Network Biology SIG Abstracts — Friday, July 11th

Invited Keynotes

Manolis Kellis – MIT – *Regulatory and network clustering of genetic variants associated with complex traits*

Marc Vidal – Harvard University – *Interactome networks and human disease*

Marian Walhout – UMass Medical School – *Interspecies systems biology: nutritional regulatory networks*

Accepted Talks

David Amar – *Pathways as robust biomarkers for cancer classification: the power of big expression data*

Tel-Aviv University, Israel

Background: Gene expression signatures, serving as biomarkers, have been used successfully for prognosis, diagnosis and patient stratification in cancer. However, such signatures are often not sufficiently robust and sometimes perform poorly on new datasets. Moreover, standard case-control studies may yield a signature that is not specific to the tested disease. In addition, the biomedical interpretation of gene set signatures is often difficult.

Methods: We collected data from 174 gene expression studies in GEO, covering 13,314 samples from 17 different array technologies covering 48 diseases. Each sample was manually annotated with Disease Ontology terms. We analyzed each sample separately by calculating two pathway features: the mean and standard deviation of the pathway genes' weighted rank. We used 1700 pathways from the Reactome, NCI, KEGG and Biocarta databases. We call the new database Peptalk (Pathway Expression, Phenotype and Tissue - A Learning Knowledgebase).

Results: Using Peptalk, we constructed a classifier for each disease by comparing the pathway features of the disease samples to those of all other samples. We used 48 diseases that have at least 5 datasets and 100 samples each. Our analysis produced high performance classifiers for 16 disease terms, which include cancer, gastric cancers, breast cancer, and immune system cancers. For these diseases, classification was accurate even when validated on new datasets of different technologies. Our cancer classifier is based on a signature that of cell cycle and DNA replication pathways. Notably, the classifier for gastrointestinal cancer is based on several pathways not associated previously with the disease, including down-regulation of aquaporin-mediated transport. We summarize our results in a network in which nodes diseases and pathways. Edges connect a pathway to a disease if the pathway is markedly differential in the disease patients compared to both healthy controls from the same study and to samples of other diseases. Hence edges indicate a pathway-disease coupling that is specific to the tested disease. The network provides an overview of associations between pathways and diseases, and is valuable for further biological interpretation.

Conclusions: Our analysis shows how large public databases can be used to infer disease specific diagnostic tools and can reveal pathways that have not been recognized before as pertinent to a specific type of cancer.

Hyunghoon Cho – *A Bayesian model for identifying context-dependent community structure across multiple networks*

MIT, Boston, USA

The dynamic and context-dependent nature of gene-gene interactions has led to the collection of a large number of biological networks based on multiple conditions (cases versus controls, or different classes of individuals), time points (disease progression), and cell types. An open challenge is to develop methods to identify transient or network-specific gene modules that can shed light into differential patterns of module activity across time points or conditions. Current methods focus mostly on identification of gene modules within individual networks, making it difficult to identify patterns across multiple networks. We have developed a hierarchical Bayesian model named multi-network mixed-membership stochastic blockmodel (Multi-MMSB) that takes as input a collection of networks, and identifies gene modules that may exist in only a subset of networks. We applied Multi-MMSB to a group of co-expression networks constructed from microarray profiles of asthma patients at three different stages: quiet, exacerbation, and two weeks after exacerbation. We discover a gene module associated with innate immune response and interferon signaling pathway that is only active during exacerbation, and another module associated with extracellular matrix disassembly that activates during exacerbation and remains active even after two weeks, suggesting the existence of a long-term molecular effect of an exacerbation event (p-value < 10E-4). In addition, we performed simulation experiments that showed Multi-MMSB to be superior at recovering gene modules compared to naive approaches that combine the networks into a single representative network before learning the modules. These results support Multi-MMSB as a tool for systematically identifying differences in community structure across multiple networks.

Traver Hart – *A human coessentiality network predicts gene function and novel cancer subtyping based on shared genetic vulnerability*

University of Toronto, Canada

Among the discoveries arising from the systematic assay of genetic interactions is that genes with correlated genetic interaction profiles tend to share biological function. To date, only yeast genetic interaction networks are large enough to derive these relationships. Human genetic interactions, on the other hand, have thus far been assayed on a relatively small scale, using either highly biased subsets of genes (Bassik et al, 2013; Laufer et al, 2013; Roguev et al, 2013) or a small number of query genes vs. a whole-genome set of array genes (Vizeacoumar et al, 2013). The resulting networks, though individually informative, are too small for correlation analyses.

Large-scale RNAi screens in human cell lines offer an alternative approach to deriving functional networks. We demonstrate that genes with correlated essentiality profiles across a panel of over 200 cell lines are highly enriched for shared biological function, providing a platform for predicting gene function as well as protein complex membership when integrated with large-scale protein co-elution profiles. Moreover, as the cell lines assayed by RNAi are generally derived from human cancers, we show that clustering profiles by shared genetic vulnerability identifies existing tumor subtypes and discovers novel metabolic relationships among cell lines derived from tumors with different tissues of origin.

Deriving the coessentiality network depends critically on a novel Bayesian method for analyzing large-scale RNAi negative selection screens. The method calculates a posterior log odds of essentiality for each gene and, combined with experimentally-derived, gold-standard reference sets of core essential and nonessential genes, represents the state of the art in analyzing genome-scale perturbation screens (and works equally well for CRISPR screens). Applying the method across two compendia of shRNA screens (Cheung et al, 2011; Marcotte et al, 2012) reveals the context-dependent essential genes in each screen as well as a surprising and counterintuitive relationship between gene expression and RNAi sensitivity for constitutively essential genes.

Salvatore Loguercio – *Network-Augmented Genomic Analysis (NAGA) applied to Cystic Fibrosis studies*

Scripps Research Institute, CA, USA

We present here a network-based method to integrate functional genomics data (e.g. siRNA screens) with interactomics datasets (e.g. AP-MS, MudPIT), useful for prioritizing novel functional targets and for identifying relevant network modules. It leverages publicly available information on protein-protein interactions and thus is readily applicable to many scenarios where a connection between functional and biochemical data is sought.

Application to CFTR Cystic fibrosis (CF) is an early onset disease characterized by a defect in the apical chloride channel, CF transmembrane conductance regulator (CFTR). The most common disease causing mutation is a 3 base pair deletion resulting in loss of Phe 508 (F508del), which leads to misfolding, endoplasmic reticulum (ER) retention and efficient ER associated degradation of the protein.
To elucidate the molecular networks influencing the folding and function of CFTR in CF, we recently screened CFBE41o- cells containing F508del-CFTR against a siRNA library of 2500 targets known to be involved in protein homeostasis. In parallel, we generated a high confidence CFTR interactome of F508del-CFTR in the same cell line.
Given a list of high-scoring siRNA hits for CFTR rescue of function, and a set of CFTR binding proteins, we sought to connect these datasets through an integrated protein-protein interaction network, and use shortest path analysis to uncover the minimal network structure consistent with both the CFTR interactome and siRNA data. The goal of this approach is to prioritize proteins connecting CFTR with siRNA hits that may act as central “hubs” in cellular processes required for CFTR functional rescue. For each protein in the subgraph, it computes the number of distinct siRNA hits that utilize the protein on its shortest path to CFTR. In order to filter out nonspecific protein hubs, this computation is repeated using a random selection of hits from the original siRNA library. The analysis identified several novel candidates for CFTR rescue of function that could be validated through targeted siRNA screens.

Tijana Milenkovic – *Novel Directions for Biological Network Alignment: MAGNA (Maximizing Accuracy in Global Network Alignment)*
University of Notre Dame, Indiana, USA
Analogous to genomic sequence alignment, biological network alignment aims to identify similar regions between networks of different species. Then, biological network alignment can be used to transfer functional knowledge across species between conserved (i.e., aligned) network regions. However, several issues exist with the current view of the problem of biological network alignment, and we aim to improve on these issues by developing two conceptually novel and superior solutions to the problem. Namely, we introduce computational frameworks for: 1) fair evaluation and comparison of existing state-of-the-art network alignment methods, which is currently lacking, and in the process, we propose a novel and improved network alignment method; and 2) maximizing accuracy in global network alignment via our another new method called MAGNA. Then, because human aging is hard to study experimentally due to long lifespan and ethical constraints, we use these two new network alignment frameworks to transfer aging-related knowledge from well-annotated model species to poorly annotated human between aligned network regions. By doing so, we produce novel aging-related information, which complements currently available information about aging that has been obtained mainly by sequence alignment, especially in human. To our knowledge, we are the first to use network alignment to learn more about aging.

Ashwini Patil – *TimeXNet: Identifying active gene sub-networks using time-course gene expression profiles*
Institute of Medical Science, University of Tokyo, Japan
Time-course gene expression profiles are frequently used to study cellular response to stimulus and to infer molecular pathways involved in cellular response. We introduce a method, TimeXNet (http://timexnet.hgc.jp/), which identifies active gene sub-networks with temporal paths using time-course gene expression profiles in the context of a weighted gene regulatory and protein-protein interaction network. TimeXNet uses a specialized form of the network flow optimization approach1 to identify the most probable paths connecting the genes with significant changes in expression at consecutive time intervals.
TimeXNet was used to study the innate immune response using time course gene expression profiles of activated immune cells2. The innate immune response is the first level of protection in organisms against invading pathogens. It is primarily mediated by the Toll-like receptors functioning through the Myd88-dependent and TRIF-dependent pathways. We classified the immune response into three consecutive time-dependent stages – early, intermediate and late - and used TimeXNet to identify the most probable paths in the molecular network between genes expressed in the early and the late phases of the immune response, while taking into account those expressed in the intervening time. The resultant network contained several new and known regulators of the innate immune response, as well as those

transiently expressed between sampled time points. The predicted temporal network suggested a role for the protein phosphatase 2a catalytic subunit α in the regulation of the immunoproteasome during the late phase of the response. An analysis of time course gene expression profiles from Myd88-knockout and TRIF-knockout dendritic cells helped clarify the differences between the Myd88-dependent and TRIF-dependent pathways in the innate immune response.
TimeXNet was extensively evaluated for its ability to predict novel regulators and their associated pathways within gene sub-networks. Compared to other similar methods, TimeXNet identified up to 40% more novel regulators from independent experimental datasets. It also predicted paths within a greater number of associated KEGG pathways with longer overlaps (up to 7 consecutive edges) within these pathways. Thus, TimeXNet is a reliable tool that can be used to study cellular response to stimuli through the identification of time-dependent active gene sub-networks in diverse biological systems.

Gerald Quon – *Context-specific regulatory networks identify key regulators of complex traits*

MIT, Boston, USA

Genome-wide association studies (GWAS) have identified thousands of single nucleotide variants associated with diverse human traits, but understanding their combined action in complex systems remains an open challenge. With more than 80% of lead GWAS SNPs located in non-coding regions of the genome rich in regulatory elements, functionally characterizing these variants necessitates knowledge of (1) the locations of cell type specific enhancers; (2) the identity of the target genes of those enhancers; and (3) the interactions between these target genes to identify disrupted pathways and subnetworks.
Using enhancer and promoter maps for 112 cell types constructed by the Roadmap Epigenomics Consortium, we have constructed directed context-specific and cell type specific networks, where nodes represent both genes and non-coding regulatory elements, and edges lead from transcription factors to regulatory regions (enhancers, promoters). These meta-networks, where nodes consist of both genes and non-coding regulatory elements, enable context-specific network analysis that can yield insight into the role of different cell types in complex traits, and to our knowledge have not been previously explored. To leverage these networks for GWAS analysis, we developed an efficient probabilistic model to map GWAS variants to candidate disrupted regulatory elements, and use each context-specific network to (1) identify trait-associated genes whose regulation is disrupted by non-coding variants; (2) identify master TF regulators of the trait-associated genes; and (3) identify other genes (not proximal to GWAS variants) involved in the trait.
We predicted modules of non-coding variants associated with brain, cardiovascular, lipid, and immune-mediated disorders, as well as their regulators. Predicted regulatory modules and transcription factors involved in HDL and LDL cholesterol levels replicated across multiple studies and are most highly expressed in liver cell types. Furthermore, gene knockouts reported in the MGI database lead to abnormalities including perturbed circulating lipid levels and susceptibility to atherosclerosis. Modules and regulators predicted for multiple sclerosis and Crohn's disease are most highly expressed in CD4+,CD8+, and CD34+ cells, and their knockouts lead to defects in B-cell and NK-cell morphology and circulating levels.
We expect our networks and probabilistic model will be a valuable resource for the community to study other complex traits and diseases.

Yu Xia – *Signatures of Pleiotropy, Economy and Convergent Evolution in a Domain-Resolved Map of Human-Virus Protein-Protein Interaction Networks*

McGill University, Canada

We construct the first global, domain-resolved map of human-virus and within-human interactome networks. Our high-resolution network biology approach enables the discovery of new mechanistic principles of host-pathogen interactions otherwise hidden in the binary network. Our analysis reveals, for the first time, global differences between viral proteins and human proteins in terms of interaction mechanisms: viruses surmount genome size constraints by convergently evolving multiple short linear motifs to effectively mimic, hijack, and manipulate complex host processes for their survival.

Accepted Posters

#	Presenter	Poster Title
1	George Acquaah-Mensah	Insights into the Alzheimer hippocampus: transcriptional regulatory networks revealing a wealth of relationships
2	Omer Basha	myGeneNet, TissueNet and ResponseNet: tools to create and explore up-to-date molecular interaction networks across tissues and organisms
3	Barry Demchak	Cytoscape Untangles the Web
4	Soheil Feizi	Spectral network algorithms reveal conserved human, fly and worm regulatory pathways
5	Somaye Hashemifar	Hubalign: an accurate and efficient method for global alignment of protein-protein interaction networks
6	Somaye Hashemifar	Analysis of tissue-specific co-expression networks
7	Michael Heuer	Visualizing consequences of genetic variation in biological networks
8	Frank Kramer	Integrating Prior Pathway Knowledge into Methods for Network Reconstruction
9	Inna Kuperstein	Visualization and analysis of data using Atlas of Cancer Signalling Networks (ACSN) and NaviCell tools for integrative systems biology of cancer
10	John Scooter Morris	clusterMaker2's "Fuzzifier": a Simple Approach to Fuzzy Clustering for Biological Networks
11	Anna Ritz	Pathway Analysis with Signaling Hypergraphs
12	Martina Summer-Kutmon	Using pathway and network analysis to investigate the mechanisms in the diabetic liver
13	Jingchun Sun	Exploring Signaling Pathway Networks to Understand Mode of Drug Action: Metformin as a Case
14	Koichi Takahashi	KEGGscape: A Cytoscape App for Mapping Metadata on KEGG Pathways
15	Alfonso Valencia	Stability of Cancer and Alzheimer's Interaction Networks, a Simulated Annealing based Approach
16	Giorgio Valentini	On the Automated Function Prediction of Big Multi-Species Networks
17	Ruisheng Wang	Network-based association of hypoxia-responsive genes with cardiovascular diseases
18	Gunaming Wu	The Reactome FI Cytoscape plugin: a tool for modeling alterations in pathway activities in experimental data sets
19	Lei Xie	Quantitative Impact of Dynamic Rewiring of Biological Network on Information Dissemination
20	Bingqing Xie	Disease gene prioritization using network and feature information