

Reconstruction of molecular genetic networks: text mining as a tool for knowledge discovery.

Vladimir A. Ivanisenko

Institute of Cytology and Genetics, Novosibirsk, Russia

Computational systems (ANDCell and ANDNanobiotech) for automated extraction of knowledge from scientific texts (PubMed abstracts) and databases were developed. The ANDCell is targeted on the reconstruction of semantic associative networks in systems biology (molecular genetic interactions, gene regulation events, catalytic processes, proteolysis, polymorphic gene – disease associations, etc.). The ANDNanobiotech is targeted on nanobiotechnology (drug-delivery systems, labeling, diagnostics, treatment, microfluidics, implants, grafts, biosensors among others.). Information is extracted with the aid of original text-mining technology. The ANDCell knowledge database contains about 5 millions of facts, the ANDNanobiotech knowledge database contains about 1,5 millions of facts. The ANDVisio program provides the access to the knowledge databases and the representation of the results in a graphic form of reconstructed associative networks. The vertices of such networks are the molecular genetic objects (genes, proteins, microRNAs, metabolites, cellular components), diseases, processes, nanomaterials and nanobioconstructs while the edges between the vertices represent various types of associations. For molecular interactions and associations, data on the cell types and organisms are represented. The systems provide a user's friendly interface implemented links to the molecular-genetic databases and the articles from which information was extracted. The developed systems may be useful for resolving a wide range of tasks in biology, biomedicine and nanobiotechnology.

#### Acknowledgements

This work was supported in part by: Ministry of Science & Education (02.740.11.0882), Interdisciplinary integrative Project of SB RAS (94, 111, 113, 119), Program of RAS (19, 21, 22, 23).

1. Bogachek MV, Zaitsev BN, Sekatskii SK, Protopopova EV, Ternovoi VA, Ivanova AV, Kachko AV, Ivanisenko VA, Dietler G, Loktev VB. Characterization of glycoprotein E C-end of west nile virus and evaluation of its interaction force with alphaVbeta3 integrin as putative cellular receptor. *Biochemistry (Mosc)*. 2010 Apr;75(4):472-80.
2. Momynaliev KT, Kashin SV, Chelysheva VV, Selezneva OV, Demina IA, Serebryakova MV, Alexeev D, Ivanisenko VA, Aman E, Govorun VM. Functional divergence of *Helicobacter pylori* related to early gastric cancer. *J Proteome Res*. 2010 Jan;9(1):254-67.
3. Bandyopadhyay A, Arora A, Jain S, Laskar A, Mandal C, Ivanisenko VA, Fomin ES, Pintus SS, Kolchanov NA, Maiti S, Ramachandran S. Expression and molecular characterization of the *Mycobacterium tuberculosis* PII protein. *J Biochem*. 2010 Feb;147(2):279-89.
4. Malygin AA, Bondarenko EI, Ivanisenko VA, Protopopova EV, Karpova GG, Loktev VB. C-terminal fragment of human laminin-binding protein contains a receptor domain for venezuelan equine encephalitis and tick-borne encephalitis viruses. *Biochemistry (Mosc)*. 2009 Dec;74(12):1328-36.
5. Ivanisenko VA, Afonnikov DA, Kolchanov NA. Web-based computational tools for the prediction and analysis of post-translational modifications of proteins. *Methods Mol Biol*. 2008;446:363-84.
6. Kochetov AV, Ahmad S, Ivanisenko V, Volkova OA, Kolchanov NA, Sarai A. uORFs, reinitiation and alternative translation start sites in human mRNAs. *FEBS Lett*. 2008 Apr 16;582(9):1293-7.

7. Demenkov PS, Aman EE, Ivanisenko VA. Associative network discovery (AND) - computer system to automatic reconstruction of the associative knowledge networks about molecular-genetic interactions (2008) Computational technology. 13, N 2. pp. 15-19. Russian
8. Mishchenko EL, Bezmaternykh KD, Likhoshvai VA, Ratushny AV, Khlebodarova TM, Yu Sournina N, Ivanisenko VA, Kolchanov NA. Mathematical model for suppression of subgenomic hepatitis C virus RNA replication in cell culture. J Bioinform Comput Biol. 2007 Apr;5(2B):593-609.
9. Ivanisenko VA, Eroshkin AM, Kolchanov NA. WebProAnalyst: an interactive tool for analysis of quantitative structure-activity relationships in protein families. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W99-104.
10. Ivanisenko VA, Pintus SS, Grigorovich DA, Kolchanov NA. PDBSite: a database of the 3D structure of protein functional sites. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D183-7.
11. Ivanisenko VA, Pintus SS, Grigorovich DA, Kolchanov NA. PDBSiteScan: a program for searching for active, binding and posttranslational modification sites in the 3D structures of proteins. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W549-54.