

Analysis of Signaling Pathways in Human T-Cells using Bayesian Network Modeling of Single Cell Data

Karen Sachs¹, Omar D. Perez^{2,3,4}, Dana Pe'er⁸, Garry P. Nolan^{2,3,4}, David K. Gifford⁵, Tommi S. Jaakkola⁵ and Douglas A. Lauffenburger^{1,6,7}

1. Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, MA 02139.
 2. Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305-5175.
 3. The Baxter Laboratory of Genetic Pharmacology, Stanford University School of Medicine, Stanford, CA 94305-5175.
 4. Department of Molecular Pharmacology, Stanford University School of Medicine, Stanford, CA 94305-5175.
 5. MIT Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139.
 6. Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139.
 7. Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.
 8. Department of Genetics, Harvard Medical School, Boston, MA 02115.
- Correspondence should be addressed to K. Sachs. Email: karens@mit.edu

Abstract

We perform network inference ('reverse-engineering') on phospho-specific multi-dimensional flow cytometry measurements of signaling molecules in human T cells using Bayesian networks. Inferred networks are found to have good agreement with known pathways derived from the literature.

We applied Bayesian networks, a probabilistic modeling tool, to flow cytometry measurements of signaling molecules in human T-cells under various stimulatory and inhibitory conditions. Recent advances in multi dimensional flow cytometry enable the measurement of multiple (~12) intracellular events at the molecular level, in single-cell resolution (1). We use Bayesian networks, a class of graphical models, to address questions of dependence and connectivity among biomolecules of interest in the flow cytometry dataset, consisting of MAPK and PI3K pathway components. These models extract probabilistic dependencies from data (Fig.1). Single-cell data of simultaneously observed biomolecules provides probabilistic power for modeling of molecular dependencies. In particular, the *acyclic* influence diagram among measured variables in a dataset can be extracted. Our resulting model contains many expected connections known from the literature, several unexpected connections that, upon closer examination, have been reported previously in the literature, and few as yet unconfirmed, putative novel connections that pose testable biological hypotheses. Our work demonstrates that Bayesian network modeling of multi dimensional flow cytometry data can help elucidate underlying signaling pathways and provide insight into the influence structure of molecules of interest. This approach can be

used to formulate testable influence hypotheses in a pathway of interest or to help elucidate pathways and points of cross-talk between pathways.

Figure 1. A small Bayesian network showing some dependencies extracted from flow cytometry measurements of representatives from the TNF and EGF signaling pathways. In the Bayesian network, the nodes (boxes and ovals) represent biomolecules while the edges represent dependencies.

References

1. L. Herzenberg, D. Parks, B. Sahaf, O. Perez, M. Roederer and L. Herzenberg, The History and Future of the Fluorescence Activated Cell Sorter and Flow Cytometry: A View from Stanford. *Clinical Chemistry*, (2002), 48:10, 1819:1827.