Usage Report for Fiscal Year 2015

Project Title:

Inference of Gene Regulatory Network from Experimental Data

Name: Prabhat Shankar

Laboratory at RIKEN: Laboratory for Physical Biology

Background

In the embryonic development, pluripotent Embryonic Stem(ES) cells which are derived from the blastocyst of the early embryo plays a crucial role. Pluripotent cells give rise to all the somatic cells of the body. Their properties and regulatory mechanisms are currently an active field of investigation, as understanding their properties have numerous applications in cellular regeneration. To uncover the regulatory mechanisms of pluripotent cells, a number of experimental approaches are being applied. These approach consist of observing the gene expression profile of the wild type cell, as well of a cell after applying various perturbations in the form of gene Knockout or Overexpression.

But owing to the complexity of the underlying phenomena, a purely experimental study is not sufficient, and this approach must be complemented by numerical analysis of the obtained data, as well as by making a mathematical model of the system under study.. Such mathematical modeling is a computationally intensive process owing to the extremely large number of parameters associated with it.

This project is aimed towards understand the functioning of the transcription regulatory network in terms of a Boolean network. This approach is promising because of the intuitive interpretation of obtained Boolean models in terms of activation and repression of the involved regulatory interactions. But as the network size increases, the number of possible Boolean networks too increases superexponentially. This makes the use of high computing resources a necessity in solving this problem.

Methods

To make a Boolean model of the Pluripotent state, we collaborated with the Laboratory of Pluripotent Studies headed by Dr Hitoshi Niwa, who provided us with a set of gene expression profile. Based on this data, we identified the possible set of networks which can explain the pluripotent cells. The total number of candidate solution networks obtained from the data was of the order of 10^{10}. As this number is very large, we selected a small subset of this network based on minimizing the network size, which came out to be of the order of 10^8. For all such networks, we simulations based on the Boolean truth tables obtained from data. For these computations, the computing facilities provided by RICC were also used.

For performing the calculations, we wrote our own C programs, and to speed up the computation, we parallelized the search among candidate networks. This parallelization also allowed us to use the various cores available at RICC.

Results

Based on the time evolution among the candidate networks, we obtained a set of approximately 5000 networks which can completely explain the observed experimental data. An analysis of these solution networks identify some new connections in the transcription regulatory networks. Many already known connections were also present in the set of connections

obtained from solution networks. At the same time, based on these networks, we can make predictions for the perturbation experiments which have not been experimentally performed till now. Verifying these predictions will ultimately provide validity to our proposed model. If validated, such predictions can help to reduce the costs of the experiments performed. The experimental validation of the predictions from our model is proposed in next few months.

Future Prospects

This project is still continuing, as we are planning to analyze some more recent data obtained by the experimenter. Once we have analyzed the new set of data obtained, and have a satisfactory model which can explain how pluripotency is maintained with sufficient detail, we plan to publish our work.