

## Gene network reconstruction and functional association studies for oncological diseases using online bioinformatics tools

Shakirova R.R.<sup>1\*</sup>, Elistratova M.G.<sup>1</sup>, Korobeynikova A.V.<sup>1</sup>, Mozyleva Y.A.<sup>1</sup>, Orlova N.G.<sup>2,3</sup>, Orlov Y.L.<sup>1,4</sup>

<sup>1</sup> *The Digital Health Institute, I.M. Sechenov First Moscow State Medical University of the Russian Ministry of Health (Sechenov University), Moscow, Russia*

<sup>2</sup> *Agrarian and Technological Institute, Peoples' Friendship University of Russia, Moscow, Russia*

<sup>3</sup> *Financial University under the Government of the Russian Federation, Moscow, Russia*

<sup>4</sup> *Moscow State Technical University of Civil Aviation, Moscow, Russia*

\* [y.orlov@sechenov.ru](mailto:y.orlov@sechenov.ru)

**Key words:** biomedicine, oncology, complex disorders, gene network, bioinformatics, databases, e-Health

*Motivation and Aim:* Functional annotation of genes and gene ontology analysis for complex diseases using available bioinformatics tools provide background to search new targets for therapy. We aimed develop computer pipelines to study set of complex heterogeneous disease such as cancers and mental disorders using online bioinformatics tools. Examples of gene lists studied are related to glioma, glioblastoma, cystic fibrosis, Kaposi's sarcoma, Parkinson's disease and dementia. Thus, cystic fibrosis a systemic hereditary disease caused by a mutation in the gene for the transmembrane regulator of cystic fibrosis and characterized by damage to the external secretion glands, severe respiratory dysfunction. Parkinson's disease and dementia present socially important complex diseases. Functional annotation of genes related to the disease could be retrieved based on genetic databases and cross-validated by integrating complementary experimental data. Gene network reconstruction for a set of genes (proteins) proved to be effective approach to study mechanisms underlying disease progression.

*Methods and Algorithms:* We used online bioinformatics tools such as GeneCards, GeneMANIA (<https://genemania.org/>) and STRING-DB (<https://string-db.org/>) for annotation of gene list for the disease, reconstruction of gene network and comparative analysis of gene ontology categories. The Internet-resource OMIM (Online Mendelian Inheritance in Man) (<https://omim.org/>) and GeneCards suit (GeneCards.org) were used for gene list reconstruction. DAVID (<https://david.ncifcrf.gov/>) and PANTHER resources (Protein Analysis THrough Evolutionary Relationships) (<http://pantherdb.org/>) were used to analyze the categories of gene ontologies. The issue of developing bioinformatics courses is related to the adaptation of the training to the educational profile of students and trainees. Examples of the tasks in medical informatics, which can be solved using only online tools – the formation of lists of genes associated with the disease, reconstruction of the associative gene network, visualization of the gene network, calculation of categories of gene ontologies, overrepresented in a given list of genes, determination of the spatial structure of the protein. These tasks students are able to solve independently, using online tools, with the preparation of abstracts and their own publications as it was discussed at series of students workshops [6] and Younf Scientists Schools on bioinformatics [8].

*Results:* We reconstructed gene networks for series of oncological diseases – glioma, lymphoma, and Kaposi's sarcoma. The work on gene network models include

glioma [2], Parkinson's disease [5], and metabolic syndrome [9]. The steps include collection of a list of genes associated with the development of the disease (e.g., Parkinson's disease), analyze gene ontology categories for such a list, and reconstruct the gene network. For the key disease genes derived from the gene network structure analysis, drug search options are considered [5]. This approach has already been tested for teaching the course "Fundamentals of Bioinformatics and Databases Management" at the Sechenov University [1, 2]. In many cases, it requires advanced training, additional education in modern genomics technologies, such as high-throughput genomic and transcriptomic sequencing, data analysis, metagenome sequencing, non-coding RNA studies, Artificial Intelligence (AI) applications in medicine, which greatly expand the range of classical bioinformatics tasks [4].

*Conclusion:* Overall, the goal of the work is to use bioinformatics online tools to solve practical tasks in biomedicine. The development of next-gen sequencing technologies poses new challenges which can be solved by already existing online tools without the use of programming skills, in remote access. The issue of developing bioinformatics courses is related to the need to adapt training depending on the educational profile of the trainees [6]. The transition to distance learning requires the development of new methods of teaching, with the ability of students to perform independent work using only Internet resources, which in turn meets a number of challenges due to limited access to a number of online services in Russia (for example, data storage in Google cloud storages) [3].

*Acknowledgements:* The study is supported by the Potanin Foundation grant for masters' teachers (ГК22-000797).

### References

1. Dergilev A.I., Orlova N.G., Dobrovolskay O.B., Orlov Y.L. Statistical estimates of transcription factor binding site clusters in plant genomes based on genome-wide data. *J Integr Bioinform.* 2021;18:20200036. doi: 10.1515/jib-2020-0036.
2. Gubanova N.V., Orlova N.G., Dergilev A.I., Oparina N.Y., Orlov Y.L. Glioblastoma gene network reconstruction and ontology analysis by online bioinformatics tools. *J Integr Bioinform.* 2021;18:20210031. doi: 10.1515/jib-2021-0031.
3. Gusev A.V., Ivshin A.A., Vladzimirsky A.V. Russian mobile health apps: systematic search in app stores. *Russ J Telemed E-Health.* 2021;7(3):21-31. doi: 10.29188/2712-9217-2021-7-3-21-31. (in Russian)
4. Koshechkin K., Lebedev G., Radzievsky G., Seepold R., Martinez N. Review of blockchain technology projects to provide telemedical services. *J Med Internet Res.* 2021;23(8):e17475. doi: 10.2196/17475.
5. Orlov Y.L. et al. Reconstruction of the Parkinson's disease gene network to search for target genes. *Biomed Chemistry.* 2021;67(3):222-230. doi:10.18097/PBMC20216703222. (In Russian)
6. Orlov Y.L., Bakulina A.Y. Development of Education in Bioinformatics Based on Student Conferences ISSC-2018, School of Molecular Modeling and Hackathon in Novosibirsk. *Vestnik Novosibirsk State University. Series: Information Technologies.* 2018;16(3):5-6. (In Russian)
7. Shaderkin I.A., Shaderkina V.A. Remote medical consultations for patients: what has changed in Russia in 20 years. *Russ J Telemed E-Health.* 2021;7(2):7-17. doi: 10.29188/2712-9217-2021-7-2-7-17. (In Russian)
8. Tatarinova T.V., Tabikhanova L.E., Eslami G., Bai H., Orlov Y.L. Genetics researches at the "Centenary of human population genetics" conference and SBB-2019. *BMC Genetics.* 2020;21(Suppl 1):109. doi: 10.1186/s12863-020-00906-7.
9. Tiis R.P., Osipova L.P., Galieva E.R. et al. N-acetyltransferase (NAT2) gene polymorphism and gene network analysis. *Biomeditsinskaya Khimiya.* 2021;67(3):213-221. doi: 10.18097/PBMC20216703213.