Hearing the Shape of Cancer: Spectral Graph Theory for Pathway Analysis of Gene Expression Data

Rosemary Braun[1]

***Short Abstract* — Gene expression profiling is a ubiquitous tool in the study of disease, yielding detailed measurements of the expression of thousands of genes for each sample. Because complex phenotypes, such as cancers, do not have single-gene etiologies, there is growing interest in techniques that provide systems–level insights by analyzing expression data in the context of biological pathways. This talk presents a novel graph–theoretic approach to pathway analysis. Because the method does not rely upon single-gene association statistics, it is able to detect pathways with aberrent signalling and transcription dynamics that would be missed in traditional pathway enrichment tests. Importantly, spectral analysis of the pathway graph enables inferences about a pathway's dynamical properties using static gene expression measurements. Details of the method and its application to cancer gene expression data set will be presented.**

***Keywords* — spectral analysis, gene expression, pathways and signalling networks**

I. Background and Significance

GENE-expression profiling experiments provide detailed measurements of a sample's transcriptomic state by simultaneously assaying 10^5–10^6 probes. Typically, each sample is assayed once, yielding a "snapshot" of its gene expression. By analyzing this static data in the context of known networks of biological interactions (pathways), it is possible infer systems that exhibit different transcription and signalling dynamics in different phenotypes [1].

In the typical analysis of gene expression data, each marker is tested independently for association with the phenotype of interest, followed by enrichment analyses (such as GSEA [2]) that detect pathways in which differentially expressed genes are over-represented. Recently, pathway analyses incorporating the topology of molecular interaction networks have been proposed, allowing gene–level statistics to be considered in context of the genes' position in the network [1].

However, the reliance of these analyses on gene–level tests of differential expression limits their ability to reveal functionally relevant pathways. Specifically, enrichment analyses will fail to detect pathways in which aberrant transcriptional signalling leads to a change in co-expression patterns across multiple genes but in which the individual genes do not exhibit independent (i.e., marginal) differental expression. Instead, methods that summarize the gene co-expressions across the pathway and then test for phenotype–conditional differences are required [3,4].

[1]Biostatistics Division, Department of Preventive Medicine and Robert H. Lurie Comprehensive Cancer Center, Northwestern University

II. Results

We present a novel graph-theoretic method to summarize the bulk characteristics of pathways and reveal systems–level differences without relying on single–marker association statistics. Using putative network topologies from curated pathway databases [5], we weight the edges by gene pair co-expression measurements from the experimental data. The weighted adjacency matrix is then used to compute the graph Laplacian [6], encapsulating both the pathway topology and the expression data.

Just as the spectrum of the Laplacian of a physical system (e.g., the shape of a drumhead) may be used to infer its dynamical properties (its sound), spectral decomposition of the graph Laplacian characterizes the network dynamics. Here, we use the spectrum to summarize the bulk dynamical properties of the pathway network in the context of gene co-expression data. Comparisons of the spectra obtained from different phenotypes provides insight into pathways with altered dynamical characteristics. (Statistical significance is assessed by permutation tests.) In addition, projection of gene expression values onto the eigenvectors of the Laplacian provides a means to characterize how alterations in the expression of specific genes impact pathway dynamics.

III. Conclusions

We have developed a novel method designed to reveal systems–level differences in biological pathways without relying on significant gene–level expression differences. By considering the gene expression data in the context of the spectral characteristics of pathway graphs, the method provides a means to extrapolate from static gene expression measurements to the dynamical properties of biological signalling networks. This talk will present the method in detail and describe its application to cancer gene expression data.

References

[1] Khatri, P., Sirota, M., and Butte, A. Ten years of pathway analysis: Current approaches and outstanding challenges. *PLoS Computational Biology*, 8(2):e1002375, 2012.

[2] Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS*, 102(43):15545–50, 2005.

[3] Braun, R., Leibon, G., Pauls, S., and Rockmore, D. Partition decoupling for multi-gene analysis of gene expression profiling data. *BMC Bioinformatics*, 12(497), 2011.

[4] Braun, R., Cope, L., and Parmigiani, G. Identifying differential correlation in gene/pathway combinations. *BMC Bioinformatics*, 9:488, 2008.

[5] Schaefer, C. F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., and Buetow, K. H. PID: the Pathway Interaction Database. *Nucleic Acids Res.*, 37:D674–679, 2009.

[6] Chung, F. *Spectral graph theory*. Amer Mathematical Society, 1997.