

Signalome-Transfactome Networks in Human B Cells

Kai Wang^{1,2}, Rune Linding³, Katia Basso⁴, Riccardo Dalla Favera⁴, Andrea Califano^{1,2,4,*}

1. Department of Biomedical Informatics, Columbia University, New York, NY, USA

2. Joint Centers for Systems Biology, Columbia University, New York, NY, USA

3. Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada

4. Institute of Cancer Genetics, Columbia University, New York, NY, USA

*email: califano@c2b2.columbia.edu

Signal transduction pathways mediate information transfer from the cell surface to its nucleus through a cascade of post-translational modification events. Eventually these processes activate or suppress specific transcriptional programs, thus determining the cell's response to environmental stimuli. Systematic mapping of signaling pathways and of their effect on transcriptional networks is a key challenge in molecular systems biology. Their understanding is increasingly crucial in the dissection of human disease and in the identification of targets for therapeutic intervention. Traditional signaling pathway studies have been focused on measuring the post-translational modification state of individual or of a handful of interacting proteins. Such studies have been limited by the difficulties in precisely measuring protein concentration and their specific post-translational modification state using high throughput approaches.

On the other hand, genome-wide transcriptional profiling can simultaneously monitor the gene expression programs of all genes in the cell, including those in signal transduction pathways, thus potentially providing a means to trace the signaling mechanisms and circuits that underline complex biological responses. Earlier studies of signal transduction network using gene expression profile data in yeast have been limited to the identification of gene sets regulated by a handful of signaling proteins [1], or to the reconstruction of signaling pathways using known protein-protein interactions as a backbone and gene co-expression as a ranking schema [2]. However, a genome-wide map of the signaling proteins affecting specific transcriptional programs, on a single transcription factor (TF) basis, has not been reported either in yeast or in higher order eukaryotes.

Previously we introduced the MINDy algorithm for the genome-wide identification of modulators of TF activity [3] and biochemically validated the predictions [4]. MINDy uses an information theoretic approach, based on the conditional mutual information, to detect changes in the transcriptional activity of a TF as a function of the modulator availability. Modulators predicted by MINDy can be both direct and indirect effectors of TF activity. We have shown that such modulators are highly enriched in signaling molecules, suggesting that the method could be used to systematically explore the interface between signaling and transcriptional pathways.

We thus applied MINDy to identify the putative modulators of a set of 595 human TFs (transfactome) among 772 signaling molecules (signalome), including 421 protein kinases, 113 phosphatases and 295 cell surface receptors, using a collection of 254 human B cell gene expression profiles. For each TF and signaling molecule pair, MINDy searches through entire dataset for potential TF-target interactions that are modulated by the signaling molecule. Among the 44,349 modulatory interactions predicted by MINDy, 71 ($p < 1.5 \times 10^{-6}$) are reported in public

databases (HPRD, BIND, DIP and IntAct) and 91 ($p < 1.3 \times 10^{-5}$) are validated in the literature. Furthermore, 115 ($p < 8.4 \times 10^{-16}$) direct kinase-substrate interactions were also reported in a recent systematic study of phosphorylation substrates using the NetworKIN method [5]. In total, 249 ($p < 2.2 \times 10^{-16}$) MINDy predictions have evidence for direct interactions. While this number seems to be small compared to the number of MINDy predictions, this is in part due to (a) given the transient nature of direct phosphorylation interactions, they are only sparsely represented in databases and literatures; (b) the space of possible overlap with networKIN predictions is small, limited by the number of kinases with known consensus substrate motifs and of experimentally identified phosphorylation sites; (c) MINDy predictions also include indirect, upstream interactions. Given the size of a typical signaling pathway, we estimate that only ~20 percent of the predicted modulators would be expected to be directly interacting with the TF layer.

In support of pathway mediated modulatory interactions identified by MINDy, we find that 1,907 ($p < 2.2 \times 10^{-16}$) of the predicted interactions are between a signaling molecule and a TF belonging to the same KEGG or GenMAPP cellular pathways. Finally, 11,787 ($p < 2.2 \times 10^{-16}$) interactions could be reduced to linear chains of events mediated by known PPIs and NetworKIN predicted phosphorylation interactions. Remarkably, as the statistical significance threshold for predicting a modulatory interaction becomes more stringent, the percentage of predictions that are validated by known pathways increases dramatically from 27% to over 60% (Figure 1), further validating the approach. We clustered the signalome-transfactome interactions based on the similarity of the transcriptional program affected by each signaling gene. This produced functionally coherent signaling molecule sets involved in known biological processes (Figure 2).

Taken together, this represents the first genome-wide computational analysis of the interface between signaling and transcriptional networks. The combination of these results and those from in-vivo experiments may significantly improve our understanding of the role of cellular signaling in the regulation of transcriptional programs and provide new targets for therapeutic intervention.

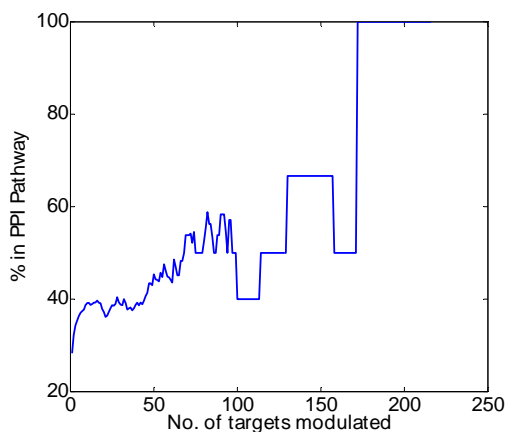


Figure 1: In-silico validation of the signalome-transfactome interactions. Plotted on the x-axis is the number of modulated TF-target interactions, which is used as a threshold for determining whether a signaling molecule is predicted to affect the TF. Y-axis records the percentage of total predicted modulatory interactions that can be mediated through known PPIs and phosphorylation interactions, corresponding to each threshold on the x-axis.

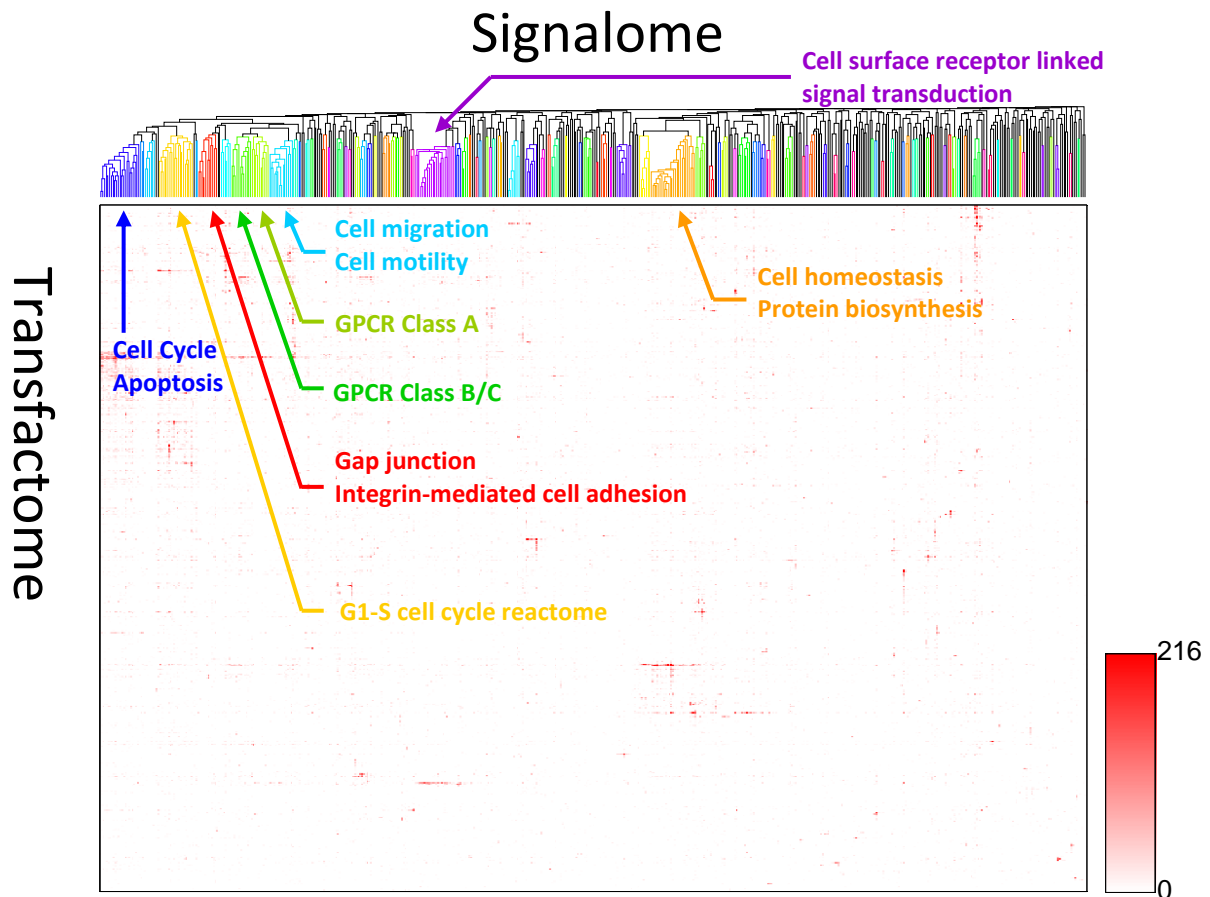


Figure 2: Clustering of human B cell signalome. In the plotted matrix, rows represent human TFs and columns signaling molecules. Each entry indicates the number of transcriptional interactions of a TF affected by a signaling molecule, and is color coded according to the color bar on the right. Hierarchical clustering of the column vectors is performed, and the columns are then rearranged according to the resulting tree. Clusters with more than 10 members are annotated with GO biological processes and KEGG pathways, and statistically significantly enriched categories (multiple testing corrected) are labeled on the graph.

Reference

1. Roberts, C.J., et al., *Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles*. *Science*, 2000. **287**: p. 873-880.
2. Steffen, M., et al., *Automated modeling of signal transduction networks*. *BMC Bioinformatics*, 2002. **3**(34).
3. Wang, K., et al., *Genome-Wide Discovery of Modulators of Transcriptional Interactions in Human B Lymphocytes*. *Lecture Notes in Computer Science*, 2006. **3909**: p. 348 - 362 (RECOMB 2006).
4. Wang, K., et al., *Genome-wide identification of transcriptional network modulators in human B cells*. Submitted, 2007 (manuscript available upon request).
5. Linding, R., et al., *Systematic Discovery of In Vivo Phosphorylation Networks*. *Cell*, 2007. **129**(7): p. 1415-1426.