

Integrating multiple genomic data types into structured models of gene regulatory networks

Daniel E. Zak¹, Rajanikanth Vadigepalli², James S. Schwaber²,
Gregory E. Gonye², & Babatunde A. Ogunnaike¹ *

¹Department of Chemical Engineering, University of Delaware, Newark, DE 19716

²Department of Pathology, Cell Biology and Anatomy, Thomas Jefferson University, Philadelphia, PA 19107

ABSTRACT

Cells are complex dynamical systems that remodel themselves over time in response to changes in their internal and external environments. The emergence of technology for acquiring genome-wide gene expression and other genomic data sets has created an opportunity to construct computational models of gene regulatory networks, with the objective of going beyond observation of changes in gene expression and quantifying causal links between gene products that underlie physiological processes. Computational models of these causal links can enable global understanding of the regulation that is fundamental to all life and generate testable hypotheses about novel drug targets for prevention and treatment of complex diseases.

A growing literature describes approaches for modeling gene regulatory networks that use gene expression data alone. This may be a fundamentally intractable problem, as it requires the determination of thousands of dynamic relationships between genes and the estimation of potentially millions of parameters from only a limited amount of relatively poor quality data. To render gene regulatory network modeling tractable, we take an alternative, structured, approach that uses multiple genomic data types and employs the assumption that gene expression is largely regulated through variation in the complement of active nuclear transcription factors (TFs).

We consider three levels of structure in our modeling approach: (1) subcellular structure that only allows active TFs to regulate gene expression (2) nuclear connectivity that specifies which TFs regulate which genes, and (3) dynamical model structure that describes how changes in the activity of specific TFs changes the transcription rates of target genes. Determining (1) requires knowing which genes are TFs and thus depends on gene annotation and literature information. Determining (2) requires knowing which transcriptional regulatory elements (TREs) are present in the promoters of which genes, and which TREs are bound by which TFs, and thus depends on genomic sequences, databases of TREs, bioinformatic tools, and literature information. Dynamic modeling in (3) allows inclusion of prior knowledge of mRNA degradation rates which dynamically

constrains the network possibilities. While our structured modeling approach trades-off dependence on gene expression data alone for dependence on additional data types, it is not problematic given the current rate of progress in genomic sequencing, annotation, and bioinformatic tool development. There is no reason to exclude these data types from any attempt to model gene regulation.

We have applied our approach to data from the yeast cell cycle [1, 6, 3] and observed how the activity of some, but not all, TFs may be regulated effectively at the level of gene expression [5]. This result indicates where more complex models are necessary to describe the experimental observations, and could not have been obtained if we did not take a structured approach. With the help of the bioinformatic tool PAINT [2], we are currently applying our approach to data for the response of neuronal cell lines to neuromodulators, and are obtaining results that both confirm previous results from the literature and generate new, testable hypotheses [4]. We observed that individual TFs may function both as transcriptional activators and transcriptional repressors for different sets of genes during a single response. This result contrasts what we observed for yeast cell cycle, where the TFs appeared to be dedicated activators or repressors, and demonstrates how a structured approach is even more essential for modeling mammalian gene regulatory networks.

Acknowledgments

We thank Praveen Chakravarthula for expert technical support with PAINT. We thank NIH/NHLBI, NIH/NIAAA IRPG, NIH/NIGMS BISTI, and DARPA BioCOMP for funding. DEZ additionally thanks the University of Delaware Department of Chemical Engineering for funding.

REFERENCES

- [1] R. J. Cho, M. J. Campbell, E. A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T. G. Wolfsberg, A. E. Gabrielian, D. Landsman, D. J. Lockhart, and R. W. Davis. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell*, 2(1):65–75, 1998.
- [2] R. Vadigepalli, P. Chakravarthula, D. E. Zak, J. S. Schwaber, and G. E. Gonye. PAINT: A promoter analysis and interaction network generation tool for genetic regulatory network identification. *Omic*s, In press.

* Author to whom correspondence should be addressed: Babatunde A. Ogunnaike, ogunnaik@che.udel.edu, (302) 831-4504, FAX: (302) 831-1048.

- [3] Y. Wang, C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown. Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. USA*, 99(9):5860–5865, 2002.
- [4] D. E. Zak, H. Liu, R. Vadigepalli, B. A. Ogunnaike, J. S. Schwaber, and G. E. Gonye. Combining structural and functional genomic data to infer transcriptional networks underlying neuromodulation in a model system. *Manuscript in preparation*, 2003.
- [5] D. E. Zak, R. Vadigepalli, G. E. Gonye, F. J. Doyle III, J. S. Schwaber, and B. A. Ogunnaike. Unconventional systems analysis problems in molecular biology: a case study in gene regulatory network modeling. *Submitted*, 2003.
- [6] J. Zhu and M. Q. Zhang. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics*, 15(7–8):607–11, 1999.