**Title:**

Modelling the ordering of cell-cycle phase in single-cell RNA-seq data

**Authors:**

Chiaowen Joyce Hsiao, Kushal K Dey, PoYuan Tung, Yoav Gilad, and Matthew Stephens

**Affiliations:**

1Department of Statistics, University of Chicago, Chicago, IL

2Department of Human Genetics, University of Chicago, Chicago, IL

**Keywords:**

single-cell, RNA-seq, cell-cycle, iPSC

Single-cell RNA sequencing (scRNA-seq) is emerging as a powerful technology for understanding transitions between cell states and markers that distinguish cell types at the genome-wide transcriptional level. The applications of scRNA-seq have expanded a variety of fundamental cellular processes including differentiation and pluripotency. Notably, scRNA-seq has proven useful for studies of embryonic stem cells (human and mouse) that investigate marker genes for transitions between pluripotency states or between developmental stages. Specifically, we can begin to ask questions such as: what is the relationship between cell state and variability in gene expression profile from cell to cell (cell-cell)? – thus moving beyond the applications of bulk-sample RNA-seq data which collapse expression signals across cells.

However, the single-cell perspective provided by scRNA-seq data comes with a cost: Currently, there is no a good way to track cell-cycle phase in scRNA-seq (G0, G1, S, G2, M), which is associated with cell-cell variability in expression profiles. For example, the primate embryonic stem cells (ESCs) cell-cycle phase structure is usually characterized by a short G1 phase and a long S1 phase, and this structure can be observed in scRNA-seq data by the small number of G1 cells and large number of S1 cells. As ESCs differentiate, the G1 phase becomes longer, and we observe a larger number of G1 cells than when cells are in an earlier differentiation stage (White and Dalton, 2005). Therefore, scRNA-seq is described as a technology that takes a “snapshot” of the cells and provides a cross-sectional view of the cell-cycle phase structure. understanding cell-cycle phase structure has become one of the most important challenges in the analysis of scRNA-seq data.

Modeling cell-cycle phase is becoming one of the most important topics in the analysis of scRNA-seq data. By doing some, we are able to study cellular processes such as differentiation while controlling for cell-cell variation due to cell-cycle phase. We propose a model-based approach to estimate the pseudo-time order of the cells based on their expression profiles that are associated with cell-cycle and cell-cycle regulation genes. In our model, we assume that the number of pseudo-time points to be the number of cells, and that the cell samples comprise of a complete cell-cycle structure spanning G0, G1, S, G2, and M. Our model follows an iterative scheme, where in each iteration, we pick a pseudo-time for each cell at random and fit a sinusoidal model with gene-specific amplitude and frequency. The full model that combines data over all cells and over all relevant genes is then iterated through sets of pseudo-time points. In our evaluations, we found that our sinusoidal approach performs similarly as non-sinusoidal approaches that include LOESS, Spline smoother and Wavelet smoother. Furthermore, simulation case studies of the ability to recover cell-cycle order under scenarios of low to high amplitude-to-noise ratios showed that the sinusoidal approach outperforms principal component analysis (PCA).

Finally, we also applied our model to 220 undifferentiated induced pluripotent stem cells. Results of this application indicated that we were able to recover the cell-cycle phase structure when including a set of genes that are related to cell-cycle or cell-cycle regulation. However, when including all genes, the model was unable to recover the cell-cycle phase structure, suggesting that cell-cycle phase is one of the many biological processes driving the gene expression profiles. Our next step is to focus on selecting a well-informed set of genes that are relevant to cell-cycle phase structure, from a statistical method of variable selection and also a controlled biological experiment of cell-cycle related genes.