

Gene Expression Variability in Single Cells Not So Dependent On Cell Cycle

September 15, 2014

ND Weber

Gene expression studies have recently begun to involve analysis at the single cell level. The ability to query gene expression variations in an individual cell will allow for co-expression pattern determination in which correlation between genes can be attributed to shared regulatory elements within each cell rather than as a response to varying biological conditions in the bulk cell population. Inherent in this analysis is the ability to separate the effects that cell cycle has on the variability in expression of a particular gene. In a recent study published in *PLoS Computational Biology*, Dr. Raphael Gottardo (Vaccine and Infectious Disease Division) and collaborators have developed a new modeling framework to better analyze data of single cell gene expression.

Another aspect of single cell gene expression analysis that is difficult to correctly compensate for in data modeling approaches is the bi-modality of individual gene expression. That is to say, the expression of an otherwise abundant gene will be either strongly positive or undetectable within an individual cell. In other words, averaging expression data among a cell population eliminates information about cell-to-cell variability. This characteristic is not easily dealt with by typical analytical tools such as linear modeling. Therefore, the researchers developed a novel computational framework by utilizing a modeling approach called the Hurdle model to better incorporate this characteristic of the data into the analysis.

"There has been a flurry of activity in efforts to model single cell RNA sequencing data," explains Andrew McDavid, a graduate student in the Gottardo Lab and the first author of the study. "The Hurdle model, like most models, has two components: the rate and strength of expression in high-expressing cells. So it appears that Hurdle models or Hurdle-like models are going to be an important part of understanding single cell gene expression."

In their analysis of the influence of cell cycle on gene expression in single cells, the researchers utilized the Hurdle model to derive better estimates of the percentage of gene expression variation that is attributable to the phase of the cell cycle. They found that for the median gene in their data set, cell cycle explains only 5%–18% of the variability, a smaller percentage than was previously thought. This implies that in attempting to analyze the single cell transcriptome, differences in cell

cycle phase will not be a major confounding factor.

"Future work needs to address the fact that the models largely have been phenomenological, which is to say the model attempts to capture features of the biology, without necessarily offering a mechanism to explain why these features appear," explained McDavid, "For example, the bimodality could be an irreducible technical artifact of aspects of the assay, or biological, or both." In order to address such questions, "careful experiments are going to be necessary to determine how much of the bimodality is technical versus biological," he explained, and this in turn, "will allow development of more mechanistic models of single cell gene expression."

This allows the results of their study to contribute to the improving of the methodologies involved in such studies. "Our finding of the 5%–18% of the variability being attributable to cell cycle should be seen in this light as well," he said, "It's a ratio of biological variability to everything that is left over after removing all the technical variability that we could measure. If future technologies can measure single cell gene expression with less technical variability, or we can ascribe more of the residual variability to technical causes, then this estimate can be revised upwards."

[McDavid A, Dennis L, Danaher P, Finak G, Krouse M, Wang A, Webster P, Beechem J, Gottardo R.](#) 2014. Modeling Bi-modality Improves Characterization of Cell Cycle on Gene Expression in Single Cells. *PLoS Comput Biol* 10(7):e1003696.

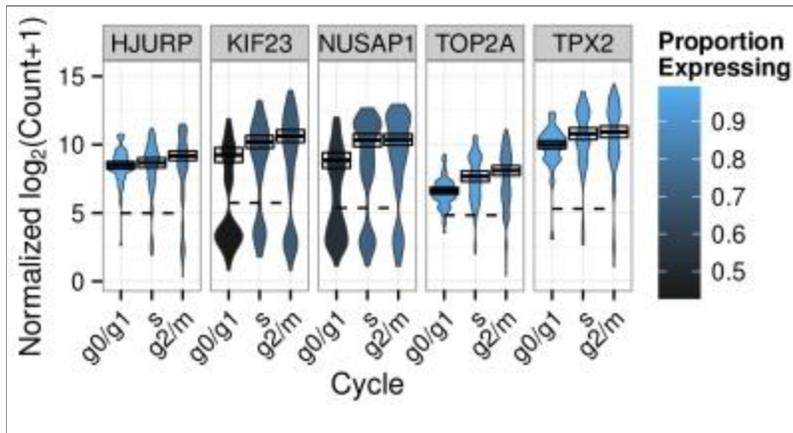


Image adapted from the manuscript

Top 5 cell cycle genes detected by the Hurdle model. A violin plot shows the density of log counts of mRNA for 5 cell cycle genes according to the three cell cycle phases. The expression threshold estimated for each gene is shown as a dashed line, so that the ratio of area above the dashed line reflects the proportion of cells expressing a gene. Blue shades of the violin depict genes with more expressing cells in a condition. The positive mean and 95% confidence interval is depicted as a box with solid line. The Hurdle model combines evidence for changes in either of these parameters, after adjusting for cell line, to determine statistical significance.