Reverse Engineering Signaling Cascade from High Throughput Data

Sayak Mukherjee[1], William Stewart[1], and, Jayajit Das[1]

***Short Abstract* — The high throughput experiments opened for the first time a possibility of inferring the cell signaling networks from data. Presence of extrinsic noise in a population as well as tonic signaling however offers a challenge for inference. Using a combination of Maximum Entropy based inference scheme and simulated annealing we have developed a method that is capable of constructing an effective linear description of the underlying biochemical network. We have validated our method for synthetic data acquired from linear and non-linear reaction cascades and used it to infer signaling networks for primary NK cells from multi parameter CyTOF data.**

***Keywords* —CyTOF data, Maximum Entropy, simulated annealing, data driven modeling, NK cell signaling**

I. Purpose

The advent of high throughput measurements ushered in a new era in systems biology. The simultaneous monitoring of expression levels of different proteins at different time points for each single cell showed us that the time evolution of the proteins are highly co-regulated and opened the door for inferring the underlying network of interactions.

There are two major sources of noise [1] in a high throughput data. First, the presence of large amount of cell-to-cell fluctuations in the protein copy numbers as well as fluctuations in the kinetic binding and unbinding rates (extrinsic noise). Second, the fluctuations those arise due to the inherent probabilistic nature of biochemical reactions (intrinsic noise). Even in the limit of large copy numbers of proteins (weak intrinsic noise), the presence of large extrinsic noise can pose a formidable challenge in inferring networks. It can dominate over the variations in time evolution of proteins coming exclusively from the internal network architecture and can potentially render the inference scheme ineffective. In addition, there can be a substantial amount activation present in the unstimulated cells due tonic or basal signaling which makes it difficult to separate out the changes in activation that truly occurred due to a signaling response to external stimuli.

The purpose of this work is to develop a method that can identify and isolate the extrinsic noise and contribution from the tonic signaling from variations arising solely due to a signaling response and help us reverse engineer the directed causal architecture of the signaling network from the single cell time series data.

II. Method

We seek for an effective linear description of the real biochemical reaction cascade. To this end we sample for a linear module characterized by a corresponding **M** matrix given by $\partial_t C = \mathbf{M}C$, where $\boldsymbol{C}$ is the concentration vector, using simulated annealing [2] that minimizes the Euclidean distance between the moments generated by the sample and once calculated from the *in-silico* networks/experiments. The simulated annealing yields an effective linear description that is endowed with causal information. A Maximum Entropy based method is used to choose between different initial network topologies that are used in simulated annealing.

III. Conclusion

The method is very general; it probes the network architecture directly, it is insensitive towards extrinsic noise fluctuation, and separates out the contribution from basal signaling. We have validated the method against *in-silico* linear and non-linear reaction cascades. We have used this method for data acquired in CyTOF experiments performed at Lewis Lanier's lab (UCSF) with Jurkat T cells and primary NK cells. The primary NK cells are stimulated with cognate ligands for NKG2D activating receptors and we use the combined computational and experimental method to search for hitherto unknown novel interactions in NK cell signaling.

References

[1] Elowitz M. B, et al. (2002) Stochastic gene expression in single cells. Science 297:1183-1186.

[2] Kirkpatrick S, et al. (1983) Optimization by simulated annealing. Science 220:671-6800

Acknowledgements: This work was funded by the grant R56AI108880-01 from NIAID.

[1]BCMM, Research Institute at The Nationwide Children's Hospital. E-mail: mukherjee.39@osu.edu