

Integration and Exploration of Biological Networks

Hendrik Mehlhorn, Falk Schreiber

Leibniz Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben, Germany and
Martin Luther University Halle–Wittenberg, Halle, Germany

mehlhorn@ipk-gatersleben.de, schreibe@ipk-gatersleben.de

Biological networks play a crucial role in solving complex biological problems in the life sciences. Modern wet lab techniques such as GC/MS, multidimensional protein gels, and microarrays produce a continuously increasing amount of biological data sets. These arise from different domains such as metabolomics, proteomics, and transcriptomics. In order to achieve aims of life science projects such as the invention of new drugs or the increase of yield in crop plants, a deep understanding of the complex interactions of biological entities from different domains is necessary. However, the integration of the underlying biological networks is a task which is yet not solved satisfactory due to the lack of conventions and ambiguities [KBT⁺06].

There are many databases which integrate data from different sources. However, these databases are often limited to a few organisms or data domains, and a comprehensive view on integrated biological networks is therefore not (or only partly) possible. As a result specific analyses have to be done independently on basis of available data sets and common biological networks. In order to integrate biological networks from different sources and different domains, an identifier mapping has to be done to infer corresponding and related entities of different biological networks, and exploration methods have to be provided to support the investigation of the data.

We present methods and an easy to use prototype (based on the Vanted system [JKS06]) for the integration and visualization of biological networks via utilizing various data sources. The idea is to employ biological network data from different sources and from different domains such as metabolic pathways, protein-protein interaction networks, signal transduction pathways, and gene regulatory networks. The identifier mappings arise from an easily extensible set of integrated IDMapper's, which are managed by the identifier mapping framework BridgeDB [vIPK⁺10]. An IDMapper is a generic mapping information source which can be implemented in several shapes such as web services, SQL databases, or flat files. The manifold IDMapper implementation possibilities as well as their easy integration enable fast and adaptive extensions of the tool to further demands. The set of all identifier mappings constitute the identifier mapping graph. A powerful management of the identifier mapping graph including the handling of identifier synonyms and transitive identifier mapping paths afford a flexible integration of biological networks. The integrated biological networks can be visualized completely or partially according to various filtering operations. Via a targeted mapping and filtered visualization of integrated biological net-

works, the user is able to prepare custom systems biology analyzes and publication ready figures.

Literatur

- [JKS06] Bjorn Junker, Christian Klukas und Falk Schreiber. VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, 7(1):109+, 2006.
- [KBT⁺06] Jacob Köhler, Jan Baumbach, Jan Taubert, Michael Specht, Andre Skusa, Alexander Rüegg, Chris Rawlings, Paul Verrier und Stephan Philippi. Graph-based analysis and visualization of experimental results with ONDEX. *Bioinformatics*, 22(11):1383–1390, Juni 2006.
- [vIPK⁺10] Martijn van Iersel, Alexander Pico, Thomas Kelder, Jianjiong Gao, Isaac Ho, Kristina Hanspers, Bruce Conklin und Chris Evelo. The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinformatics*, 11(1):5+, 2010.