Network Biology SIG Abstracts — Friday, July 10th

Invited Keynotes

Ron Shamir – Tel Aviv University – *Do networks help in making sense of the omics data deluge?*

Henning Hermjakob – EBI – Open community standards in proteomics and systems biology

Florian Markowetz – University of Cambridge – Functional analysis of genetic interaction networks

Brenda Andrews – University of Toronto – Global Mapping of Genetic Networks in Yeast

Accepted Talks

Barry Demchak – Cytoscape Cyberinfrastructure: leveraging microservices to the cloud and beyond
UCSD, USA

Cytoscape is an indispensable tool for network data analysis and visualization. One of Cytoscape's greatest strengths is that it is powered by a vibrant array of developer-contributed apps. However, as network biologists' requirements evolve, Cytoscape is challenged not only to keep pace, but to lead new and existing developers to create even greater value. Currently, multiscale and multifaceted networks push the memory limits of a Cytoscape workstation, while complex calculations such as Network Based Stratification and Network Based GWAS strain workstation processors. Increasingly, users demand support for collaborative projects, reproducible workflows, and interoperability with external tool chains. Finally, economic pressures favor solutions that promote code and algorithm reusability and evolvability.

In response, we have created the Cytoscape Cyberinfrastructure (CI), which is both an Internet-scale distributed system (based on Microservices [1]) and the network biology community it serves. Its mission is to enable and encourage network biologists to create and deploy high quality, innovative and scalable services focusing on network-based computation, collaboration and visualization.

Microservices can be written in any language, and are highly testable and evolvable. They can run on servers ranging from a single thread to a large cloud-based cluster. They can easily be reused in reproducible workflows or can serve as components in larger services. The CI links microservices via a light weight REST-based aspect-oriented interchange protocol (called CX), which enables tailored data streams while supporting service innovation via evolvable standards. CI infrastructure services support user authentication, long duration job execution, and a service repository that enables researchers to publish their services or discover services published by others. This model builds on the successful Cytoscape app community, which is based on similar mechanisms though at the scale of individual workstations.

Prominent examples of microservices include NDEx [2] (a repository for biological networks), NodeWalker (which uses heat dispersion to identify the most relevant subnetworks containing a given set of genes), cyNetShare [3] (which visualizes a network in a browser) and Cytoscape itself (which can also call CI services). Interfaces are available for Python, IPython, R and Matlab. Future work includes adding clustering, analysis, layout, publishing and display microservices and interfaces to Galaxy and Taverna workflows.

Already though, our researchers are leveraging existing microservices in multiple contexts, thus demonstrating the increased flexibility and productivity available in the CI.

[1] http://martinfowler.com/articles/microservices.html
[2] http://home.ndexbio.org/about-ndex-2/
[3] http://idekerlab.github.io/cy-net-share/

Fazle Faisal – Exploring the structure and function of temporal networks with dynamic graphlets: implications for aging
University of Notre Dame, USA

Since susceptibility to diseases increases with age, studying human aging is important. However, doing so experimentally is hard due to long lifespan and ethical constraints. Therefore, human aging-related knowledge needs to be inferred computationally. In this context, computational analyses of genomic sequence or gene expression data have been indispensable for investigating human aging. Since proteins carry out cellular processes by interacting with each other, computational analyses of protein-protein interaction (PPI) networks could further our understanding of the processes of aging. Most of current methods for analyzing systems-level PPI networks deal with their static representations, due to limitations of biotechnologies for PPI data collection, even though cellular functioning is dynamic. Here, we aim to complement the existing static network efforts to study human aging via dynamic network analysis. Namely, we integrate current static PPI network data with aging-related gene expression data to computationally construct dynamic, age-specific PPI networks. Then, we aim to study cellular changes with age from such networks. We hypothesize that such dynamic network analysis framework provides a valuable model of cellular functioning that can complement the existing aging-related knowledge. As a result, we propose novel computational approaches for efficient dynamic network analysis in the context of human aging.

Inna Kuperstein – Data visualization and modeling using Atlas of Cancer Signaling Network predicts clinical outcome
Institut Curie, U900 INSERM, France

Studying reciprocal regulations between cancer-related pathways is essential for understanding signaling rewiring during cancer evolution and in response to treatments. With this aim we have constructed the Atlas of Cancer Signaling Network (ACSN, http://acsn.curie.fr), a resource of cancer signaling maps and tools with interactive web-based environment for navigation, curation and data visualization (http://navicell.curie.fr). The content of ACSN is represented as a seamless ‘geographic-like’ map browsable using the Google Maps engine and semantic zooming. The associated blog provides a forum for commenting and curating the ACSN maps content. The atlas contains multiple crosstalk and regulatory circuits between molecular processes implicated in cancer. The integrated NaviCell web-based tool box allows to import and visualize heterogeneous omics data on top of the ACSN maps and to perform functional analysis of the maps. NaviCell web-based tool box is also suitable for computing aggregated values for sample groups and protein families and mapping this data onto the maps. The tool contains standard heatmaps, barplots and glyphs as well as the novel map staining technique for grasping large-scale trends in numerical values projected onto a pathway map. The combination of these flexible features that provides an opportunity to adjust the modes of visualization to the data type and achieve the most meaningful picture. The NaviCell web service provides a server mode, which allows automating visualization tasks and retrieve data from maps via RESTfull (standard HTTP) calls. There is also a possibility of bindings to several programming languages as Python, R, Java.

To demonstrate the application of ACSN and NaviCell we show a study on drug sensitivity prediction using the networks. A comprehensive map of cell cycle and DNA repair

signaling network, from the ACSN collection, has been used for deciphering genetic interactions of drug resistance. We looked for synthetically interacting combinations of genes on the comprehensive map of cell cycle and DNA repair signaling network by: (1) deriving a state transition graph from the map and retrieving all paths from a damaged DNA to the repaired DNA state; (2) applying an algorithm for searching the minimal cut sets to model the effect of invalidating a set of genes involved in regulating DNA repair on the overall efficiency of DNA repair machinery; (3) considering genes that regulate each state transition as potential target for interference. Using this approach we have retrieved all possible gene sets whose knock-out halts DNA repair. Integrating expression and mutation data from drug resistant breast cancer cell lines and patients resistant to genotoxic treatment allowed prioritizing synthetically lethal gene sets and predicting specific intervention set for restoring sensitivity to drugs in each individual drug resistant cell line or patient.

In additional study we show how epithelial to mesenchymal transition (EMT) signaling network from the ACSN collection has been used for finding metastasis inducers in colon cancer through network analysis. We performed structural analysis of EMT signaling network that allowed highlighting the network organization principles and complexity reduction up to core regulatory routs. Using the reduced network we modeled single and double mutants for achieving the metastasis phenotype. We predicted that a combination of p53 knock-out and overexpression of Notch would induce metastasis and suggested the molecular mechanism. This prediction lead to generation of colon cancer mice model with metastases in distant organs. We confirmed in invasive human colon cancer samples the modulation of Notch and p53 gene expression in similar manner as in the mice model, supporting a synergy between these genes to permit metastasis induction in colon.

Juliane Perner – Functional and co-evolutionary analysis of chromatin and cytosine modification networks in mouse embryonic stem cells
Max Planck Institute for Molecular Genetics, Germany

The cell-type specific regulatory state of the genome can be observed by investigating the local chromatin environment. Specific combinations of histone modifications and cytosine modifications reveal the current chromatin state. With these combinations individual regulatory elements, e.g. promoters or enhancers, and their regulatory states can be identified. However, the functional impact of the different combinations of chromatin modifications is not yet clear. Especially their implications on the recruitment of chromatin modifiers to specific chromatin environments and their relationships to transcription factors are mostly unknown.

We have integrated publicly available epigenomic data of mouse embryonic stem cells (mESC) combining 139 experiments (ChIP-Seq and MeDIP) of 77 epigenomic features to investigate the interactions between chromatin modifications and chromatin modifiers. We applied network reconstruction methods to the genome-wide location data to derive chromatin state-specific regulatory programs. The resulting interactions are then compared across the various chromatin states to reveal potential commonalities in the different regulatory programs. Using this novel approach we identified modules of specific interactions characterizing particular chromatin states. We show that, although some interaction modules are shared among chromatin states, others are very context-specific. Our comparative analysis results in data-driven, novel hypotheses on the regulatory mechanisms defining the various chromatin states. Most importantly, we propose a novel role of the recently discovered hydroxymethylcytosine (5hmC) and 5-formylcytosine (5fC) in the chromatin signaling network.

Maria Sorokina – A novel metabolic network representation for the discovery of conserved modules of chemical transformations
CEA, IG, Genoscope, France

The identification of functional modules in metabolic networks allows the improvement of functional annotation and the discovery of new metabolic pathways and enzymatic reactions. In this context, we propose a novel representation of a global metabolic network where reactions sharing a same chemical transformation type are grouped in reaction molecular signatures (RMS) (Carbonell, Carlsson, & Faulon, 2013, http://molsig.sourceforge.net). RMS have the advantage of being an automatic and expert-independent reaction classification that is much finer and has a wider coverage than the Enzyme Commission classification.

Starting from a directed reaction network, reaction vertices sharing the same RMS are grouped in single vertices and edges are established from the initial reaction connectivity. Several scores are then computed for each path in the RMS network in order to evaluate known metabolic pathway conservation and to discover new ones. The first score, scoreRea, is computed using the average number of reactions and represents the chemical conservation of the paths in the whole metabolism. The second one, scoreProt, is based on the number of proteins associated to each RMS and represents the enzymatic conservation among the tree of life. The next one, scoreTopo, is based on the PageRank centrality and depicts the topological importance of the RMS sequence in the metabolic network. The last score, scoreMeta, is the number of different reaction paths among known metabolic pathways grouped in a RMS path and represents the chemical transformation conservation across the known part of the metabolism. The most conserved RMS paths are then identified and used to understand the linkage between the path conservation types (chemical, enzymatic and topological) and the type of metabolic pathway (biosynthesis, degradation, detoxification, etc.).

We show that our representation of metabolism has an interesting predictive potential and can be used to identify most conserved parts of the metabolism and to find new metabolic modules. Furthermore, the combination of different scores can be used to predict the metabolic role of new pathways using supervised machine learning. Associated to genomic context data like operons and syntenies, conserved paths of chemical transformations will be a useful tool for functional annotation of genes and groups of genes of unknown function.

Georg Summer – Neo4j as a useful addition to the network biology toolkit
Maastricht University, The Netherlands

We are working on different heart failure projects with varying requirements. Neo4j serves as a storage backend for the many different networks we work on. To integrate the domain information with other biological knowledge data sources we developed the Network Library, a tool set to merge these data sources. For visualization and analysis purposes we use cyNeo4j, a Cytoscape app to connect to Neo4j. We present these tools and our experience with Neo4j while investigating heart failure related projects.

Junwen Wang – Prediction of determining factors and underlying gene regulatory network for cell fate conversion
University of Hong Kong

Conversion of somatic cells into another cell type by forced expression of defined factors is a powerful approach in the researches for regenerative medicine. Yamanaka, the 2012 Nobel Prize winner, showed in 2006 that the overexpression of four transcription factors, Oct4, Sox2, Klf4 and Myc, could convert mature, differentiated cells into pluripotent stem cells that can proliferate and self-renew indefinitely and give rise to all types of cells in the body. Patient-specific induced pluripotent stem cells have many potential uses to cure degenerative diseases, such as diabetes and neurodegeneration. Direct conversions between differentiated cells are also succeeded in many cases these years. However, searching the factors for cell fate conversion needs tedious experimental efforts and the efficiency is usually very low. Recent advancement of various high-throughput omics technologies promotes the generation of big data in biology and medicine. Systematic

analysis on these genome-wide omics data makes it feasible to predict the determining factors for cell fate conversion. Shmulevich and colleagues used expression rank difference to identify transcription factor pairs that may control cell lineages. But the top cell-specific reversal transcription factor pairs may not be the driven factors for cell conversion since how these factors interact or regulate other genes are still unknown. For example, in mouse embryonic stem cell, only two Yamanaka factors ranks in the top 20 transcription factors identified by this method. Network comparison between engineered cells and their in vivo counterparts shows that cell conversion may fail to silence or activate proper gene expression programs even though the targeted cell phenotype and marker expression are shown. Thus, both the factors and their target networks should be taken into consideration to achieve better prediction performance. In this study, we designed a novel bioinformatics method to improve the prediction of determining factors for cell fate conversion by inferring the factors and their networks simultaneously via a sparse-and-group-sparse optimization algorithm. Integration of transcriptome and transcription factor binding data, as well as other omics data, to construct gene regulatory network has shown a higher accuracy than using transcriptome data alone in our previous studies. To access the robustness of our method, we verified it by the simulated data and applied it to cell types with known factors for successful cell fate conversions. Our method has shown a better performance than previous methods. To predict the determining factors for more cell types that are valuable in medicine, such as regenerative cell therapies, we collected multiple omics data of over 100 cell types or tissues in human and mouse from public databases, and applied our method to predict the determining factors and their networks in each cell type or tissue. Results from this study will benefit the researches in regenerative medicine by enhancing the successful rate and reducing the experimental cost of cell fate conversion.

Accepted Posters

#	Authors	Poster Title
1	Erich Baker, Charles Phillips, Jason Bubier, Michael A. Langston and Elissa Chesler	Efficient hierarchical biological network comparison in GeneWeaver using a bipartite edge list store
2	Francois-Michel Boisvert	AP-MS and PP-MS as complementary approaches for the interactome mapping of protein complexes.
3	Glyn Bradley	CausalR – Extracting mechanistic sense from genome scale data
4	Guillaume Brysbaert, Jérôme de Ruyck, Mélany Tanchon, Ralf Blossey, Marc Aumercier and Marc Lensink	Identification of novel targets for the inhibition of Ets-1 related cancer progression
5	Keqiang Li, Gang Su, Guanming Wu, Kevin Davis, Barbara Mirel and Fan Meng	CoolMap-Reactome Integration for Hypothesis Development with Sample Level Data Using Curated Pathway Diagrams
6	Laura O'hara, Alessandra Livigni, Sz-Hau Chen, Mark Barnett, Tim Angus, Derek	A graphical and computational system for modelling the dynamics of biological

	Wright, Lee Smith and Tom Freeeman	pathways
7	Luca Erculeiani, Francesca Galante, Caterina Gallo, Francesco Asnicar, Luca Masera, Paolo Morettin, Nadir Sella, Thomas Tolio, Giulia Malacarne, Kristof Engelen, Andrea Argentini, Valter Cavecchia, Claudio Moser and Enrico Blanzieri	Discovering Candidates for Gene Network Expansion by Variable Subsetting and Ranking Aggregation
8	Malte D Luecken, Matthew Page, Andrea Crosby, Sean Mason, Gesine Reinert and Charlotte M Deane	A Network-based Functional Validation Method for Protein Sets
9	Martin H Schaefer and Luis Serrano	Cellular environment shapes tissue specificity of cancer genes
10	Martina Kutmon, Susan L Coort and Chris T Evelo	Integrative network-based analysis of mRNA and microRNA expression in vitamin D3-treated cancer cells
11	Michaela Bayerlová, Frank Kramer, Annalen Bleckmann and Tim Beissbarth	Integrated network analysis for identification of functional modules in cancer
12	Minsuk Kim, Jeong Sang Yi and Byung-Gee Kim	Designing antibiotic overproducing strain of Streptomyces coelicolor using the model of genome-scale metabolic network
13	Mohammad Sadeh and Martin Vingron	Estimating Causal Effect Strength between Histone Modifications and Modifiers in Chromatin Signaling Networks
14	Omer Nebil Yaveroglu, Tijana Milenkovic and Natasa Przulj	Proper Evaluation of Alignment-free Network Comparison Methods
15	Pavel Sumazin	Genomic alterations dysregulate cancer genes by modulating microRNA activity
16	Scooter Morris, Daniel Keedy, Henry van den Bedem and James Fraser	contactApp: Using Cytoscape to explore CONTACT Networks
17	Sergey Nepomnyachiy, Nir Ben-Tal and Rachel Kolodny	CyToServer: A developer tool to bridge between network and 3D-structure
18	Shérazade Braham, Christine Brun and Andreas Zanzoni	Inference of the protein interaction network between Fusobacterium nucleatum putative secretome and the human host
19	Stefan Avey and Steven Kleinstein	Network-Regularization Improves Prediction of Influenza Vaccination Response
20	Teal Guidici, George Michailidis, Amy	Detecting differentially expressed

	Rothberg and Charles Burant	metabolic pathways with adjustments for macronutrient intake
21	Theresa Schacht, Alexandra Poos, Marcus Oswald and Rainer Koenig	Employing Mixed Integer Linear Programming to elucidate gene regulation
22	Vipin Vijayan, Vikram Saraph and Tijana Milenkovic	MAGNA++: Maximizing Accuracy in Global Network Alignment via both node and edge conservation
23	Wiktor Jurkowski	Network analysis guided integration of omics data to unravel healthy prostate response in broccoli intervention study