World Scientific
www.worldscientific.com

Introduction to the 9th Young Scientists School on Systems Biology and Bioinformatics (SBB'2017)

Published 13 February 2018

This special issue presents materials from the 9th International Young Scientists School on Systems Biology and Bioinformatics (SBB'2017), organized in June 2017 in Yalta, Russia.

The Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences (ICG SB RAS) hosts the International Multi-conference on Bioinformatics of Genome Regulation and Structure\Systems Biology (BGRS\SB) every two years beginning from 1998. From BGRS\SB'2008 onwards, the Young Scientists School on Systems Biology and Bioinformatics (SBB) runs as a satellite event following the BGRS\SB conference or as a standalone annual event (http://conf.bionet.nsc.ru/sbb2017/en/archive/). Since the first meeting, the SBB has grown to a large international event. Gradually, the initial focus has been extended from systems biology and classical bioinformatics topics to gene network analysis and reconstruction, and omics technologies.[1]

The *Journal of Bioinformatics and Computational Biology* (*JBCB*) publishes special issues on bioinformatics, algorithms, network analysis dedicated to BGRS\SB. The first *JBCB* special issue in 2006 highlighted BGRS\SB-2006.[2] Then *JBCB* published special issues on the 2012, 2014, and 2016 conferences.[3–5] Additionally, the journal publishes reports from the SBB schools. For instance, *JBCB* has published proceedings of SBB-2015 on modeling of gene network based on material presented at earlier BGRS\SB meetings.[2,4,5]

To continue traditions of the BGRS conference series in 2017, the Institute of Cytology and Genetics SB RAS organized the Belyaev Conference-2017 on genetics and evolution, dedicated to the 100th anniversary of Academician, Professor Dmitry K. Belyaev (1917–1985), an outstanding scientist, evolutionist and geneticist.[6] The works of Belyaev Conference-2017 and SBB'2017 School on computational biology were recently covered in special issues of several international journals: the *BMC Evolutionary Biology*,[6] *BMC Genetics*,[7] *BMC Plant Biology, BMC Genomics*, *BMC Neuroscience and Vavilov Journal of Selection and Breeding* (http://vavilov.elpub.ru/jour/issue/view/32/showToc).

The current *JBCB* issue presents full papers on bioinformatics and computational biology from SBB'2017, reporting research on promoter prediction, algorithm optimization, mRNA modeling and database applications. These papers are briefly described in what follows.

Krasnov *et al.*[8] presented MethyMer, a Python-based tool aimed at selecting primers for amplification of complete CpG islands. These regions are difficult in terms of selecting appropriate primers because of their low complexity and high GC content. MethyMer has a flexible scoring system that optimizes the balance between various characteristics such as nucleotide composition, thermodynamic features, presence of CpG sites and primer specificity, which is assessed with aligning primers to the bisulfite-treated genome using Bowtie. MethyMer incorporates ENCODE genome annotation records (promoter/enhancer/insulator), The Cancer Genome Atlas (TCGA) CpG methylation data derived with Illumina Infinium 450K microarrays, and records on correlations between TCGA RNA-Seq and CpG methylation data for 20 cancer types.

Ryasik *et al.*[9] used a machine learning approach for bacterial promoter prediction. Predicting promoter activity of a DNA fragment is an important task for computational biology. To select an adequate set of physical properties of DNA for training a classifier, various characteristics of DNA molecule should be taken into consideration. A systematic approach allowing selection of the least correlated properties for classification by means of both correlation and cophenetic coefficients as well as concordance matrices was presented. The proposed classifier uses not only sequence and static physical properties of DNA fragment, but also dynamic properties of DNA open state. Therefore, the best performing models with accuracy values up to 90% for all types of sequences were obtained. It was demonstrated that the classifier can serve as a reliable tool enabling promoter DNA fragments to be distinguished from promoter islands despite the similarity of their nucleotide sequences.

Vishnevsky *et al.*[10] presented optimization approaches for high-performance GPU technologies. Their Argo_CUDA web service was designed to process massive DNA data. It is a program for the detection of degenerate oligonucleotide motifs of fixed length written in 15-letter IUPAC code. Despite different tools and stochastic approaches, the efficiency of the motif discovery programs dramatically declines as query set size increases. Argo_CUDA is a full-exhaustive approach based on high-performance GPU technologies. The authors showed effective prediction of ChIP-Seq sequences on simulated sets.

Kiseleva *et al.*[11] describe tools for mass-spectrometry analysis. The necessity of single-thread analysis of bulky data emerged during interpretation of HepG2 proteome profiling results for proteoforms searching. The authors compared contribution of each of eight search engines (X!Tandem, MS-GF+, MS Amanda, MyriMatch, Comet, Tide, Andromeda, and OMSSA) integrated in an open-source graphical user interface SearchGUI into total result of proteoforms identification and optimized set of engines working simultaneously. Most efficient combination of the tools was selected.

Antonov *et al.*[12] described a new computational pipeline, ASSA, combining sequence alignment and thermodynamics-based tools for efficient prediction of RNA–RNA interactions between long transcripts. The discovery of thousands of long noncoding RNAs (lncRNAs) in mammals raises a question about their functionality including prediction of RNA–RNA interactions. Sequence alignment tools have been used for transcriptome-wide prediction of RNA–RNA interactions; but they have poor prediction accuracy since they ignore RNA secondary structure. The novelty of ASSA is its ability to quickly estimate the statistical significance of the observed interaction energies. The authors showed that ASSA outperforms eleven other tools.

Ovchinnikov *et al.*[13] presented a case study for transcription factor binding site search in promoters of intronic and intergenic microRNAs. Some exogenous compounds or xenobiotics may affect microRNA expression. Authors hypothesized that such chemicals can affect miRNA expression through the activation of aryl hydrocarbon receptor (AHR), the nuclear receptor CAR, and estrogen receptors (ESRs). To support this claim, they used *in silico* methods to find potential binding sites for these receptors, and predicted AhR, CAR and ESRs binding sites in rat, mouse and human promoters of miRNA-coding genes that open a new strategy for ongoing experimental studies and will contribute to further investigation of epigenetic mechanisms of carcinogenesis.

Kazantsev *et al.*[14] presented the MAMMOTh: the database for curated MAthematical Models of bioMOlecular systems. MAMMOTh database entries are organized as building blocks in a way that the model parts can be used in different combinations to describe systems with higher organizational level (metabolic pathways and/or transcription regulatory networks). The tool supports export of a single model or their combinations in SBML or Mathematica standards. The database currently contains 110 mathematical sub models for *Escherichia coli* elementary subsystems (enzymatic reactions and gene expression regulatory processes) that can be combined in at least 5100 complex/sophisticated models concerning more complex biological processes as *de novo* nucleotide biosynthesis, aerobic/anaerobic respiration and nitrate/nitrite utilization in *E. coli*.

Bezhentsev *et al.*[15] presented a network pharmacology approach to search potential drug targets for treatment of refractory epilepsy. Epilepsy is the fourth most common neurological disease after a migraine, stroke and Alzheimer's disease. Approximately one-third of all epilepsy cases are refractory to the existing anticonvulsants. Discovery of newer antiepileptic drugs for treating refractory epilepsy is a challenging problem due to unclear etiology of this disease. In this regard, network pharmacology as an area of bioinformatics is gaining popularity. It combines methods of network biology and polypharmacology, which makes it a promising approach for finding new molecular targets. This paper is dedicated to the discovery of new pharmacological targets for treatment of refractory epilepsy using network pharmacology methods.

Overall, this issue collected works on bioinformatics and biomedical applications presented by young scientists at SBB'2017, thus extending traditions of post-conference special journal issues of the BGRS\SB conference series.

Lastly, it is our pleasure to invite you, the readers, to BGRS\SB'2018, 20–25 August 2018, and the Young Scientists School on Systems Biology and Bioinformatics (SBB'2018), 26–30 August 2018 in Novosibirsk, Russia (http://conf.bionet. nsc.ru/bgrssb2018/en/). We look forward to welcoming you in Novosibirsk!

Yuriy L. Orlov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
Institute of Marine Biology Researches of the RAS, Sevastopol, Russia
orlov@bionet.nsu.ru

Tatiana V. Tatarinova
University of La Verne, Los Angeles, CA, USA

Maksim V. Zakhartsev
Norwegian University of Life Sciences, Ås, Norway

Nikolay A. Kolchanov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

Acknowledgments

Hefei, China), Xiaodong Zhao (Shanghai Jiao Tong University, China), Anastasia Anashkina (Engelhardt Institute of Molecular Biology RAS Moscow, Russia), Liang Sun (University of Delaware, USA), Anna Kudryavtseva (Engelhardt Institute of Molecular Biology RAS, Moscow, Russia), Bjorn Sommer (University of Konstanz, Germany), Junaid Gamieldien (University of the Western Cape, South Africa), Vladimir Babenko (Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia) and Piramanayagam Shanmughavel (Bharathiar University, Coimbatore, India).

References

1. Baranova AV, Orlov YL, The papers presented at 7th young scientists school "systems biology and bioinformatics" (SBB'15): Introductory note, *BMC Genetics*, **17**(1):20, 2016.
2. Kolchanov NA, Hofestädt R, Orlov YL, Introduction to the selected papers of BGRS'06, *J Bioinform Comput Biol* **5**(2b):vii–viii, 2007.
3. Kolchanov NA, Orlov YL, Introductory note for BGRS-2012 special issue, *J Bioinform Comput Biol* **11**(1):1302001, 2013.
4. Orlov YL, Hofestädt R, Kolchanov NA, Introductory note for BGRS\SB-2014 special issue, *J Bioinform Comput Biol* **13**(1):1502001, 2015.
5. Orlov YL, Kolchanov NA, Hofestädt R, Wong L, Editorial - Bioinformatics development at the BGRS\SB conference series: 10th anniversary, *J Bioinform Comput Biol* **15**(2):1702001, 2017a.
6. Orlov YL, Baranova AV, Herbeck YE, Evolutionary Biology at Belyaev Conference-2017, *BMC Evol Biol* **17**(2):260, 2017b.
7. Orlov YL, Baranova AV, Tatarinova TV, Kolchanov NA, Genetics at Belyaev Conference-2017: Introductory note, *BMC Genetics* **18**(1):116, 2017c.
8. Krasnov GS, Melnikova NV, Lakunina VA, Snezhkina AV, Kudryavtseva AV, Dmitriev AA, MethyMer: Design of combinations of specific primers for bisulfite sequencing of complete CpG islands, *J Bioinform Comput Biol* **16**(1):1840004, 2018.
9. Ryasik A, Orlov M, Zykova E, Ermak T, Sorokin A, Bacterial promoter prediction: Selection of dynamic and static physical properties of DNA for reliable sequence classification, *J Bioinform Comput Biol* **16**(1):1840003, 2018.
10. Vishnevsky OV, Bocharnikov AV, Kolchanov NA, Argo_CUDA: Exhaustive GPU based approach for motif discovery in large DNA datasets, *J Bioinform Comput Biol* **16**(1):1740012, 2018.
11. Kiseleva O, Poverennaya E, Shargunov A, Lisitsa A, Proteomic Cinderella: Customized analysis of bulky MS/MS data in one night, *J Bioinform Comput Biol* **16**(1):1740011, 2018.
12. Antonov I, Marakhonov A, Zamkova M, Medvedeva Y, ASSA: Fast identification of statistically significant interactions between long RNAs, *J Bioinform Comput Biol* **16**(1):1840001, 2018.
13. Ovchinnikov VY, Antonets DV, Gulyaeva LF, The search of CAR, AhR, ESRs binding sites in promoters of intronic and intergenic microRNAs, *J Bioinform Comput Biol* **16**(1):1750029, 2018.
14. Kazantsev F, Akberdin I, Lashin S, Ree N, Timonov V, Ratushny A, Khlebodarova T, Likhoshvai V, MAMMOTh: A new database for curated mathematical models of biomolecular systems, *J Bioinform Comput Biol* **16**(1):1740010, 2018.
15. Bezhentsev V, Ivanov S, Kumar S, Goel R, Poroikov V, Identification of potential drug targets for treatment of refractory epilepsy using network pharmacology, *J Bioinform Comput Biol* **16**(1):1840002, 2018.