Network Biology Software: Integration of *-omics* data to discover functional sub-network markers of human disease

John L.H. Schenkel, Jr.[1] , Rod K. Nibbe[1,2], Mark R. Chance[1,2]
[1]NeoProteomics Inc., 11000 Cedar Ave ,Cleveland, Ohio, [2]Center for Proteomics and Bioinformatics, Case Western Reserve University, Cleveland, Ohio

Overview

•Identification of high-value sub-networks and/or gene combinations with a functional role in polygenic human diseases.

•Drive *in silico* and *in vivo* validation.

Background

Many complex human diseases (e.g. cancer, diabetes, schizophrenia etc.) have correspondingly complex, polygenic genotypes that initiate and sustain disease progression. Despite significant progress over the past few decades identifying genes critical to mediating phenotype, our understanding of the functional basis of molecular phenotype for complex diseases is insufficient. Signaling pathways that consist of a few proteins interacting in a serial fashion oversimplify, and provide inadequate models for, the behavior mediated by multiple interacting gene products. Partly revealed by rigorous studies of increasingly well-annotated protein-protein interaction (PPI) networks, it has become clear that many of the proteins in these canonical signaling pathways engage in "crosstalk" with, and are modulated by, an ontologically diverse set of additional proteins, where this crosstalk is frequently mediated in a tissue and/or disease specific manner.

Researchers, both in the academic and commercial realms, lack analytical tools to integrate this information in the context of their research. There exists a need for software tools to search and score networks of interactions in a *quantitative, disease-specific* fashion, with flexibility to "seed" the search using a set of candidate disease targets or with high throughput, unbiased *–omics* data. **The goal is to provide a rank order for the targets/interactions in the sub graph to identify the highest value targets,** thereby productively directing the research enterprise, both in its pre-clinical phases as well as in clinical research and clinical trial management, through enhanced bioinformatics tools for correlative studies.

NeoProteomics is currently developing an integrated suite of software tools for the academic and commercial research community to fulfill the unmet demand for quantitative network analysis that can drive practical translational research and validation.

Conclusions

We are developing a robust suite of software that integrate a variety of –omics datasets in the context of the PPI for predicting candidate sub-networks with a functional role in disease. The sub-networks present an ideal context in which to propose perturbation experiments to validate the *in silico* findings. **NeoProteomics has a range of biomarker discovery capabilities to offer on a contract basis and has a number of biomarker candidates in diabetes, metabolic diseases and cancer available for commercialization (see www.neoproteomics.net).**

Acknowledgements

The authors would like to acknowledge Mehmet Koyuturk, Salim Chowdhury, Rob Ewing, Sanford Markowitz for considerable help in various phases of the implementation and associated publications. This work is supported in part by the Translational Technologies Resources Core of the Center for Translation Science UL1-RR-024989 at Case Western Reserve University and by NeoProteomics, Inc.

References

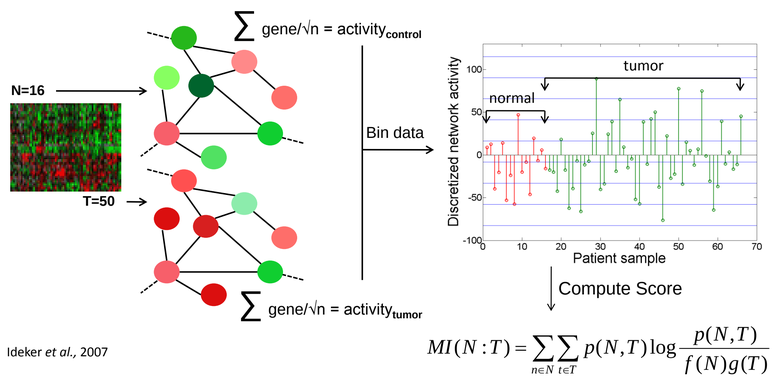
1. Nibbe RK, Markowitz S, Myeroff L, Ewing R, Chance MR. Discovery and scoring of protein interaction subnetworks discriminative of late stage human colon cancer. *Molecular & Cellular Proteomics* **2009**, 8:827–845. PMCID: PMC2667362
2. Nibbe RK, Koyuturk M, Chance MR. An integrative -omics approach to identify functional sub-networks human colorectal cancer. *PLoS Computional Biology* **2010**, 6:e1000639.
3. Chowdhury SA, Nibbe RK, Chance MR, Koyuturk M. Subnetwork state functions define dysregulated subnetworks in cancer, *Journal of Computational Biology*, *epub*, **2010**.

Featured Algorithms

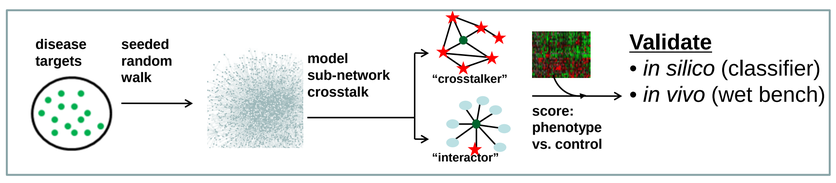
Overview

With these algorithms a researcher may 1) using genome-wide expression data, score sub-networks that one has reason to believe are significant for a particular disease phenotype (SASSy), or 2) perform a seed-guided search of a PPI for candidate sub-networks with a functional role in a particular polygenic disease (Crosstalker), or 3) using genome-wide expression data, perform an unbiased ("seedless") search of a PPI for functional sub-networks in a disease based on sub-network state functions (CRANE).

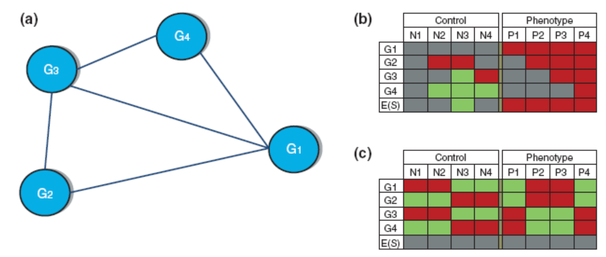
SASSy: **S**ub-network **A**nalysis and **S**coring **Sy**stem. **Synopsis**: SASSy provides functionality to score (i.e. rank-order) any sub-network of interacting proteins based on gene expression data. **Inputs**: Sub-network(s), microarray. **Outputs**: Top scoring *n* proteins in each sub-network, p-values.



Crosstalker: **Synopsis**: Uses a "seed" of disease targets (found by GWAS, proteomic profiling, etc.) and a model of network crosstalk to search any PPI for candidate sub-networks with a functional role in disease. Candidate sub-networks may be scored with SASSy. **Inputs**: Global; PPI database (e.g. HPRD), disease gene seed. **Outputs**: Sub-networks classified by significance.



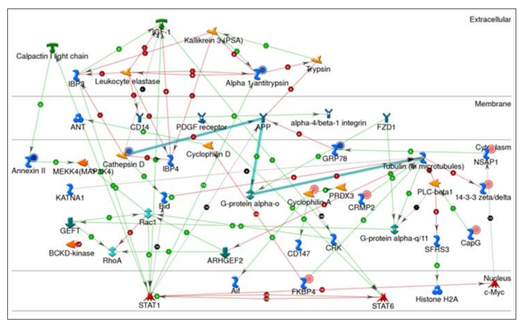
CRANE: **C**ombinatorially dys**R**egul**A**ted sub**NE**tworks. **Synopsis**: CRANE searches a global PPI for small sub-networks that maximize the sub-network state function between test & control., using binarized mRNA expression data (microarray) as a proxy for the state. CRANE sub-networks are optimal as features to train a classifier compared to sub-networks identified by comparable strategies. **Inputs**: Global PPI, microarray w/test & control. **Outputs**: Top scoring *n* proteins in each sub-network, p-values.



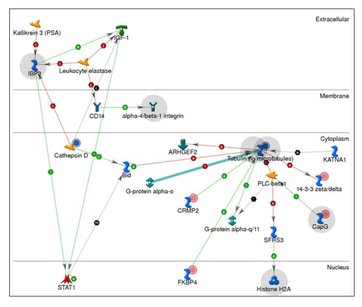
Select Implementations

SASSy[1]

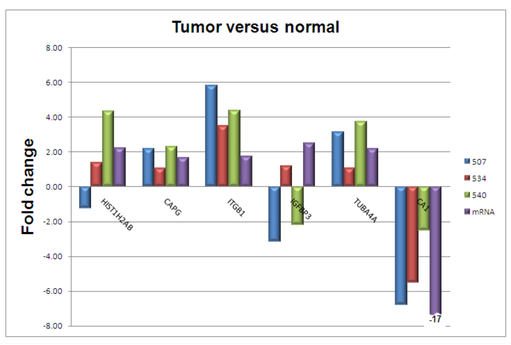
Unscored, candidate colon cancer sub-network (MetaCore).



Scored by SASSy, pruned.

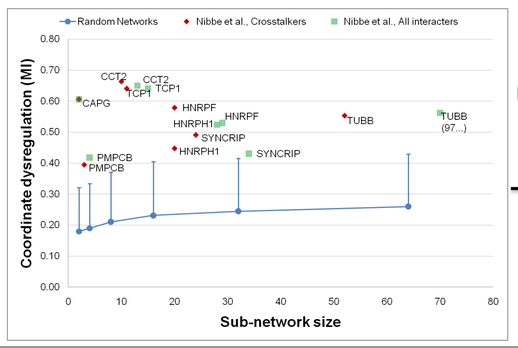


Dysregulation validated in independent patient cohort.

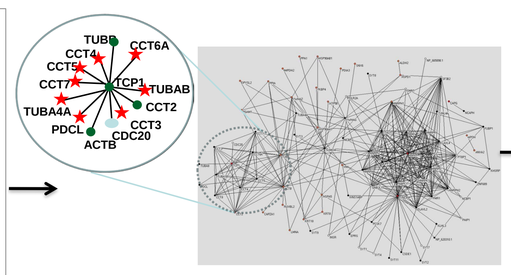


Cross-talker[2]

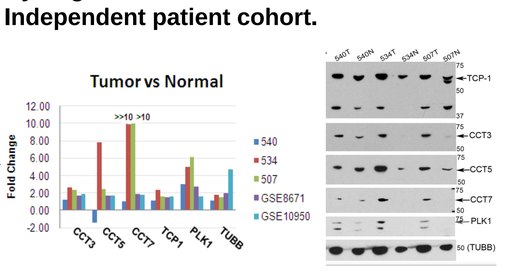
Candidate dysregulated sub-networks in colon cancer.



High-scoring crosstalker.

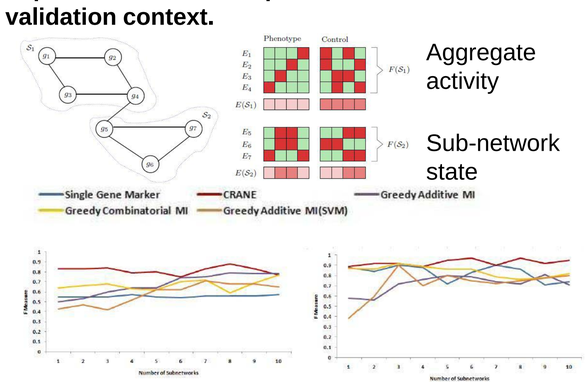


Dysregulation validated in Independent patient cohort.



CRANE[3]

State functions capture what aggregate methods miss. CRANE sub-networks show superior classification performance in a cross-validation context.



Hypothesis-driven sub-network significant for colon cancer metastasis.