Kevin Chau; Austin Crinklaw

The principal question of our main paper was, “how can time-course gene expression data be used to augment older differential expression methods?” To answer this question, the team behind the paper wrote a tool called Network Analysis Based on Co-expression Patterns (NACEP). This tool incorporates temporal expression patterns RNAseq analysis by applying novel clustering methods on the analyzed genes.

NACEP employs an infinite-mixture model to cluster gene time-course data. First, clusters are generated from a Chinese Restaurant Process. Next, a mixed-effects model is applied to the data such that each cluster mean temporal expression pattern can be derived from the observed gene expression patterns, factoring in biological fluctuations and random noise. From this, a generative probabilistic model can be calculated. This model is a Bayesian formulation of a Dirichlet Process such that posterior probabilities for every given gene’s membership to the calculated clusters can be calculated, factoring in the Chinese Restaurant Process as the prior probability. This forms the basis of the Gibbs Sampling algorithm employed by NACEP, and this algorithm is repeated for a given number of iterations, ideally converging on values. These probabilities are then used as weights for differential expression of genes between experimental conditions. Genes are ranked based on their statistical significance.

Validation of this tool consisted of analysis of retinoic-acid (RA)-induced embryonic stem (ES) cell differentiation. NACEP was run on data by Ivanova *et al.,* which consisted of mouse ES cell (control and RA-induced) time-course expression data. Top-ranked RA-induced genes were identified; these included Gli3, Zic3, etc. It was found that Shh and Gli3 were clustered together, providing evidence that the shh pathway is RA-activated and is a part of neuronal differentiation, as shown in other works. Additionally, Etv4 was identified by NACEP to be one of the top-repressed genes; Etv4 downregulates the shh pathway, which is consistent with this finding. Thus, new information was gained using NACEP, relative to those previous works.

We have elected to also perform our own analysis using NACEP. We had found publicly available prostate epithelial time-course expression data from the Gene Expression Omnibus. This data consisted of gene expressions from immortal prostate epithelial cells (iPrEC) and tumorigenic prostate epithelial cells (EMP). Our goal is to elucidate differentially expressed genes between the two conditions, factoring in temporal expression patterns. To do this, we first filtered the gene sets for expression as well as variance across their samples. We are now currently running NACEP on these datasets and will subsequently analyze the results for statistically significant genes.