

Automated reconstruction and analysis of gene networks using cognitive system ANDSystem

Ivanisenko V.A.^{1, 2, 3*}, Demenkov P.S.^{1, 2, 3}, Ivanisenko T.V.^{1, 2, 3}, Antropova E.A.¹, Makarova A.L.A.^{1, 2, 3}, Adamovskaya A.V.^{1, 2, 3}, Venzel A.S.^{1, 2, 3}, Kolchanov N.A.^{1, 2, 3}

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

² *Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

³ *Novosibirsk State University, Novosibirsk, Russia*

* *salix@bionet.nsc.ru*

Cognitive computer systems are gaining increasing interest in the field of biomedicine for the reconstruction and analysis of gene networks. These systems offer a complete cycle of knowledge engineering, including entity recognition, relationship establishment, knowledge integration, and representation in the form of semantic networks or knowledge graphs. Additionally, they provide a range of tools for data analysis and interpretation, such as gene prioritization and identification of overrepresented biological processes or diseases.

We have developed a cognitive software and information system, ANDSystem [1], which considers 12 types of objects, including organisms, cells/tissues, diseases, genes, metabolites, drugs, etc. The system extracts knowledge about physical interactions, catalytic reactions, regulatory relationships, and more from scientific texts. The ANDSystem knowledge base, created based on large-scale automatic analysis of over 30 million PubMed abstracts and full-text articles from PMC, as well as hundreds of factual databases in the field of biology and biomedicine, is a unique resource containing formalized knowledge about more than 50 million interactions between molecular-genetic objects of different organisms. The developed system surpasses similar known systems in the world in terms of the number of considered types of relationships and types of objects.

ANDSystem has been used to solve a wide range of problems in the reconstruction of molecular mechanisms of diseases, the interpretation of omics data, and other problems in the field of biomedicine. In particular, we used ANDSystem to study the molecular genetic mechanisms of pathogen-host interaction.

Reconstruction of signaling pathways for the regulation of cellular biological processes by viral proteins can help in the search for new pharmacological targets. Using ANDSystem, we reconstructed associative gene networks describing the potential regulation of the apoptosis process by HCV viral proteins. Expanding the given gene network with consideration of information about aberrantly methylated genes in hepatocellular carcinoma enabled the prioritization of new promising pharmacological targets for liver cancer therapy. Furthermore, we reconstructed potential molecular mechanisms of dysfunction of signaling pathways in hepatocellular carcinoma.

The application of ANDSystem to the interpretation of metabolomics data, in particular, allowed us to investigate the regulatory role of SARS-CoV-2 non-structural proteins in the disruption of metabolic processes in patients diagnosed with COVID-19. The application of ANDSystem to the analysis of blood plasma metabolomic screening data from patients with postoperative delirium allowed us to reconstruct the molecular genetic mechanisms of sphingolipid metabolism disorders.

In conclusion, ANDSystem is a powerful cognitive system for the reconstruction and analysis of gene networks with a unique knowledge base that surpasses similar known systems in the world. The system has been successfully applied to solve a wide range of problems in the field of biomedicine.

Funding: The bioinformatics analysis was supported by the budget project, FWNR-2022-0020. The authors express their gratitude to Evgeniy V. Gaisler, Artem D. Rogachev, Nikita V. Basov, Andrey G. Pokrovsky for obtaining and analyzing the metabolomics data.

References

1. Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics*. 2019;20(Suppl. 1):34