Gabi: Network inference from antibody-based proteomics data

Alexander L. R. Lubbock[1], Michael W. Irvin[1], David J. Harrison[2], Carlos F. Lopez[1], Ian M. Overton[3]

***Short Abstract* — Gabi is a novel algorithm for inference of small-scale networks from human tumor tissue samples scored for protein expression using quantitative antibody-based technologies. These signed, directed networks provide insight into pleiotropy, complexity and context-specificity. Inferred networks successfully recover the information flow between proteins on synthetic data generated by the PySB simulation framework. Directionality predictions have high precision (79%) if input network connectivity is accurate. The Gabi algorithm was applied to study multiple carcinomas (renal, breast, ovarian), providing novel insights into the relationships between epithelial–mesenchymal transition (EMT) players and fundamental processes dysregulated in cancers e.g. apoptosis and proliferation.**

***Keywords* — network biology, systems medicine**

I. Introduction

THE biological network has become a key concept within systems medicine. Network analysis shows particular promise for cancer biology, where the underlying causes are complex and heterogeneous. Dysregulation of networks/pathways is key to oncogenesis and clonal selection within tumors [1]. Indeed, changes in network structural properties can be predictive of patient outcome [2].

Functional proteomics platforms, including tissue microarrays (TMAs) and reverse-phase protein arrays (RPPAs) are particularly relevant for understanding cancer signaling activity, wherein protein abundance and post-translational modifications are key determinants. The antibody-based TMA and RPPA platforms both enable study of *ex vivo* tissue from carcinomas of interest, providing insight into the context specific and pleiotropic activity of proteins.

Here we present Gabi: a bespoke method to infer biological networks from functional proteomics data typically containing 6–100 protein markers. Key improvements over existing methods include detection of a broad array of coexpression patterns by combining Spearman correlation and symmetric uncertainty (normalized mutual information), relevance thresholding using an automated parametric approach, and directionality inference based on conditional independence detection, including graph theoretic evidence weighting based on the maximum clique algorithm.

[1]Department of Cancer Biology, Vanderbilt University, Nashville, TN
[2]School of Medicine, University of St Andrews, UK
[3]MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, UK

II. Results

A. Incorporation of functional scaffolds

During Gabi calibration on data from the DREAM4 network inference competition [3], it was noted that directionality precision improved from 56% to 79% if the correct network connectivity was supplied. Thus, we added the ability to include high confidence prior knowledge edges [4] into the procedure.

B. Benchmarking

A tool was developed for extracting gold standard information flow networks from models in the PySB framework [5]. The Apoptosis Necrosis Reaction Model (ANRMv2.0) was used. Gabi achieved similar directionality precision to methods pcalg [6] and ggm [7] but with twice and four times the directionality recall respectively.

C. Carcinoma networks

Gabi networks were generated from renal, breast and ovarian cancer datasets from Western General Hospital, Edinburgh, UK and The Cancer Genome Atlas (TCGA) project [8]. Existing knowledge is recapitulated e.g. clusters of epithelial adhesion markers and mesenchymal/invasion markers. Novel insights include connections between tumor weight and key proteins in the TCGA breast cancer network, and elucidation of ER-β2's role in high grade serous ovarian cancer.

III. Conclusion

Gabi is a network inference algorithm for small-scale proteomics data providing key improvements in connectivity and directionality inference. The algorithm was calibrated and benchmarked using separate synthetic modeling approaches. Performance on synthetic data from ANRM exceeded rival methods overall. Known cancer biology and novel insights are observed in Gabi networks generated on cancer proteomics datasets.

References

[1] Vogelstein B, Kinzler KW (2004) Cancer genes and the pathways they control. *Nat Med* **10** 789-799
[2] Taylor IW et al. (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotech* **27** 199-204
[3] Marbach D et al. (2010) Revealing strengths and weaknesses of methods for gene network inference. *PNAS* **107** 6286-6291
[4] Linghu et al. (2009) Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network. *Genome Biol* **10** R91
[5] Lopez et al. (2013) Programming biological models in Python using PySB. *Mol Syst Biol* **9** 646
[6] Kalisch et al. (2012) Causal Inference Using Graphical Models with the R Package pcalg. *J Stat Soft* **47** 1-26
[7] Marchetti GM, Drton M, Sadeghi K (2014) ggm: A package for Graphical Markov Models. http://CRAN.R-project.org/package=ggm
[8] The Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. *Nature* **490** 61-70