Exploring early drivers of stem-cell reprogramming by an integrative statistical approach

Louella J. Vasquez[1,2,3,4] and Sach Mukherjee[2,3]

[1]Institute of Advanced Study, University of Warwick, Coventry CV47AL, UK
[2]Department of Statistics, University of Warwick, Coventry CV47AL, UK
[3]Centre for Complexity Science, University of Warwick, Coventry CV47AL, UK
[4] *L.J.Vasquez@warwick.ac.uk*

Abstract:

It is well established that induction of certain combinations of transcription factors (TFs) can inducepluripotency in somatic cells. However, the regulatory events by which TF induction brings about epigenetic reprogramming remain incompletely understood. Here, we carry out an exploratory, genome-wide analysis with the aim of highlighting a small number of genes that may be important in the early stages of reprogramming. We integrate gene expression data taken from mouse fibroblasts, obtained after 4 days under combinatorial expression of the four factors Oct4, Sox2, Klf4 and c-Myc, with ChIP-on-chip data for these factors. Gene expression is modelled as a function of the presence/absence of the factors, including higher-order terms to account for possible combinatorial effects. Genes are then ranked by evidence of influence by one or more factors, singly or in combination. Inference and integration with ChIP-on-chip data is carried out within a Bayesian statistical framework. Our results suggest testable hypotheses and may aid in the development of richer, dynamic models of reprogramming that focus on a smaller number of genes. (In collaboration with K. Saha & J. Hanna, Jaenisch Lab, Whitehead Institute, MIT)