HyperModules: identifying clinically and phenotypically significant network modules with disease mutations for biomarker discovery

Alvin Leung, Gary D. Bader* and Jüri Reimand*

The Donnelly Centre, University of Toronto, 160 College Street, M5S 3E1 Toronto, Ontario, Canada

ABSTRACT

Summary: Correlating disease mutations with clinical and phenotypic information such as drug response or patient survival is an important goal of personalized cancer genomics and a first step in biomarker discovery. HyperModules is a network search algorithm that finds frequently mutated gene modules with significant clinical or phenotypic signatures from biomolecular interaction networks.

Availability and implementation: HyperModules is available in Cytoscape App Store and as a command line tool at www.baderlab.org/software/HyperModules.

Contact: Juri.Reimand@utoronto.ca or Gary.Bader@utoronto.ca

Received on December 20, 2013; revised on February 25, 2014; accepted on March 25, 2014

1 INTRODUCTION

Establishing functional links between genetic variation and human disease is a key goal of cancer genome sequencing (Gonzalez-Perez et al., 2013) and genome-wide association studies (Hardy and Singleton, 2009). Complex diseases like cancer are often driven by infrequent changes in multiple genes in pathways (Vogelstein et al., 2013). Network analysis helps interpret mutations in systems context and find disease genes, pathways and biomarkers for precision medicine (Barabasi et al., 2011).

Discovery of modules (subnetworks) in biological networks helps isolate systems with disease-related properties and reduces interactome complexity. A growing number of methods are available for this purpose. A landmark paper combines gene expression signatures with protein–protein interactions (PPI) to find predictive modules of cancer outcome (Chuang et al., 2007). The NETBAG method studies genetic associations and copy number variants to find autism-related modules (Gilman et al., 2011). HotNet detects frequently mutated pathways in networks (Vandin et al., 2011, 2012). Net-Cox builds prognostic cancer signatures in network analysis of gene expression data (Zhang et al., 2013). The Reactome FI Cytoscape plugin uncovers prognostic gene modules from networks and gene expression data (Wu and Stein, 2012). Network-based stratification predicts tumor subtypes from mutations in network regions (Hofree et al., 2013). Such modules maximize a feature of genes such as differential expression, disease mutation frequency or enrichment of interactions.

Because clinical profiles of patients are increasingly available in cancer genomics efforts such as the The Cancer Genome Atlas (TCGA) pan-cancer project (Weinstein et al., 2013), new methods are needed to discover multivariate biomarkers in networks. We recently analyzed cancer mutations in phosphorylation signaling and found that kinase–substrate networks are informative of patient survival and therapy response (Reimand and Bader, 2013; Reimand et al., 2013). In particular, we found network modules with rare mutations in ovarian cancer patients with improved prognosis. We created the HyperModules method to systematically discover clinically correlated modules from gene and protein networks (Reimand and Bader, 2013) based on our earlier work on functional subnetwork discovery (Altmae et al., 2012; Reimand et al., 2008). Here we present the previously unavailable software in open-source Java as a command line tool for automated work and a Cytoscape app for interactive graphical analysis.

2 SOFTWARE

HyperModules assumes that clinically informative mutations of complex disease occur in systems of closely interacting genes. The greedy network search algorithm focuses on a local network area, defined by a central seed node (a mutated gene) and its surrounding subnetwork. All mutated genes are sequentially considered as seeds in module discovery. Search starts from the seed and grows the module toward increased benefit by adding connected genes that best improve clinical significance. This objective is driven by statistical tests where patients defined by the module are compared with other patients. Categorical clinical variables are studied with Fisher’s exact test and survival times with log-rank test. Cox regression is currently not supported; however, we plan to add this feature in the future. To establish statistical significance of detected modules, we build a null distribution by searching networks with permuted gene names. Each module of the true network is quantified with an empirical P-value reflecting the fraction of seed-specific modules from shuffled networks exceeding the significance of the true module. This removes artifacts of the greedy strategy and corrects for topological features such as highly connected nodes.

The analysis pipeline is outlined in Figure 1. Interaction networks are loaded into Cytoscape using standard features (Shannon et al., 2003). HyperModules requires gene mutations and patient clinical information in two tables. The user selects

*To whom correspondence should be addressed.
HyperModules is a biological network-mining algorithm that reveals modules of interacting genes with clinically informative disease mutations. Diverse biomolecular interaction networks can be analyzed, including PPI networks, gene regulatory networks and curated biological pathways. Disease mutations are also broadly defined. Although we initially studied cancer point mutations, other types of alterations such as copy number and gene expression changes can be used. HyperModules finds correlations with groups of genes where mutations may be infrequent but the signature strengthens through network integration. Such modules are not often directly usable as biomarkers because of small sample size; however, we believe that our approach helps discover genes and pathways as potential multivariate biomarkers for further experiments.

ACKNOWLEDGEMENTS

The authors thank Jason Montojo and Harold Rodriguez for Cytoscape help.

Funding: NRNB (US NIH NCRR grant P41 GM103504) and Google (Google Summer of Code 2013).

Conflict of Interest: none declared.

REFERENCES

Altmae,S. et al. (2012) Research resource: interactome of human embryo implantation: identification of gene expression pathways, regulation, and integrated regulatory networks. Mol. Endocrinol., 26, 203–217.

Barabasi,A.L. et al. (2011) Network medicine: a network-based approach to human disease. Nat. Rev. Genet., 12, 56–68.

Cancer Genome Atlas Research Network. (2008) Integrated genomic analyses of ovarian carcinoma. Nature, 474, 609–615.

Chuang,H.Y. et al. (2007) Network-based classification of breast cancer metastasis. Mol. Syst. Biol., 3, 140.

Gilman,S.R. et al. (2011) Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron, 70, 898–907.

Gonzalez-Perez,A. et al. (2013) Computational approaches to identify functional genetic variants in cancer genomes. Nat. Methods, 10, 723–729.

Hardy,J. and Singleton,A. (2009) Genomewide association studies and human disease. N. Engl. J. Med., 360, 1759–1768.

Hofree,M. et al. (2013) Network-based stratification of tumor mutations. Nat. Methods, 10, 1108–1115.

Reimand,J. and Bader,G.D. (2013) Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers. Mol. Syst. Biol., 9, 637.

Reimand,J. et al. (2008) GraphWeb: mining heterogeneous biological networks for gene modules with functional significance. Nucleic Acids Res., 36, W452–W459.

Reimand,J. et al. (2013) The mutational landscape of phosphorylation signaling in cancer. Sci. Rep., 3, 2651.
Shannon, P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13, 2498–2504.

Turner, B. et al. (2010) iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database (Oxford)*, 2010, baq023.

Vandin, F. et al. (2011) Algorithms for detecting significantly mutated pathways in cancer. *J. Comput. Biol.*, 18, 507–522.

Vandin, F. et al. (2012) Discovery of mutated subnetworks associated with clinical data in cancer. *Pac. Symp. Biocomput.*, 1, 55–66.

Vogelstein, B. et al. (2013) Cancer genome landscapes. *Science*, 339, 1546–1558.

Weinstein, J. N. et al. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, 45, 1113–1120.

Wu, G. and Stein, L. (2012) A network module-based method for identifying cancer prognostic signatures. *Genome Biol.*, 13, R112.

Zhang, W. et al. (2013) Network-based survival analysis reveals subnetwork signatures for predicting outcomes of ovarian cancer treatment. *PLoS Comput. Biol.*, 9, e1002975.