

ANDSystem: cognitive computer system for automated gene networks reconstruction and analysis

Ivanisenko V.A.^{1,2,3*}, Demenkov P.S.^{1,2}, Ivanisenko T.V.^{1,2}, Ivanisenko N.V.^{1,2}, Antropova E.A.¹, Adamovskaya A.V.^{1,3}, Revva P.M.^{1,3}, Lavrik I.N.^{1,4}, Kolchanov N.A.^{1,3}

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*

⁴ *Translational Inflammation Research, Medical Faculty, Center of Dynamic Systems, Otto von Guericke University Magdeburg, Magdeburg, Germany*

* salix@bionet.nsc.ru

The main source of data for the reconstruction of gene networks are scientific publications. Among the existing software and information systems that use methods of automatic knowledge extraction from scientific texts, cognitive computer systems are of particular interest. A distinctive feature of these systems is the full cycle of knowledge engineering, which includes the stage of identifying entities in the text, establishing relationships between them, integrating the acquired knowledge and presenting it in the form of semantic networks or a knowledge graph. In addition, a characteristic feature of such systems is the availability of a wide range of tools designed for data analysis and interpretation, including gene prioritization, identification of overrepresented biological processes or diseases, etc. Examples of cognitive systems in the field of biomedicine are Watson [1], Ingenuity Pathway Analysis (IPA) [2], STRING [3], Pathway Studio [4]. We have developed a cognitive software and information system ANDSystem [5-7]. The ANDSystem considers 12 types of objects, including organisms, cells/tissues, diseases, genes, metabolites, drugs, etc. For these objects, knowledge about physical interactions, catalytic reactions, regulatory relationships, such as regulation of expression, regulation of activity, regulation of stability, etc. is extracted from texts. There are more than 20 types of interactions in total.

ANDSystem has been used to solve a wide range of problems in the reconstruction of the molecular mechanisms of diseases, the interpretation of omics data, and other problems in the field of biomedicine. In particular, an urgent problem is the study of 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 the ANDSystem, the associative gene networks describing the potential regulation of the apoptosis process by HCV viral proteins were reconstructed. Another direction of research was related to the study of the molecular mechanisms of pathological processes in hepatocellular carcinoma. Induction of apoptosis by the cell response to mechanical stress caused by tumor tissue compaction is important antitumor mechanism [8, 9]. The analysis of gene networks made it possible to identify potential signaling pathways linking genes involved in the cell response to mechanical stress with key genes of the external pathway of apoptosis. Potential molecular mechanisms of dysfunction of these signaling pathways in hepatocellular carcinoma have also been reconstructed.

Acknowledgements: Work was funded by the Ministry of Science and Higher Education of the Russian Federation project “Kurchatov Center for World-Class Genomic Research” No. 075-15-2019-1662 from 2019-10-31.

References

1. Ying C., Argentinis J.D., Weber G. IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. *Clin Ther.* 2016;38:688-701.
2. Krämer A., Green J., Pollard Jr.J., Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics.* 2014;30(4):523-530.
3. Szklarczyk D., Gable A.L., Lyon D. et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019;47(D1):D607-613.
4. Nikitin A., Egorov S., Daraselia N., Mazo I. Pathway studio – the analysis and navigation of molecular networks. *Bioinformatics.* 2003;19(16):2155-7.
5. Ivanisenko T.V., Saik O.V., Demenkov P.S., Ivanisenko N.V., Savostianov A.N., Ivanisenko V.A. ANDDigest: a new web-based module of ANDSystem for the search of knowledge in the scientific literature. *BMC Bioinformatics.* 2020;21:228.
6. 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.
7. Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. *BMC Systems Biol.* 2015;9(Suppl 2):S2.
8. Levayer R. Solid stress, competition for space and cancer: The opposing roles of mechanical cell competition in tumour initiation and growth. In: *Seminars in cancer biology.* 2020;63:69-80.
9. Hong E., Lee E., Kim J., Kwon D., Lim Y. Elevated pressure enhanced TRAIL-induced apoptosis in hepatocellular carcinoma cells via ERK1/2-inactivation. *Cell Mol Biol Letters.* 2015;20(4):535-548.