Chapter 3: Multi-level data analysis in cancer: tools and approaches

3.1 The Cytoscape platform for network analysis and visualization

Benno Schwikowski, Systems Biology Lab, Institut Pasteur and CNRS USR 3756, Paris, France

benno.schwikowski@pasteur.fr

Summary

Cytoscape is an open source software platform that supports the visualization and analysis of molecular profiling data in the context of functional interaction networks. It is developed by several research groups that are actively involved in the development of technologies around the generation and integrative analysis of molecular profiling data in the context of biological and biomedical research. Here, we outline the rationale behind the use of functional interaction networks, introduce the Cytoscape platform, and present an example in which data analysis and visualization using Cytoscape has led to a discovery of previously unknown disease biology.

Introduction

Biology is organized hierarchically, from lower levels that comprise DNA, RNA and proteins, over cells, to multicellular organisms, and ecosystems. Changes in lower levels lead to changes in higher levels, and evolutionary forces that act on higher levels shape, over longer time spans, lower levels. Many research questions concerning complex diseases, such as cancer, thus require data across different genes, and across several biological levels. Functional interaction networks offer a meaningful context for molecular measurements, and thus interpretability of the strongly increasing number of molecular profiling datasets that are becoming accessible, in particular for cancer (e.g. Tomczak *et al*, 2015).

Molecular pathways and functional interaction networks

Cancer is characterized by a distinct set of biological capabilities, for example resistance to cell death, or the ability to sustain proliferative signalling [2]. Most of these capabilities can be understood at the level of cellular and molecular physiology, and their acquisition is thought to arise through genetic instability.

To study how genetic instability leads to physiological change, it is natural to consider the intermediate level of molecular pathways that implement subcellular functions using systems of interacting molecules: DNA, RNA, proteins, and metabolites. Measuring, or even defining 'pathway activity' at a large scale is difficult. Based on global profiling technologies, such as next-generation sequencing, one can, for example, try to derive measures of pathway activity from integrated measures of abundance of the underlying molecules (e.g., Martignetti *et al*, 2016). Databases, such as KEGG [4] and Reactome [5] provide access to the molecules that constitute pathways, and their molecular interactions. However, our knowledge of pathways is still highly incomplete, and systematic approaches to determine unknown pathways do not exist. Therefore, global and exhaustive characterizations of biology at the pathway level, as it is now possible for DNA, RNA, and proteins, are currently out of reach.

An alternative mode of analysis at the level between molecules and cellular physiology is based on *functional interaction networks.* Two molecules are said to interaction functionally, whenever they potentially occur together in some functional context. Globally one can assemble these interactions into functional interaction networks that contain molecular pathways as connected subnetworks. Thus, to explore the possible bases of change in cellular physiology, one can look for aggregate changes in data about molecular subnetworks as reflections of change in molecular pathways.

Functional interaction networks can be based on various types of evidence. Genes and metabolites involved in sequential metabolic interactions provide one type of functional interaction. The resulting metabolic networks are particularly useful for integrated analysis, as they are relatively complete, and quantitative aspects of their function are relatively well understood. Popular databases of functional interactions [6,7] gather evidence for metabolic interactions, and a variety of other functional interactions, for example, physical protein-protein interactions from high-throughput, computational predictions, and evidence from automated text mining.

As of today, most types of functional interaction networks have to be considered highly incomplete. Still, placing molecular profiling data in the context of functional networks does offer practical opportunities to identify and study changes in molecular pathways as changes in — potentially fragmentary — subnetworks of functional interaction networks.

Approach and application example

A brief tour of Cytoscape

The Cytoscape software platform [8,9] offers a wide range of functionality around the visualization and analysis of functional and other networks and network-associated data, with no need for programming. Cytoscape has been developed as an open-source, community-driven software platform since its first release in 2002. Initiated at the Institute for Systems Biology in Seattle by the need to explore some of the first large-scale protein interaction datasets, Cytoscape is now a software platform whose core is maintained and developed by a handful of different systems biology research teams and extended a worldwide user community. Cytoscape is used across a wide range of applications, but its focus remains on biomedical research, and specifically, on molecular interaction networks.

Cytoscape consist of two parts: The core software, which provides a graphical user interface and a basic set of features for analysis and visualization, and apps, software with specific additional functions that can be added and removed as required. Cytoscape can hold one or several *Cytoscape networks*. Each Cytoscape network consists, on the one hand, of a *graph* that consists of a set of *nodes* and *edges* that connect pairs of nodes. Edges can be undirected or directed towards one of the nodes. On the other hand, edges and nodes have *attributes* in user-accessible *Node* and *Edge Tables*. Attributes are numerical (e.g., values for gene expression), categorical (e.g., from a limited number of strings that code different kinds of protein-protein interaction), or strings (e.g., URLs pointing to publications with in-depth information about genes and their interactions). Attributes can be interpreted by user-configurable apps, thus allowing a wide and flexible range of use cases.

Cytoscape can import data directly from local files and from the internet. Cytoscape and its apps provide access to a wide range of internet databases with networks, experimental data. File formats include human-readable and -editable formats for the network and associated data. Cytoscape can store and retrieve the current state of a session in *session files* that can be shared over the internet.

The Cytoscape user interface provides a number of panels for interacting with networks, node, edge, and network tables, and the set of all networks in memory. The main network panel graphically represents the nodes and edges of the currently selected network. The graph layout can be modified directly in the network panel or optimized by a large selection of graph layout algorithms. Data in node and edge tables can be represented using various visual node and edge attributes. How each type of data is mapped to visual attributes can be configured and controlled by the user by means of *styles*, combinations of parameterized visual properties whose parameters correspond to node- or edge-specific data in the node and edge tables. Styles can be saved independently of data, and, once defined, reused in subsequent Cytoscape sessions. More than 20 accessible node properties include fill color, shape, label, label color, and shape. The way in which node and edge attributes are mapped to visual properties, i.e., color gradients for underlying numerical data, can be configured by the user. These features allow for flexible, largely automated and uniquely customized and information-rich interactive visualization of larger networks, beyond biological applications.

To provide further flexibility and extensibility by the Cytoscape user community for the rapidly growing use cases, Cytoscape uses a modular software architecture that allows most Cytoscape functionality to be extended by *app*s, optional software components. While apps can be installed, upgraded, and uninstalled from within Cytoscape, most apps are accessed using the Web-based Cytoscape App Store (Lotia *et al*, 2013; http://apps.cytoscape.org). The App Store offers browsing apps by category, such as data import/export, visualization, data analysis, and automation. For each of the over 300 apps available today, the App Store provides an author-curated web page with a searchable description and release and download information. Apps can be installed directly into a running Cytoscape session upon mouse click. App pages also contain links to other app-specific information, such as journal articles and on-line discussions among users.

Cytoscape beginners can find links to manuals, tutorials, journal articles, and presentations on the main Cytoscape website (http://cytoscape.org). Beginning app developers find introductions to the Cytoscape architecture and APIs. Several searchable mailing lists provide exchange platforms for users and developers, and Cytoscape core developers participate to help with the hardest technical questions. Regular Cytoscape workshops and symposia bring user and developer communities together.

We now illustrate the analysis and visualization of subnetworks by a discovery of a new cell-physiological effect in a developmental disorder [11]. Briefly, the hallmarks of *Cavernous Cerebral Malformation* (CCM), can be recapitulated in murine models by inactivating homologs of the genes CCM1, CCM2, and CCM3, which carry the causative mutations for CCM [12,13]. Transcriptomic profiles had been obtained from the relevant venous tissue in which causative genes were invalidated, and from controls.

The Figure below shows a functional gene subnetwork, visualized in Cytoscape, and, on the right, images of tissue sections showing a physiological effect through labeling the protein corresponding to the center gene of the subnetwork.

Subnetwork nodes are colored by differential expression P-value, indicating change of the corresponding transcripts in CCM2-perturbed mice, relative to controls. The border color of each node indicates transcript fold change. The subnetwork, centered around the Van Willebrand Factor (VFW) gene, had been identified from the transcriptomic data by the LEAN method [11] to identify regions of strong aggregate differential expression P-value. The aggregates in the image on the right-hand side of the Figure shows a dysfunction of the VWF pathway in a cerebral tissue section of the CCM mouse model. The observed dysfunction may play a role in the pathophysiology of the disease. For details, we refer the reader to the original publication [11].

Discussion and perspectives

Towards comprehensive understanding of complex disease

Diseases, such as cancer, are rarely interpretable as the consequence of an abnormality in a single gene. Concepts and tools around functional interaction networks are still evolving, but it is clear that networks offer an important basis for interpreting complex disease, and treating it in the future [14]. For instance, functional interaction networks may be helpful in helping understand the otherwise often cryptic sets of similarities and differences between related diseases. The reason for this may simply be that the sets of genes that play important roles in related diseases strongly overlap (e.g., Smoller *et al*, 2013). Thus, understanding disease at the level of functional interaction networks may be essential to develop drug targets and biomarkers, and to reposition existing drugs.

As technologies for molecular profiling continue to proliferate, functional interaction networks for biology and medicine will become more complete, and thus more powerful. In parallel, better, and more comprehensive data at the molecular level will also lead to better models. Novel technologies to profile single cells at the molecular level are starting to provide sufficient measurement precision to deduce regulatory networks [16]. Functional interaction networks are thus poised to play an increasingly important role for the development of predictive and preventive approaches to medicine.

References

1. Tomczak, K., Czerwińska, P. & Wiznerowicz, M. The Cancer Genome Atlas (TCGA): An immeasurable source of knowledge. *Wspolczesna Onkol.* **1A,** A68–A77 (2015).
2. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144,** 646–674 (2011).
3. Martignetti, L., Calzone, L., Bonnet, E., Barillot, E. & Zinovyev, A. ROMA: Representation and quantification of module activity from target expression data. *Front. Genet.* **7,** 1–12 (2016).
4. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkv1070
5. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gkx1132
6. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47,** 1–7 (2018).
7. Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47,** D529–D541 (2019).
8. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13,** 2498–504 (2003).
9. Cline, M. S. *et al.* Integration of biological networks and gene expression data using cytoscape. *Nat. Protoc.* **2,** 2366–2382 (2007).
10. Lotia, S., Montojo, J., Dong, Y., Bader, G. D. & Pico, A. R. Cytoscape app store. *Bioinformatics* **29,** 1350–1351 (2013).
11. Gwinner, F. *et al.* Network-based analysis of omics data: the LEAN method. *Bioinformatics* **33,** 701–709 (2017).
12. Bergametti, F. *et al.* Mutations within the Programmed Cell Death 10 Gene Cause Cerebral Cavernous Malformations. *Am. J. Hum. Genet.* (2005). doi:10.1086/426952
13. Boulday, G. *et al.* Developmental timing of CCM2 loss influences cerebral cavernous malformations

in mice. *J. Exp. Med.* (2011). doi:10.1084/jem.20110571

14. Loscalzo, J., Barabási, A.-L. & Silverman, K. E. *Network medicine: Complex Systems in Human Disease and Therapeutics*. (2017).

15. Smoller, J. W. *et al.* Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381,** 1371–1379 (2013).

16. Rubin, A. J. *et al.* Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell* **176,** 361–376.e17 (2019).

Figure. Left: Cytoscape visualization of a functional subnetwork. Right: Physiological effect discovered by fluorescent labeling of the protein in the center of the subnetwork[11]