Evaluation and Visualization of Pathway Efficiency based on Subcellular Protein Localizations

Olga V. Popik[1,*], Björn Sommer[2], Ralf Hofestädt[2], Vladimir A. Ivanisenko[1]

[1] Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia

[2] Bio-/Medical Informatics Department, Bielefeld University, Bielefeld, Germany

1 Introduction

Intracellular protein localization plays an important role in cell functioning as biological processes are usually distributed over different intracellular components. In some cell components processes might proceed with a high rate, and in others the rate is low.

For visualizing and analyzing subcellular localizations, we presented the subcellular localization charts in 2013 [1]. After a concrete localization scenario is found, it can be visualized in the context of a virtual cell environment. To acquire the heterogeneous localization data, two information sources where combined with the CELLmicrocosmos 4.2 PathwayIntegration, DAWIS-M.D. and ANDCell [2], [3].

Here, instead of analyzing only a subset of proteins or genes, a new approach is proposed to obtain a more global view of protein localizations. For this purpose, the localizations of all KEGG pathways were analyzed [4]. The results are combined in a localization matrix, showing cell component-internal reactions and those reactions occurring between two proteins localized in two different cell components.

2 Localization-based PPI frequency matrices

There are various databases containing data about PPI and intracellular localizations. ANDSystem integrates protein intracellular localization data and information on PPI, extracted from various databases such as UniProt, IntAct, KEGG, etc. as well as text-mining-based information derived from PubMed abstracts [5]. ANDSystem internally uses unique keys, mapping different synonyms to specific terms by using a mapping table.

By using data from ANDSystem we constructed a frequency matrix of PPI for all pair-wise combinations out of 14 intracellular localizations. In order to get a global view of subcellular reaction rates, the intra-localization rates and inter-localization rates are summarized in our approach. We propose that the resulting sum is an appropriate approach to estimate the overall pathway rate. The intracellular localization efficiency matrix $M_{i,j}$: is defined as

$M_{i,j} = \frac{K_{i,j}}{N_i * N_j}$, where

j, *i* are the indexes of the localizations, N_i, N_j are the number of proteins localized in *i* and *j* localizations, and $K_{i,j}$ is the number of interactions between proteins from localization *i* and *j*. The diagonal elements of each matrix show the localization-internal PPI frequency, and all remaining elements show the PPI frequency between different localizations. Initial analysis of the matrices revealed that the frequency of intra-localization PPI is higher than the one of inter-localization PPI.

*popik-olga@bionet.nsc.ru

3 Localization Visualization

The final matrix was used to generate a subcellular localization scenario by using the CELLmicrocosmos 4.2.1 PathwayIntegration [1], [6]. Figure 1 shows the result. The human KEGG pathways were evaluated using the matrix, ranked and compared with random pathways. The efficiency values can elucidate the problem on process flow differentiation, but to gain a more comprehensive picture, such data should be preferably visualized in 3D. The subcellular visualization is showing which regions are more preferable for highly efficient reactions. In the future, the proposed methods can be completed by data from other databases, such as BRENDA or BioModels, and extended with data concerning protein distribution over tissues.

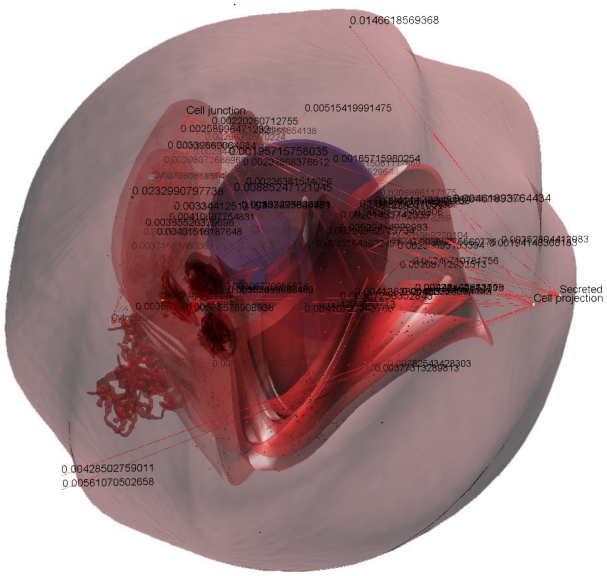

Figure 1: The matrix visualization in the context of an animal cell model using CELLmicrocosmos 4.2.1 PathwayIntegration, correlating the matrix values with the interrelated cell components

References

[1] B. Sommer, B. Kormeier, P. S. Demenkov, P. Arrigo, K. Hippe, Ö. Ates, A. V. Kochetov, V. A. Ivanisenko, N. A. Kolchanov, and R. Hofestädt. Subcellular Localization Charts: A new visual methodology for the semi-automatic localization of protein-related data sets. Journal of Bioinformatics and Computational Biology, 11(1):1340005, 2013.

[2] O. A. Podkolodnaya, E. E. Yarkova, P. S. Demenkov, O. S. Konovalova, V. A. Ivanisenko, and N. A. Kolchanov. Application of the ANDCell computer system to reconstruction and analysis of associative networks describing potential relationships between myopia and glaucoma. Russian Journal of Genetics: Applied Research, 1(1):21–28, 2011.

[3] K. Hippe, B. Kormeier, T. Töpel, S. Janowski, and R. Hofestädt. DAWIS-MD - a data warehouse system for metabolic data. GI Jahrestagung, 2:720–725, 2010.

[4] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic acids research, 40(D1):D109–D114, 2012.

[5] P. S. Demenkov, T. V. Ivanisenko, N. A. Kolchanov, and V. A. Ivanisenko. ANDVisio: A new tool for graphic visualization and analysis of literature mined associative gene networks in the ANDSystem. In silico biology, 11(3):149–161, 2012.

[6] B. Sommer, J. Künsemöller, N. Sand, A. Husemann, M. Rumming, and B. Kormeier. CELLmicrocosmos 4.1: an interactive approach to integrating spatially localized metabolic networks into a virtual 3D cell environment. in *BIOSTEC BIOINFORMATICS 2010*, 2010:90–95.