Inferring Topology and Dynamical Properties of Genome-wide Regulatory Networks

Aviv Madar[1], Alex Greenfield[2], Devorah Kengmana[1], Harry Ostrer[4], Eric Vanden-Eijnden[3], Richard Bonneau[1,3]

Inferelator is a network reconstruction algorithm that uses expression and proteomics datasets to infer genome wide regulatory networks. Specifically, it learns a system of ordinary differential equations that describe the rate of change in transcription, of each gene, as a function of relevant predictors. Here we extend the method by: 1) implementing a resampling technique to produce an ensemble of networks that encapsulate potentially different network topologies and kinetic parameters; and 2) developing methods to refine this ensemble, via a MCMC dynamical modeling method, Inferelator-2. Initial results indicate significant improvement in our ability to both learn topology and model dynamics.

***Keywords* — Inferelator, network reconstruction, resampling, regulatory networks.**

LEARNING and characterizing regulatory networks, responsible for the remarkable ability of organisms to adapt to changing environment is a key problem in modern biology with applications spanning bioengineering, drug development, and many other biological fields. Detailed regulatory networks (RNs) can be modeled as a system of ordinary differential equations (ODEs), describing the rate of change in mRNA concentrations as a function of relevant predictors (e.g. transcription factors). We have recently described a network reconstruction algorithm, Inferelator-1[1], which infers regulatory influences for genes and gene clusters. The typical input is: 1) a microarray compendium composed of time-series and equilibrium measurements, and 2) prior information, such as a list of considered predictors. The output is a sparse dynamical model for each gene or gene cluster, i.e. an ODE describing the rate of change in transcription as a function of just a few predictors.

We have shown that RNs learned using Inferelator-1, at least for small genomes (e.g. *halobacterium*), could explain observed mRNA measurements (training-set), as well as predict unobserved mRNA measurements[2] (a large test-set). Inferelator-1, however; 1) approximates the predictors level to be constant throughout a time interval, which becomes a crude approximation as the time-interval length increases, 2) solves the system of ODEs as an uncoupled system which is not a realistic model for the underlying RN, and 3) produces only one set of ODEs, describing a single RN without an associated confidence estimate. These limitation diminish our ability to model more complex systems over long time intervals.

Here, we use resampling to create an ensemble of datasets as input for Inferlator-1. Thus, producing an ensemble of RNs as output. Using this ensemble we derive empirical-distributions for many putative regulatory-interactions together with their corresponding kinetic-parameters. To take advantage of these empirical distributions we have developed Inferelator-2 a Bayesian, dynamical modeling approach. Inferelator-2 searches for regulatory-interactions and kinetic-parameters, defining a set of ODEs, that maximizes the probability of this ODE set, given the observations (the posterior). In the core of Inferelator-2 is an importance sampling Markov Chain Monte Carlo algorithm, designed to efficiently sample the posterior. Importantly, to compare the merit of considered RNs, Inferelator-2 solves the system of ODEs as a coupled system, providing a more realistic model for the underlying RN.

Initial results, using synthetic and real datasets suggest that our resampling approach together with the new Inferelator-2 method, significantly boost our ability to correctly learn topology, and to model dynamics over longer time scales. These changes to our framework provides an essential step toward learning more complex RNs, such as mammalian RNs, and over longer time scales, such as the time scales required to model cell differentiation.

REFERENCES

[1] Bonneau R, et.al.(2006) The Inferelator: an algorithm for learning parsimonious regulator networks from systems biology data sets de novo. *Genome Biology* **7**, R36

[2] Bonneau R, et. al. (2007) A Predictive Model for Transcriptional Control of Physiology in a Free Living Cell. *Cell* **131**, 7:1354-1365

[1] Center for Genomics and Systems Biology, New York University madaraviv@nyu.edu, bonneau@nyu.edu
[2] Computational Biology Program, New York University ag1868@nyu.edu,
[3] Courant Institute of Mathematical Sciences, New York University eve2@cims.nyu.edu
[4] Human Genetics, New York University School of Medicine ostreh01@nyu.edu