome-wide linkage analysis genes (*RPL6* and *CDK2AP1*) can be also involved in pathogenesis of SMA.
Acknowledgements: Supported by RSF (17-75-10041).

References

1. Valetdinova K.R. Spinal muscular atrophy model system based on induced pluripotent stem cells: Abstract dis. ... Ph.D. in Biology: 03.02.07. Novosibirsk, 2016 (in Russian).
2. Alrafiah A., Karyka E., Coldicott I., Iremonger K., Lewis K.E., Ning K., Azzouz M. Plastin 3 promotes motor neuron axonal growth and extends survival in a mouse model of spinal muscular atrophy. *Mol. Ther. Methods Clin. Dev*. 2018;31(9):81-89.
3. Zheleznyakova G.Y., Nilsson E.K., Kiselev A.V., Maretina M.A., Tishchenko L.I., Fredriksson R., … Schiöth H.B. Methylation levels of SLC23A2 and NCOR2 genes correlate with spinal muscular atrophy severity. *PLoS ONE*. 2015;10(3):1-14.
4. Riessland M., Kaczmarek A., Schneider S., Swoboda K.J., Löhr H., Bradler C., … Wirth B. Neurocalcin delta suppression protects against spinal muscular atrophy in humans and across species by restoring impaired endocytosis. *American J. Human Gen*. 2017;100(2):297-315.

DOI 10.18699/SBB-2019-33

Dynamical modeling of the floral transition in legumes

Pavlinova P.A.[1], Gursky V.V.[1, 2], Kozlov K.N.[1], Samsonova M.G.[1]*
[1] *Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*
[2] *Ioffe Institute, St. Petersburg, Russia*
** e-mail: m.samsonova@spbstu.ru*

Key words: *flowering gene network, FLOWERING LOCUS T, chickpea, pea*

Flowering time is an important stage of plant development, representing a transition from the vegetative to reproductive phases. Understanding processes controlling transition to flowering is significant for plant selection. The protein FLOWERING LOCUS T (FT) provides the key flowering signal mediating floral transition in the model plant *Arabidopsis thaliana*. This transcription factor is expressed in leaves and transfers to apical meristem of the plant, where it activates expression of the floral meristem identity genes *LEAFY* and *APETALA1*. In legumes, multiple *FT* duplication occurred early in evolution, so that all legumes most probably have *FT* orthologs from the following three clades: *FTa*, *FTb*, and *FTc*. This raises a question about the organization of the mobile flowering signals in legumes. We developed a dynamical model of the core gene network controlling transition to flowering and applied it to the *previously published expression data* for two closely related legume species, pea (*Pisum sativum*) and chickpea (*Cicer arietinum, the early flowering ICCV 96029 cultivar*). Both species have five *FT*-like genes, but the data-driven modeling results predict drastically different flowering signal organizations in pea and in the studied chickpea cultivar. For *ICCV 96029, the model shows evidence that* the inputs from different *FT*-like genes combine additively in the activation of flowering [1]. The pea orthologs of *FT* have essentially different expression profiles, which resulted in a hypothesis about more complicated functional roles of *FT*-like genes in the flowering gene network in pea [2, 3]. Our model confirms this hypothesis at the quantitative level. The comparative analysis of the regulatory architectures driving floral transition in pea and chickpea provides a first step in elucidating evolution of this regulatory module in legumes.

Acknowledgements: The study was supported by the Russian Foundation for Basic Research, grant 18-29-13033.

References

1. Gursky V.V., Kozlov K.N., Nuzhdin S.V., Samsonova M.G. Dynamical modeling of the core gene network controlling flowering suggests cumulative activation from the FLOWERING LOCUS T gene homologs in chickpea. *Front. Genet*. 2018;9:547.
2. Hecht V., Laurie R.E., Vander Schoor J.K. et al. The Pea GIGAS Gene Is a FLOWERING LOCUS T homolog necessary for graft-transmissible specification of flowering but not for responsiveness to photoperiod. *Plant Cell*. 2011;23:147.
3. Sussmilch F.C., Berbel A., Hecht V., Vander Schoor J.K., Ferrándiz C., Madueño F., Weller J.L. Pea VEGETATIVE2 is an FD homolog that is essential for flowering and compound inflorescence development. *Plant Cell*. 2015 Apr;27(4):1046-1060.

DOI 10.18699/SBB-2019-34

No association between SCN9A gene rs6746030 polymorphic variant and psychoactive drug addiction

Poltavskaya E.*, Savochkina D.
Mental Health Research Institute of Tomsk NRMC, Tomsk, Russia
** e-mail: egboyarko@mail.ru*

Key words: SCN9A, gene polymorphism, addiction, psychoactive substance

Motivation and Aim: SCN9A gene encodes voltage-gated sodium channel NaV1.7, which expressed in the nervous system. Shown that NaV1.7 is present within neurons in the hypothalamic supraoptic nucleus, specifically within vasopressin- and oxytocin-producing magnocellular neurosecretory neurons [1]. Alterations to the development of the oxytocin system plays a pivotal role in the regulation of social behaviours and emotion [2]. Dysregulation of this system also contributes to increased susceptibility to develop drug addiction. The aim of this study was to study the possible association of rs6746030 polymorphism with psychoactive drug addiction in Russian population of Siberian region.

Methods and Algorithms: The groups for study were formed on the basis of psychoneurologic dispensaries (patients) and higher schools (healthy persons) in two cities of West Siberia (Russia). We examined 213 persons – 106 participants with psychoactive drug (PAD) addiction (F1x.2 according to ICD-10), and 107 healthy person (control group). Genotyping of rs6746030 SCN9A has been carried out using ABI StepOne Plus with TaqMan1 Validated SNP Genotyping Assay (Applied Biosystems). The Hardy-Weinberg equilibrium of genotypic frequencies was tested by the chi-square test. Statistical analyses were performed using SPSS software, release 21, for Windows; $P<0.05$ was considered as significant.

Results: The genotype distribution of rs6746030 SCN9A polymorphism were in agreement with Hardy-Weinberg equilibrium in patient group and in control group. Comparison of genotypes frequencies of SCN9A gene rs6746030 polymorphism in groups of patients with PAD addiction and healthy persons has not revealed statistically significant difference. The genotypes distribution has been shown as follows: GG (69.8 and 73.8 %), AG (27.4 and 22.4 %), AA (2.8 and 3.7 %) in patients with PAD addiction and healthy persons, respectively.

Conclusion: Thus, there is no association between SCN9A gene rs6746030 polymorphic variant and PAD addiction in Russian population of Siberian region.

Acknowledgements: This work has been completed under support of RFBR research project No. 17-36-01113.

References

1. Black J.A. et al. NaV1.7: stress-induced changes in immunoreactivity within magnocellular neurosecretory neurons of the supraoptic nucleus. *Mol. Pain*. 2013;8(9):39.
2. Baracz S.J. et al. The impact of early life stress on the central oxytocin system and susceptibility for drug addiction: applicability of oxytocin as a pharmacotherapy. *Neurosci Biobehav Rev*. 2018; 30:pii:S0149-7634(18)30276-8.

DOI 10.18699/SBB-2019-35

Induced expression of precursor, mature and unprocessed forms of BDNF in the rat neonatal cortex

Shaburova E.[1, 2]*, Lanshakov D.[1]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: sliza.nsk@gmail.com*

Key words: BDNF, dexamethasone, natural antisense BDNF, viral transduction

Motivation and Aim: We have previously found that even a single injection of dexamethasone on the third day of life led to changes in behavior and enhanced expression of *Bdnf* mRNA in the rat cortex within 5 days after the injection [1]. Changes in the expression of processed and non-processed forms of BDNF that have opposite effects on the formation of the neonatal cortex can be a possible reason for these hormone's effects.
Methods and Algorithms: To test this hypothesis, we created lentiviral vectors which contain different cDNA: normal proBDNF, mBDNF without propeptide sequence, mutant mutBDNF, which is not processed due to mutation of the propeptide cleavage site, and control without insertion. These TET-ON-controlled vectors were injected into the cortex of rat pups on the first postnatal day. Expression of the vectors was activated by doxycycline treatment from 3rd to 8th days of life. After the 8th day, the material was taken to test DOX induction of expression with RT-PCR analysis. Remaining rats were raised without doxycycline up to one month. Statistical differences were determined by one-way or two-way ANOVA followed by Fisher's least significant difference post hoc analysis.
Results: On the 8th day of life, in the animals with the injection of mBDNF or mutBDNF vector, level of *Bdnf* mRNA expression in the cortex was significantly doubled, compared to control. At the same time, there was no increase in this level after the injection of proBDNF vector.
Conclusion: To explain the obtained results, a new hypothesis was proposed. There is a mechanism of *Bdnf* mRNA suppression by antisense RNA, which have a region of complementarity with the sequence encoding the propeptide. This mechanism is able to function only with the presence of proBDNF cDNA. Later, similar antisense RNA was found in humans and mice [2]. Our results suggest the existence of such a mechanism of *Bdnf* expression regulation in rats. Identification of the sequence of this antisense RNA in rats elucidates the functioning of this mechanism in this species and its participation in the development of the psychophysiological properties of the individual in early ontogenesis.
Acknowledgements: Supported by the RFBR (16-34-60103).

References

1. Bulygina V. et al. Dexamethasone upregulates levels of proapoptotic proteins proBDNF and cleaved caspase-3 in the frontal cortex and the hippocampus of neonatal rats. *European Neuropsychopharmacol*. 2017;27(4):S609-S610.
2. Modarresi F. et al. Inhibition of natural antisense transcripts *in vivo* results in gene-specific transcriptional upregulation. *Nature Biotechnol*. 2012;30(5):453.

DOI 10.18699/SBB-2019-36

Genetic aspects of seed longevity in barley

Shvachko N.A.[1]*, Gerasimova T.V.[1], Khlestkina E.K.[1, 2]
[1] *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
** e-mail: n.shvachko@vir.nw.ru*

Key words: seed, longevity, germination, aging, QTL, gene mapping, *Hordeum*

Motivation and Aim: Seed longevity is a famous factor for conservation of genetic resources and biodiversity. The maintenance of viability during storage is an important trait of agriculture crops including barley (*Hordeum vulgare* L.). This feature is impaired by reactive oxygen species (ROS) during seed storage and germination. Longevity of the germplasm in seed banks remains for a limited time in storage. The aim of this analysis was the detection of new candidate genes associated with seed longevity in barley according to Scopus database.
Methods and Algorithms: The development of DNA-markers was made on the basis of known allelic differences of the genes putatively related with seed longevity. The BLAST-search and annotation of nucleotide sequences of the target genes was carried out in IPK BLAST Server according to known genes of *Zea mays*, *Oryza sativa* and *Arabidopsis thaliana*. Seed viability tests were performed on >90 barley accessions maintained in VIR Genbank since 1977 using method described [1].
Results: Genetic variability among barley accessions was revealed for viability level of the seeds maintained since 1977 in Genbank cold storage. Based on the known data about relation of seed longevity with major QTLs on barley chromosomes 2H, 5H and 7H [2], coincided with location of the genes *Zeo1* (compact spikes with long awns), *Ale* (barley vacuolar thiol protease) and *nud* (naked caryopsis) respectively, we estimated correlation of seed longevity with allelic state of the genes *Zeo1*, *Ale* and *nud*. In addition, we performed *in silico* analysis of the genes families for glutathione synthase and glutathione reductase, since glutathione disulfide accumulation is associated with oxidative response mechanism, oxidative stress and programmed cell death seeds [3].
Conclusion: Correlation of some allelic variants in the candidate genes with barley seed longevity is established. Accessions with high seed viability not explained by known genes will be used for genetic dissection aimed on finding novel genes underlying longevity.

References

1. Khoroshaylov N.G. (Ed.). Metodicheskie ukazaniya po dliteľnomu khraneniyu semyan [Guidelines for long-term seed storage]. Leningrad, 1981 (in Russian).
2. Nagel M., Vogel H., Landjeva S., Buck-Sorlin G., Lohwasser U., Scholz U., Börner A. Seed conservation in ex situ genebanks – genetic studies on longevity in barley. *Euphytica*. 2009;170(1-2):5-14.
3. Rouhier N., Lemaire S.D., Jacquot J. The role of glutathione in photosynthetic organisms: emerging functions for glutaredoxins and glutathionylation. *Ann. Review Plant Biol*. 2008;59(1):143-166.

DOI 10.18699/SBB-2019-37

Creation and analysis of an agent-based computer model of the AIDS epidemic using an algorithm for explicit calculation of the HIV replicability

Smirnova A.A.[1, 2]*, Ponomarenko M.P.[1], Lashin S.A.[1, 2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: asmirnova@bionet.nsc.ru*

Key words: agent-based model, prediction, simulation, TBP/TATA, HIV, AIDS

Motivation and Aim: The epidemic of HIV spread is a serious problem for the whole world. A large number of models describe HIV at different levels of the biological organization, from population-epidemiological level to molecular-genetic one. However, we have not met models that combine genetic and population levels together. The aim of the study was to develop an agent-based computer model of the HIV epidemic, which would take into account both the genetic variability of the virus and its corresponding variability in replication, as well as socio-epidemic aspects.
Methods and Algorithms: We used the method of computer prediction of affinity of TBP/TATA complex [1], which made it possible to assess the replication and virulence of different genetic variants of HIV. Python programming language was used to process the data. The agent-based model was implemented in the C++ programming language.
Results: The analysis of the diversity of virulence and replicativeness of various genetic variants of HIV-1 was carried out using statistical methods for predicting the affinity of the TBP / TATA complex. Clustering of 2500 HIV strains into 3 visible groups was obtained. An agent-based model for the spread of HIV in the human population has been developed.
Acknowledgements: The study was supported by the Budget Project No. 0259-2019-0008.

References

1. Suslov V.V., Ponomarenko P.M., Efimov V.M., Savinkova L.K., Ponomarenko M.P., Kolchanov N.A. SNPs in the HIV-1 TATA box and the AIDS pandemic. *Journal Bioinformatics Computational Biol*. 2010;8(3):607-625.

DOI 10.18699/SBB-2019-38

Transcriptional control of flavonoid biosynthesis in polyploid plant species

Strygina K.V.[1, 2]*, Mikhailova A.S.[1, 3], Khlestkina E.K.[1, 2]
[1] *N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia*
[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[3] *St. Petersburg State University, St. Petersburg, Russia*
** e-mail: k.strygina@vir.nw.ru*

Key words: flavonoids, MBW complex, cotton, wheat

Motivation and Aim: Gene duplication is the main evolution mechanism leading to the emergence of new gene functions and new species. Genomes of polyploid plants have an increased number of genes copies. Thereby, the studying of polyploid genomes is of interest to determine the functional and evolutionary features of the duplicated genes. Here, we report results of study on paralogous and homeologous copies of regulatory *Myb*, *bHLH* and *WD40* genes involved in flavonoid biosynthesis in monocot and dicot plant species – allotetraploid cotton *Gossypium hirsutum* L. (AD_1), allohexaploid bread wheat *Triticum aestivum* L. (BAD) and its relatives.
Methods and Algorithms: Homologous sequences of *Myb*, *bHLH* and *WD40* genes were found in databases for not annotated wheat (Triticeae tribe) and cotton (*Gossypium* genus) sequences using BLAST search. The *in silico* analysis of identified sequences using Multalin, FGENESH+, New PLACE and InterPro software was carried out. The construction of the phylogenetic trees, the calculation of *Ka*/*Ks* ratio and calculation of divergence time was made with MEGA software with 1000 bootstrap replicates.
Results: Using a wide range of bioinformatics tools, in Triticeae tribe we revealed three homologous clusters of *Myb* genes, two clusters of *bHLH* genes and two clusters of *WD40* genes based on known sequences of *TaC1* (*Myb*), *TaMyc1* (*bHLH*) and *ZmPAC1* (*WD40*) genes. On the other hand, we identified tree paralogous clusters of Myb-coding genes, one group of *bHLH* genes and two *WD40* paralogous groups in *Gossypium* genus – orthologs of *GhTT2*, *GhTT8* and *AtTTG1*, respectively. The results of investigation of phylogenetic relationship, divergence time and the analysis of structural organization of duplicated genes are represented. It was shown that members of *WD40* gene family are mostly supported by selection than *Myb* and *bHLH*. Additionally, the transcriptional activity of *Myb*, *bHLH* and *WD40* genes in different parts of wheat plant was investigated.
Conclusion: The results obtained in the current research revealed some genotype-specific features of phylogenetic relationship and structure organization of duplicated regulatory *Myb*, *bHLH* and *WD40* genes. In the future, the discovered genes could be used for the development of diagnostic DNA-markers in marker-assisted selection of cotton and wheat with naturally coloured tissues.

DOI 10.18699/SBB-2019-39

Revision of phylogenic relationships between several Acrididae subfamilies

Sukhikh I.[1]*, Ustyantsev K.[1], Bugrov A.[2, 3], Sergeev M.[2, 3], Vavilova V.[1], Blinov A.[1]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Systematics and Ecology of Animals, SB RAS, Novosibirsk, Russia*
** e-mail: igor3419@gmail.com*

Key words: Orthoptera, Acrididae, molecular phylogeny, insects

Motivation and Aim: Acrididae, with more than 6,500 species, is the largest, cosmopolitan family of short-horned orthopteran insects (Orthoptera, Caelifera). Due to their importance as agricultural pests, Acrididae had long attracted the attention of many taxonomists. However, there are many disagreements about position and status of several subfamilies of Acrididae family. Analysing sequences of several phylogenetic markers from the NCBI GenBank database, with the addition of experimentally obtained data, we try to determine the phylogenetic relation and taxonomic rank of the problematic groups.
Methods and Algorithms: Sequences of complete mitochondrial genomes, COI, COII, CytB and NADH mitochondrial genes as well as ITS2 ribosomal region from NCBI database were used in order to construct the phylogenetic trees of Acrididae family. In addition, we experimentally obtained sequences of COI, COII, CytB genes and ITS2 region for more than 40 species in order to compliment the database information. In order to assign the right substitution model to each coding gene, their alighments were parsed by PartitionFinder 2 v2.1.1. All phylogenetic trees were build using Maximum likelihood method realized in IQ-Tree online program v 1.6.10. In addition, we performed the Bayesian analysis using MrBayes program v 3.2.7.
Results: In order to establish phylogenetic relationships between Acrididae subfamilies, we compared complete coding mitochondrial and nuclear ITS2 DNA sequences for more than hundred Acrididae grasshoppers. In addition, we analysed four concatenated mitochondrial genes, COI, COII, CytB and NADH. Preliminary results show that all studied acridids fall into three main phylogenetic groups that include 13 subfamilies: (I) Oxyinae, Hemiacridinae, and Spathosterninae; (II) Melanoplinae; (III) Acridinae, Calliptaminae, Catantopinae, Cyrtacanthacridinae, Eyprepocnemidinae, Euryphyminae, Gomphocerinae, Oedipodinae, and Pezotettiginae. Additional comparison of recent studies to our data reveals a fourth group embracing five more subfamilies: (IV) Marelliinae, Pauliniinae, Leptysminae, Rhytidochrotinae, and Ommatolampidinae. Subfamilies Acridinae, Gomphocerinae and Oedipodinae appears to be polyphyletic and form several independent branches inside one main group that join all three subfamilies together.
Conclusion: Our analysis show that the commonly recognized subfamilies – Acridinae, Gomphocerinae and Oedipodinae appears to be polyphyletic, forming at least 2 new groups of species that may have a subfamily status.
Acknowledgements: The project was supported by the Russian Foundation for Basic Research (No. 17-04-01615).

DOI 10.18699/SBB-2019-40

Gene ontology enrichment and network analysis for differently expressed genes related to aggressive behavior

Tiys E.[1], Chadaeva I.[1, 2]*, Oshchepkov D.[1], Shichevich S.[1], Kozhemyakina R.[1]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: ichadaeva@bionet.nsc.ru*

Key words: differently expressed genes, behaviour, tame rats, aggressive rats, ANDVisio, FunGeneNet

Motivation and Aim: The purpose of this work was to analyze the co-expression, possible physical interactions, metabolic pathway, and co-localization of genes significantly different in expression levels between animals with contrasting behavioral phenotypes, gray rats, selected for tame and aggressive behavior for over 40 years (since 1972, more than 90 generations) [1].
Methods and Algorithms: Using bioinformatic analysis of RNA sequencing data, lists of differentially expressed genes for several brain regions of aggressive and tame rats (hippocampus, hypothalamus, mesencephalic tegmentum and periaqueductal gray) were compiled. Using the gene network research programs ANDSystem [2] and FunGeneNet [3], we analyzed 78 genes, the expression level of which was statistically different.
Results: The search for overrepresented processes using PANTHER and the analysis of networks for groups of these genes using FunGeneNet did not reveal a significant excess of the number of intergenic links in comparison with random networks.
Conclusion: The results of our analysis can be interpreted as an evidence of a known phenomenon, the polygenic determination of behavioral phenotype. In our case, these phenotypes are “aggressiveness towards a human” and “non-aggressiveness towards a human” in the rats of the studied lines. The resulting absence of direct connections between genes with the differential expression in the tame and aggressive rats confirms our assumption about a nonlinear interaction of these genes and their complex coordination in the manifestation of hereditarily fixed tame or aggressive behavior. Further work is planned to identify links of the differentially expressed genes.
Acknowledgements: The research is supported by RFBR (grant 18-34-00496).

References

1. Belyaev D.K., Borodin P.M. The influence of stress on variation and its role in evolution. *Biol. Zent*. 1982;100:705-714.
2. Ivanisenko V.A. et al. ANDSystem: an associative network discovery system for automated literature mining in the field of biology. *BMC Syst Biol*. 2015;9(Suppl. 2):S2.
3. Tiys E.S. et al. FunGeneNet: a web tool to estimate enrichment of functional interactions in experimental gene sets. *BMC Genomics*. 2018;19:76.

DOI 10.18699/SBB-2019-41

How heat stress drives the expression of LTR retrotransposons in the flatworm model organism *Macrostomum lignano*

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

Key words: heat shock, LTR retrotransposons, transgenesis

Motivation and Aim: There is a growing body of evidence suggesting that some transposable elements are activated under stress conditions. However, the exact molecular mechanism underlying the activation is rarely known. Recently, we have found that two groups of *Ty3/Gypsy* LTR retrotransposons (LTR-RTs) are highly expressed under heat shock (37 ºC) in the genome of a free-living flatworm *Macrostomum lignano*, a model species to study stem cell biology, regeneration and ageing [1]. Here, we present new evidence for the activation mechanism of these retrotransposons, confirming our previous hypothesis with the new experimental data.
Methods and Algorithms: We utilized the transgenesis technology [2] to investigate whether the long terminal repeat (LTR) sequence of the LTR-RTs alone can induce expression of the reporter fluorescent protein in *M. lignano* after heat shock, and used the promoter of the heat shock protein gene (*Hsp*) 20 as a positive control [3].
Results: Three new *M. lignano* transgenic lines were produced. The heat shock treatment resulted in the expression of the fluorescent reporter controlled both by the LTRs and the *Hsp20* promoters of the transgenic worms. The LTR-driven expression was mostly absent from gonads, while it was ubiquitous in the *Hsp20*-line.
Conclusion: The functional elements of the heat stress activation of the *M. lignano* LTR-RTs are located within their LTRs and regulated by the same mechanism as that for *Hsp* genes.
Acknowledgements: The study was supported by the RFBR grant No. 18-34-00288.

References

1. Vavilova V. et al. Heat shock response elements are present in the promoters of the heat stress activated LTR retrotransposons in the free-living regenerative flatworm *Macrostomum lignano*. Systems Biology and Bioinformatics (SBB-2018). The Tenth International Young Scientists School. – Novosibirsk, 2018;46.
2. Wudarski J. et al. Efficient transgenesis and annotated genome sequence of the regenerative flatworm model *Macrostomum lignano*. *Nat Commun*. 2017;8:2120.
3. Wudarski J. et al. Influence of temperature on the development, reproduction and regeneration in the flatworm model organism *Macrostomum lignano*. *Zoological Letters.* 2019;5:7.

DOI 10.18699/SBB-2019-42

CYP75 gene family in barley

Vikhorev A.V.[1, 2]*, Gerasimova S.V.[1], Strygina K.V.[1, 3], Khlestkina E.K.[1, 3]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *N.I. Vavilov Research Institute of Plant Industry, St. Petersburg, Russia*
** e-mail: vikhorev@bionet.nsc.ru*

Key words: anthocyanin biosynthesis, CYP75, flavonoid pigments, gene evolution, Hordeum, P450

Motivation and Aim: Anthocyanins are secondary metabolites of plants playing multiple biological functions. These compounds could be accumulated in many plants, including barley (*Hordeum vulgare* L.). Important structural genes in the pathway of anthocyanin biosynthesis are *flavonoid 3'-hydroxylase* (*F3'H*, *CYP75B*) and *flavonoid 3',5'-hydroxylase* (*F3'5'H*, *CYP75A*). These genes catalyze the hydroxylation of B-ring of anthocyanins and direct the pathway of biosynthesis towards the formation of magenta pigments cyanidins and blue pigments delphinidins, respectively [1]. However, these genes are poorly studied in barley. The aim of this study was to identify and characterize duplicated copies of the *F3'H* and *F3'5'H* genes in the barley genome.
Methods and Algorithms: Homologous sequences of *F3'H* and *F3'5'H* was found in IPK Barley BLAST Server database using BLASTN search. *In silico* analysis of the genes was carried out using Multalin, FGENESH+, New PLACE and InterPro software. The construction of the UPGMA tree, the calculation of *Ka*/*Ks* ratio and calculation of divergence time was performed using MEGA v.606 software with 1000 bootstrap replicates to assess the branch support. Analysis of expression of genes was made using qPCR.
Results: In barley genome we identified one homologous copy belonging to the family *CYP75B* (chromosome 6HS) and three homologous copies belonging to the family *CYP75A* (chromosomes 6HL, 6HS and 7HS). All found genes have a cytochrome P450 domain, except for the pseudogenic copy designated *F3'5'H-3* (6HS). It was shown that genes *F3'H-1* and *F3'H-2* are responsible for the accumulation of magenta pigments in the pericarp and stem, respectively. The gene *F3'5'H-1* is responsible for the accumulation of blue pigments in the aleurone layer of barley grain. Genes *F3'5'H-2*, *F3'5'H-3* and *F3'5'H-4* do not participate in the biosynthesis of anthocyanins in the studied tissues.
Conclusion: In current work one new *F3'H* and three *F3'5'H* genes were described for the first time. The subfunctionalization of *F3'H-1*, *F3'H-2* and *F3'5'H-1* genes expressing in tissue-specific activity and pseudogenization of *F3'5'H-3* gene were shown. The results obtained are important for understanding the features of anthocyanin biosynthesis in barley and could be used for the manipulation of the content of diverse anthocyanins in different parts of barley.
Acknowledgements: This study was partially by the RFBR (No. 18-416-543007).

References

1. Shoeva O.Y., Mock H.P., Kukoeva T.V., Börner A., Khlestkina E.K. Regulation of the flavonoid biosynthesis pathway genes in purple and black grains of *Hordeum vulgare*. *PloS ONE*. 2016; 11(10):e0163782.

DOI 10.18699/SBB-2019-43

Adaptation of the CRISPR/Cas9 system for targeted manipulations of the human mitochondrial genome

Zakirova E.G.[1]*, Vereshchagina N.A.[2], Mazunin I.O.[2], Orishchenko K.E.[1, 2]
[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*
[2] *I. Kant Baltic Federal University, Kaliningrad, Russia*
** e-mail: zakirova@bionet.nsc.ru*

Key words: mitochondrial DNA editing, cybrids, MitoCRISPR/Cas9

Motivation and Aim: Repopulation of healthy copies of mtDNA in a multi-copy mitochondrial genome can be done using various technologies based on site-specific nucleases. Mitochondria-imported restriction endonucleases, ZFN and TALE-nucleases have already demonstrated the possibility of eliminating mutant copies of mtDNA, but the well-known CRISPR/Cas9 system have not been adapted for this task yet. Here we propose a strategy for modifying the components of the CRISPR/Cas9 to manipulate mitochondrial DNA (mtDNA) haplotype level in a cell.
Methods and Algorithms: A stable and uniform distribution of Cas9 nuclease among cells was obtained by integrating the Cas9 gene, that contained the mitochondrial localization signal on 5'-end, into the genome of cybrid cell lines NARP3-1 and NARP3-2. To transfer a guide RNA (gRNA) – the second component of the CRISPR / Cas9 system – we added mitochondria-specific determinants of import of nucleic acids into the scaffold of gRNA. We checked, whether these modifications affected the functional activity of the system. Genetic constructs were designed in SnapGene software. The plasmids were assembled using the classical methods of molecular cloning and the Gibson assembly cloning method (NEB, USA). Integration of the nuclease gene was performed by Sleeping Beauty transposon system.
Results: We confirmed intramitochondrial localization of the modified nuclease, expressed from the cell nucleus, by immunocytochemistry analysis of the obtained stable cell lines NARP3-1-MitoCas9 and NARP3-2-MitoCas9. *In vitro* cleavage analysis of the template DNA using ribonucleoprotein complex showed that both of the guide RNA, without determinant in the RNA scaffold (NEG), and the modified variants of gRNA do not change the functional activity of the CRISPR/Cas9 system.
Conclusion: The derived transgenic cybrid cell lines allow performing further analysis of the effect of the MitoCRISPR / Cas9 system on the functional level without additional transfection with a plasmid encoding nuclease. Also we found that the CRISPR / Cas9 system remains functional activity after modifications in gRNA structure. Our results contribute to the further study of innovative technology of mitochondrial pathologies treatment – the MitoCRISPR/Cas9 system.
Acknowledgements: The work was supported by the Russian Foundation for Basic Research (19-015-00084) and the State budget project (0324-2019-0042).

DOI 10.18699/SBB-2019-44

Gene networks of type 2 diabetes and Alzheimer's disease. Reconstruction and analysis

Zamyatin V.[1, 2]*, Mustafin Z.S.[1, 2], Matushkin Y.G.[1, 2], Klimontov V.V.[1, 2], Lashin S.A.[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: gene networks, SNP, T2D, diabetes, AD, Alzheimer's disease

Motivation and Aim: Type 2 diabetes (T2D) and Alzheimer's disease (AD) are multifactorial diseases caused by a combination of genetic and environmental factors. Today, systems biology approaches and network approaches are powerful tools for studying multifactor diseases. They are based on knowledge of physical and/or functional interactions between molecules, usually represented as networks. Such networks not only report connections between individual nodes, but also implicitly organize high-level cellular connections. In this study, we consider networks of transcriptional regulation, protein-protein interactions, and functional connectivity, leading them to the general meaning of gene networks – molecular-genetic systems ensuring the phenotypic properties development of organisms based on genetic information [1].
Methods and Algorithms: Based on an analysis of literature data, as well as a number of databases, sets of genes associated with AD and T2D were formed. Among them, transcription factors (TF) were identified. HOCOMOCO database was used for search the binding sites of TF in 5'-noncoding regions of genes. The phylostratigraphic age and the divergence index were calculated for each gene using Orthoscape software [2]. The intrapopulation variability was estimated using the data on the SNP of the project "1000 human genomes". Based on the transcriptomes found, differential gene expression was analyzed. Based on the "raw" data from type 2 diabetes mellitus and Alzheimer's disease researches, a search was made for genome-wide associations. Gene networks were reconstructed using Cytoscape based on interaction data extracted from GeneMANIA and StringDB.
Results: After applying the information obtained earlier, an analysis of the gene networks was carried out. Were identified regulatory circuits, violations in the elements of which can lead to the diseases. They contain new genes that were not previously associated with T2D.
Conclusion: Analysis of gene networks is an effective tool for a comprehensive theoretical study of human diseases and it can be used to not only search for and prioritize genes associated with diseases, but also to explain the molecular genetic mechanisms behind them.
Acknowledgements: The study was supported by the Project 0324-2018-0021 SB RAS.

References

1. Introduction to informational biology and bioinformatics: study guide: eds N.A. Kolchanov, O.V. Vishnevsky, D.P. Furman. – Novosibirsk, 2015;3 (in Russian).
2. Mustafin Z.S. et al. Orthoscape: A cytoscape application for grouping and visualization KEGG based gene networks by taxonomy and homology principles. *BMC Bioinformatics*. 2017;18.

DOI 10.18699/SBB-2019-45

Metagenomic search for the protozoa in atherosclerotic plaques

Zarubin A.[1]*, Markov A.[1], Sharysh D.[2], Puzyrev V.[1,2], Nazarenko M.[1,2]
[1] *Research Institute of Medical Genetics of Tomsk NRMC, Tomsk, Russia*
[2] *Siberian State Medical University, Tomsk, Russia*
** e-mail: aleksei.zarubin@medgenetics.ru*

Key words: metagenome, protozoa, atherosclerotic plaques

Motivation and Aim: Several reports have shown the relevance of infection in development of atherosclerosis. However, the bacterial DNA of atherosclerotic plaques mainly was investigated by metagenomic approaches. To search for the protozoa by using genome-wide sequencing data of human atherosclerotic plaques (AP).
Material and Methods: We used the results of whole genome sequencing of DNA isolated from 12 atherosclerotic plaques (SRA149235) and 2 blood samples from 1000 Genomes Project (HG00096, HG00099). Sequence reads were aligned to the human reference genome (hg19) using Bowtie2. The taxonomic classification of the reads was performed with Kraken2 using a curated microbial genome database containing archaea, bacteria, fungi, protozoa, viruses reference genome sequence.
Results: Pre-alignment on the human genome, even with "very-sensitive" option of Bowtie2, cannot completely clean up human gene sequences. Kraken2 determined 13 to 92 % of the human gene sequences in unmatched reads of different AP samples. We found sequences of Toxoplasma gondii genome in 8 samples (about 1678 reads per sample). These sequences were shown to be mapped to contigs (NW_017384310.1, NW_017384809.1). These contigs were also found in samples from 1000 Genomes Project.
Conclusion: There were no protozoa in genome-wide sequencing data of human atherosclerotic plaques. The reads that mapped to the Toxoplasma genome are false positive. Some microbial reference genomes probably to contain human DNA sequences not presented in the human reference genome assembly.

DOI 10.18699/SBB-2019-46

Inverse problem of information diffusion in online social networks

Zvonareva T.[1]*, Krivorotko O.[1, 2], Kabanikhin S.[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Computational Mathematics and Mathematical Geophysics, SB RAS, Novosibirsk, Russia*
** e-mail: zvonareva-tanyushka@mail.ru*

Key words: inverse problems, partial differential equations, social processes, diffusive logistic model, optimization, particle swarm optimization

Motivation and Aim: Social network are rapidly progress at the present time: access for informing relatives (Facebook's Safety Check service), publishing data about missing friends (the Google Person Finder service) or promptly informing users of an impending threat and their further actions in case of emergencies (Alerts from Twitter). Such processes can be described by the diffusive logistic mathematical model based on the partial differential equation (PDE) and characterized information dissemination in social networks [1]. The type of information is determined by the coefficients of the mathematical model and the initial conditions of the problem. To control and predict of information in social networks it is necessary to refine the model coefficients and initial data by some additional measurements in fixed time points (the inverse problem) [2].
Methods and Algorithms: One way to solve the inverse problem for PDE is to reduce it to an optimization problem, where the misfit function characterizes the quadratic deviation of the model data from the experimental one. A particle swarm optimization method (PSO) [3] is applied to solve the optimization problem. The algorithm works with a population (called a swarm) of candidate solutions (called particles). It solves a problem by moving these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity.
Results: The reconstructed coefficients and initial data of mathematical models of social networks allow one to refine the information dynamic and give the recommendation for control it.
Acknowledgements: Supported by the Russian Science Foundation (project No. 18-71-10044).

References

1. Wang H., Wang F., Xu K. Modeling information diffusion in online social networks with partial differential equations. *arXiv*. 2013;1310.0505.
2. Kabanikhin S.I. Definitions and examples of inverse and ill-posed problem. *Journal Inverse Ill-Posed Problems*. 2008;16(4):317-357.
3. Zhang Y. A comprehensive survey on particle swarm optimization algorithm and its applications. *Mathematical Problems Engineering*. 2015;931256.

Author index

Научное издание

SYSTEMS BIOLOGY
AND BIOINFORMATICS (SBB-2019)

The Eleventh International Young Scientists School

Abstracts

Printed without editing

СИСТЕМНАЯ БИОЛОГИЯ
И БИОИНФОРМАТИКА (SBB-2019)

Одиннадцатая школа молодых ученых

Тезисы докладов

Публикуется в авторской редакции

Выпуск подготовлен информационно-издательским отделом ИЦиГ СО РАН

Подписано к печати 05.06.2019. Формат 70 × 108 $^{1}/_{16}$. Усл. печ. л. 4.9.
Тираж 50 экз. Заказ № 129

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

Отпечатано в типографии Новосибирского государственного университета
630090, Новосибирск, ул. Пирогова, 2

From 24 to 28 June, 2019 Institute of Cytology and Genetics, SB RAS (Novosibirsk, Russia) holds the Annual 11th International Young Scientists School “Systems Biology and Bioinformatics” – SBB-2019. Bioinformatics is a rapidly developing field of knowledge. Each of the past Young Scientists Schools concerned topics urgent at those times. To preserve this tradition, the SBB-2019 will be dedicated to bioinformatics and systems biology in such areas as:

- Genomics and transcriptomics
- Evolutionary bioinformatics
- Structural biology: bioinformatics and molecular dynamics
- Systems biology: analysis of molecular-genetic systems and gene network modeling
- Biomedicine
- Genetics, genomics, bioinformatics and biotechnology of plants.

ISBN 978-5-91291-046-3