Network Biology SIG 2012
July 13th, Long Beach, CA

Invited Keynotes

Chris Evelo - Head of Department of Bioinformatics, BiGCaT at Maastricht University, The Netherlands
Hiroaki Kitano - Director of Systems Biology Institute in Tokyo; Professor, Okinawa Institute of Science and Technology Graduate University, Japan
Chris Sander - Chair of Computational Biology at Sloan-Kettering Institute, New York, New York, USA
Josh Stuart - Assoc. Professor of Biomedical Engineering at UC Santa Cruz, USA

Accepted abstracts

Anna Bauer-Mehren – *"Discovering latent 'modules of care' for comparative effectiveness research via network analysis of electronic health records"*
Stanford University, CA, USA

Mining of electronic health records (EHRs) has recently gained importance. However, most efforts are restricted to analyzing drugs, diseases and their associations. In biomedical research, network analysis has provided the conceptual framework to interpret protein-protein interactions or gene-disease association networks via large-scale network maps. We analyze associations between drugs, diseases, devices and procedures mined from EHRs using network analysis to extract hidden "modules of care" for hypothesis generation. In particular, we annotated the textual notes of the EHRs of one million patients in the Stanford Clinical Data Warehouse with disease, drug, procedure and device terms using ontologies such as SNOMED-CT or RxNorm. We then used standard co-occurrence statistics to establish associations between these clinical concepts and to construct networks. Hidden modules of care - clusters of diseases, drugs, procedures, devices – useful for hypothesis generation are extracted through network analysis approaches and visualized using Cytoscape.
We present a study for comparative effectiveness of Cilostazol vs. a control group in peripheral artery disease (PAD) patients (see Figure 1) and compare our results derived from the network analysis against standard methods such as regression analysis. We believe that network analysis allows us to uncover hidden ("latent") modules of care not detected through standard approaches, which do not account for the connectivity of the clinical events and entities.

Ugur Dogrusoz – *"Network Visualization and Analysis in the cBio Cancer Genomics Portal"*
Bilkent University, Turkey

The cBio Cancer Genomics Portal (http://cbioportal.org) is an open-access resource for interactively exploring multidimensional cancer genomics data sets. It provides simple and intuitive integrated access to cancer genomics data, including copy number, mutation, mRNA and microRNA expression, methylation and protein and phosphoprotein data, on more than 5,000 tumor samples from 20 cancer studies (including 16 TCGA cancer types). During the past year, we have added network visualization and analysis features to the cBio Portal. These new features enable researchers to analyze genomic alterations in the context of known biological pathways and interaction networks, and to more easily mine data generated by the TCGA. A network of interest is derived from the Pathway Commons project, based on the query genes specified by the user. Multidimensional genomic data are overlaid onto each node of the network, highlighting the frequency of somatic mutation and copy number alteration (and optionally mRNA up/down-regulation). Users can manage the

complexity of the network by filtering by total alteration frequency of genes or by type and source of the interactions. This provides an effective means of managing network complexity, while automatically highlighting those genes most directly relevant to the cancer type in question. In addition, drugs and drug target data can optionally be shown in relation to the network of interest. In this talk, we would like to illustrate the main network analysis features using data from the TCGA project. We will also discuss our future plans for the network view.

Janusz Dutkowski – *"From Networks to Ontologies of Gene Function"*

University of California, San Diego, USA

Ontologies are of key importance to many domains of biological research. The Gene Ontology (GO), in particular, has been instrumental in unifying knowledge about biological processes, cellular components, and molecular functions through a hierarchy of concepts and their interrelationships. However, given only partial biological knowledge and inconsistency in manual curation, unbiased construction, extension and validation of GO encounters significant challenges. Here we show that the existing collection of high-throughput network maps for Saccharomyces cerevisiae can be analyzed to automatically assemble an ontology of gene function that is comparable to manually curated efforts. Our systematic computational approach combines evidence from physical, genetic and transcriptional networks to produce an ontology comprised of 4,123 biological concepts and 6,872 hierarchical concept relations (Figure 1). Using a new ontology alignment procedure, we find that the network-based ontology captures the majority of known cellular components in budding yeast and identifies approximately 600 new cellular components and component relations. We will discuss the multiscale analysis performed by our framework including automatically identifying, annotating and visualizing the complete hierarchical structure of biological networks. We will also illustrate how it provides a powerful tool to uncover new biological knowledge and errors of manual curation. Finally, we will argue for a new role for ontologies in bioinformatics: rather than merely being used as a gold-standard for performing functional enrichment, ontologies should serve as evolvable models that are validated, revised, and expanded based on new genomic data.

Idit Kosti – *"A transcription-splicing integrated network reveals pervasive cross-regulation among regulatory proteins"*

Technion, Israel Institute of Technology, Israel

Traditionally the gene expression pathway was regarded as being composed of independent steps, from RNA transcription to protein translation. To-date there is increasing evidence for coupling between the different processes of the pathway, specifically between transcription and splicing. Given the extensive cross-talk between these processes, we derived a transcription-splicing integrated network. The nodes of the network included experimentally verified human proteins belonging to three groups of regulators: Transcription factors (TFs), splicing factors (SFs) and kinases. The nodes were wired by instances of predicted transcriptional and alternative splicing regulation. Analysis of the network indicated a pervasive cross-regulation among the nodes, specifically; SFs were significantly more often regulated by alternative splicing relative to the two other subgroups, while TFs were more extensively controlled by transcriptional regulation. In particular, we found a significant preference of specific pairs of TF-TF and SF-SF to regulate their target genes, SFs being the most regulated group via independent and combinatorial binding of SFs. Consistent with the extensive cross-regulation among the splicing and transcription factors, the subgroup of kinases within the network had the highest density of predicted phosphorylation sites. The prevalent regulation of the regulatory proteins was further supported by computational analysis of the protein sequences, demonstrating the propensity of these proteins to be highly disordered relative to other proteins in the human proteome. Overall, our systematic study reveals that an organizing principle in the logic of integrated networks favor the regulation of regulatory proteins by the specific regulation they conduct. Based on these results we propose a new regulatory paradigm, postulating that fine-tuned gene expression regulation of the master regulators in the cell is commonly achieved by cross-regulation.

Allan Kuchinsky & Anya Tsalenko – *"ENViz: a Cytoscape Plugin for Integrative Statistical Analysis and Visualization of Sample-Matched Data Sets with Multiple Data Types"*

Agilent Technologies, Santa Clara, CA, USA

Summary: ENViz performs enrichment analysis for pathways and gene ontology (GO) terms in matched datasets of multiple data types (e.g. gene expression and metabolites or miRNA), then visualizes results as a Cytoscape network that can be navigated to show data overlaid on pathways and GO DAGs.

Background: Modern genomic, metabolomics, and proteomic assays produce multiplexed measurements that characterize molecular composition and biological activity from complimentary angles. Integrative analysis of such measurements remains a challenge to life science and biomedical researchers. We present an enrichment network approach to jointly analyzing two types of sample matched datasets and systematic annotations, implemented as a plugin to the Cytoscape [1] network biology software platform.

Approach: ENViz analyses a primary dataset (e.g. gene expression) with respect to a 'pivot' dataset (e.g. miRNA expression, metabolomics or proteomics measurements) and primary data annotation (e.g. pathway or GO). For each pivot entity, we rank elements of the primary data based on the correlation to the pivot across all samples, and compute statistical enrichment of annotation sets in the top of this ranked list based on minimum hypergeometric statistics [2]. Significant results are represented as an enrichment network - a bipartite graph with nodes corresponding to pivot and annotation entities, and edges corresponding to pivot-annotation pairs with statistical enrichmentscores above the user defined threshold. Correlations of primary data and pivot data are visually overlaid on biological pathways for significant pivot-annotation pairs using the WikiPathways resource [3], and on gene ontology terms. Edges of the enrichment network may point to functionally relevant mechanisms. In [4], a significant association between miR-19a and the cell-cycle module was substantiated as an association to proliferation, validated using a high-throughput transfection assay. The figures below show a pathway enrichment network, with pathway nodes green and miRNAs gray (left), network view of the edge between Inflammatory Response Pathway and mir-337-5p (center), and GO enrichment network with red areas indicating high enrichment for immune response and metabolic processes (right).

Tijana Milenkovic – *"Biological networks reveal pathogen-interacting proteins"*

University of Notre Dame, Indiana, USA

Sequence-based computational approaches have revolutionized our understanding of biology. Since proteins aggregate to perform function instead of acting in isolation, the connectivity of a protein-protein interaction (PPI) network is expected to deepen biological insights over and above sequences of individual proteins. The advancement relies on developing sensitive graph-theoretic methods for extracting biological knowledge from PPI networks. By devising such a method that summarizes complex wirings around a node in the network and compares the topological similarity of extended network neighborhoods of two nodes, we showed that biological function of a protein and its network position are closely related. Since a network consists of nodes and edges, and since there is no reason to favor nodes over edges, we generalize the above measure to allow for computing topological similarity of extended network neighborhoods of two edges. Clustering of topologically similar edges according to this measure results in clusters of better quality compared to popular existing clustering methods. We apply our clustering strategy to the human PPI network to identify new pathogen-interacting proteins from the clusters, as these proteins represent candidates for therapeutic intervention.

Michael Smoot – *"Cytoscape 3.0: Architecture for Extension"*

University of California, San Diego, USA

Cytoscape is a popular, open source desktop application for visualizing and analyzing biological networks. Cytoscape 2.X consists of a core application that provides a visualization and analysis capabilities along with an API for extending Cytoscape's functionality through "plugins." Scientists and other Cytoscape users benefit from the analytical depth provided by the plugins, while plugin authors benefit from the core Cytoscape functionality and the framework for distributing and advertising plugins. This mutually beneficial relationship has resulted in over 150 plugins (http://cytoscape.org/plugins.html) along with dozens of publications about the plugins themselves.

Cytoscape 3.0 represents an attempt to refactor Cytoscape to make app writing (plugins will be renamed “apps” in 3.0) simpler while at the same time providing more stability, power, and flexibility to the system as a whole. First and foremost, Cytoscape 3.0 has been modularized with the API cleanly separated from the implementation. This modularity is being facilitated and enforced with OSGi (http://osgi.org), a popular Java modularization framework. OSGi’s micro service architecture allows private implementation code to remain private by registering micro services, which rely only on the public API. This means that any app only has the opportunity to depend on the public API, which will hopefully clarify and simplify what is needed to write an app. We have also begun using the Semantic Versioning standard (http://semver.org) for Cytoscape code to make clear how and when a public API may change. This will all go towards helping the Cytoscape core maintain backwards compatibility, which will greatly increase app stability. All “apps” in 3.0 can be written as OSGi bundles, just like the Cytoscape core modules. This means that apps will now have the opportunity to register their own public API, eliminating the distinction between core and app and resulting in a much more powerful and flexible system. While the architecture of Cytoscape 3.0 relies on OSGi, very little code does.
While not without risk, we believe that Cytoscape 3.0 will enable a new generation of Cytoscape apps as well as much greater opportunity for collaboration with different systems. This talk will elaborate on the new Cytoscape architecture including its benefits, challenges, and risks.

Chao Zhang – *“Network-Ontology Visualization and Analysis”*

University of Missouri, USA

With the improvement of high-throughput technology, the dramatic increase of large-scale data in both biomolecular concentration and biomolecular interactions has resulted in many biological networks, such as protein interaction networks, gene regulatory networks, and metabolic networks. Although functional analysis is the fundamental step of better understanding biological networks, utilizing vast wealth of data and huge amount of knowledge to annotate and analyze the function of biological networks is still challenging in nowadays bioinformatics. Many software tools are available to visualize and analyze function-derived biological networks, but most of them are isolated with simple functions. One challenge faced by these visualization tools is how to make sense of such networks often represented as massive “hairballs.” Many network analysis algorithms filter or partition networks based on topological features, or mathematically model networks rely on their statistical properties, sidestepping the issue of making sense of the network itself altogether. On other hand, traditional functional enrichment analysis methods regard a network as a list of genes, and annotate networks with gene set enrichment methods. However, it does not consider the topological dynamics of network which might lead to the different functions under different conditions. Therefore, it is necessary to consider molecular interactions to correctly and specifically annotate biological networks.
As one of the most successful open source frameworks in bioinformatics, Cytoscape is a powerful network visualization platform that actively supports independent plugin development. By integrating model-view-controller design pattern and Cytoscape techniques, it makes possible an integrated ontology-annotated biological network visualization and analysis platform. In the first stage of the project, we successfully developed two interactive plugins -- Mosaic (http://nrnb.org/tools/mosaic) and NOA (http://nrnb.org/tools/noa) -- to address both visualization and analysis respectively. Mosaic supports interactive network annotation and visualization that includes partitioning, layout and coloring based on biologically-relevant ontologies. It shows slices of a given network in the visual language of biological pathways, which are familiar to any biologist and are ideal frameworks for integrating knowledge, and also provides researchers with an interactive tool to evaluate biological interactions within the context of well-defined processes, functions and cellular localization while retaining all original network information. NOA first introduced link ontology that assigns functions to interactions based on the known annotations of joint genes via optimizing two novel indexes ‘Coverage’ and ‘Diversity’. Then, NOA generates two alternative reference sets to statistically rank the enriched functional terms for a given biological network. It has been proved to be more efficient not only in dynamic transcription regulatory networks but also in rewiring protein interaction networks.
Mosaic and NOA prove that function-driven visualization, biological network analysis and advanced software design pattern can be integrated organically. With introducing more

analysis modules and advanced visualization algorithms into this platform step by step, it will become a complete solution for network-ontology visualization and analysis.

Panelists

Tanwir Habib – Engineer Research and Development Center (ERDC), USA
Paul Pavlidis – University of British Columbia, Canada
Erik Sonnhammer – Stockholm Bioinformatics Centre, Sweden

Posters

Marcio Acencio – *"Discovery of potential conditions to gene morbidity in yeast by a machine learning approach based on network centralties"*

Sreedevi Chandrasekaran – *"A Network View on Parkinson's Disease Related Genes"*

Tanwir Habib – *"Linking physiological measurements to microarray data through network inference to understand endocrine-disrupting effects"*

Michael Heuer – *"Venn and Euler diagrams for Cytoscape"*

Yeongjun Jang – *"MONGKIE: Modular Network Generation and Visualization Platform with Knowledge Integration Environment"*

Atanas Kamburov – *"ConsensusPathDB: assembling a more complete picture of cell biology"*

Frank Kramer – *"Integration of Pathway Knowledge into Methods for Network Reconstruction"*

Jun Li – *"iPPServer: an Interactive Gene Expression Pattern and Pathway Analysis Server for Gene Regulatory Network Discovery"*

Augustin Luna – *"Using Molecular Interaction Maps as Formal Representations for Pathway Interaction Diagrams: The Notation and Software Tools"*

Shahin Mohammadi – *"Sweet-talking yeast: systematic identification of TOR downstream effectors under dietary restriction"*

Yulia Newton – *"Computational Discovery of Master Regulators in Basal vs. Luminal Breast Cancer"*

Paul Pavlidis – *"Visualizing specificity in coexpression networks"*

Deanna Petrochilos – *"Using Random Walks to Identify Cancer-Associated Modules in Expression Data"*

Oksana Riba-Grognuz – *"Quality Assessment Using Network Representation of Genome Assemblies"*

Zhiao Shi – *"NetGestalt: Data integration over hierarchically and modularly organized networks"*

Lin Song – *"Should mutual information, correlation, or model based co-expression measures be used for defining network modules?"*

Erik Sonnhammer – *"Comparative interactomics with Funcoup 2.0"*

Jing Zhu – *"Deciphering genomic alterations in colorectal cancer through subtype-specific driver networks"*