Mechanism Inference from Single Cells (MISC)

Rachel A. Haggerty[1], Jeanette Baran-Gale[1] and Jeremy E. Purvis[1]

***Short Abstract* — Mechanism Inference from Single Cells (MISC) is a method we have developed to extract the structure and strength of kinetic interactions among two or more signaling molecules. The algorithm considers all possible network structures of a certain size that could relate the measurements and then tests each network structure against the single-cell data. Well performing network structures are used to find a consensus model that suggests potential underlying signaling mechanisms.**

***Keywords* — Signal transduction, single-cell, mechanism inference, networks, machine learning.**

I. Background

Network motifs in cellular signaling pathways are indicative of the underlying function of the signaling network [1]. Sometimes the interactions between two signaling molecules can be easily inferred, for example, if one directly activates or represses another. However, there are many instances where multiple molecules are involved in a signaling response, but the exact mechanism in which they interact with each other is not clear. To address this problem, we developed a computational method called MISC (Mechanism Inference from Single Cells) that uses paired time-series measurements from individual cells to predict the underlying network structure with no prior knowledge about the network architecture.

II. Detailed Algorithm

The input for MISC is two sets of time series traces from individual cells, such as those generated from fluorescent biosensors used in live-cell imaging experiments. The output is a ranked list of signaling mechanisms from a complete list of network topologies of a certain size that describe the kinetic relationships among the biosensors.

The steps of the algorithm are as follows: (1) All possible network structures of a given size that relate an input signal and output signal are enumerated. This is to say that every possible network with a set number of nodes or less with a path from the input signal to the output signal is generated. If it is unknown which signal is upstream, the algorithm may be run both ways. The ability to include additional nodes beyond those observed allows for the possibility of other unknown factors to influence the network. (2) Ordinary differential equations are automatically generated. MISC allows the user to specify the functional form of the equations for positive and negative links, so the method can be applied to a large range of systems. (3) Signals are simulated using the single-cell data from the input biosensor. These are fed through all networks and evaluated for how well each reproduces the output biosensor. (4) All possible networks are given a score based on how well it fits the output biosensor signal and ranked. (5) A consensus network is calculated to best describe the network.

III. Method Validation

A. Synthetic Data

To validate that our method works correctly, we generated a network with three nodes involved, but only associated biosensors with two nodes. Using this network, we generated ordinary differential equations to describe the system and used them to simulate single-cell data for the input and output biosensor by varying the initial conditions and adding noise.

B. Results

We tested the synthetic biosensor data on MISC to determine whether we could recover the original network we used to create the synthetic data. We ran the algorithm with the possible number of nodes set to three or less. Links that were present in the original network were enriched in the top ranked models. In addition, the consensus model, in this case created by clustering the top 10% of models and finding the centroid of the cluster with the lowest error, was very similar to the original network. Additionally, we then generated synthetic signals for if we had a biosensor on the third node, and the behavior of this signal was very similar for the original network and the consensus model network.

IV. Conclusion

MISC is an algorithm that allows us to discover the mechanism by which one factor is influenced by another via single-cell data. There is a large range of potential applications for which this method could also be applied. In general, it searches for plausible interaction networks between any two (or more) signals. Because it ranks all possible networks, it also suggests which sets of signaling motifs may have equivalent functions. Moreover, the ability of MISC to account for unobserved signaling molecules in the network allows for the discovery of novel factors and interactions.

References

[1] Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet. 2007 Jun;8(6):450-61.*

[1]Department of Genetics, Bioinformatics & Computational Biology Graduate Program, University of North Carolina at Chapel Hill. Correspondence E-mail: haggerty@unc.edu