Tracing noisy biological progression and gene network rewiring between cell metastable states in static single-cell transcriptomes

Pablo Cordero and Joshua M. Stuart

UC Santa Cruz Genomics Institute, University of California, Santa Cruz, California, USA

Background

Understanding the dynamics of gene expression and regulatory networks as a cell undergoes biological processes is crucial for dissecting the molecular mechanistic underpinnings of complex biological processes such as differentiation and oncogenesis. Static transcriptome measurements of cell populations at the single-cell level have recently emerged as promising tools to dissect these dynamics by inferring the underlying progression. However, it remains unclear how to adequately elucidate cell types from these noisy data, simultaneously pinpoint their regulatory networks, and how gene expression is rewired during cell state transitions.

Results

To address the above challenges, we propose a strategy, Single Cell Inference of Morphing and Interdependent Trajectories and their Associated Regulatory networks (SCIMITAR), for inferring gene expression network dynamics throughout biological progression from static, single-cell, transcriptomes. SCIMITAR's approach is top-down. First, we focus on detecting recurrent, metastable transcriptional states and any connections between them supported by transitioning cells. Second, we give a detailed, full probabilistic description of each path in the metastable state graph, explicitly accounting for heteroscedastic noise in the data and detecting gene-to-gene expression correlations at each point in the progression. To achieve this, we extend Gaussian mixtures with discrete components to a smooth, continuous mixture of ‘morphing’ distributions. The inferred model allows tracking the rewiring of gene regulatory networks between metastable states and can elicit predictions on data that it wasn’t trained on, such as mapping new samples, including experimental replicates, to the model. Further, the probabilistic nature of SCIMITAR transition models allows for evaluating the shape of the multivariate gene expression distribution as a function of biological progression, which we show can be used to pinpoint stable and transitional cell states.

We tested whether SCIMITAR could yield insights in the developmental trajectory of human fetal neurons by analyzing recent, publicly-available, single-cell transcriptomic measurements, focusing on 578 expressed transcription factors. SCIMITAR pinpointed factors that were expressed in various ways across three metastable states: some went up at the beginning of the transition (in replicating neurons), others were expressed only in the middle of the quiescent state or in the end. A likelihood ratio test designed for the SCIMITAR model revealed 35 genes that significantly varied throughout the progression but that were missed by standard differential expression between cells grouped in

supervised and unsupervised ways. The genes revealed by SCIMITAR involved in the Jak-STAT pathway that presented a coordinated expression pattern in the middle of the developmental trajectory. Further, the SCIMITAR model pinpointed a previously unidentified transitional state between fetal replicating and fetal quiescent neurons. Finally, SCIMITAR also revealed regulatory network rewiring events as gene co-expression degrees changed through the progression, unveiling coordinated regulation of MAP kinases, morphogenesis, and STAT factors throughout the progression as well as potential master regulators.

Conclusions

Static, single-cell transcriptomic measurements hold great promise for revealing the cell state dynamics of a multitude of biological processes. Inferring biological progression from these data requires computational methods that can model the individual cells as an evolving gene expression distribution, a feat that can only be achieved by fully embracing the heteroscedacity of the data. Our proposed method, SCIMITAR, leverages this heteroscedacity to track gene expression rewirings and cell state switches in a continuous model of biological progression. These rich models allowed dissecting the progression dynamics in the transition between human fetal replicating and quiescent neurons, revealing Jak-STAT related genes missed by traditional, population-based differential expression and rise and fall of co-expression networks enriched with diverse kinases and developmental factors. We expect SCIMITAR to be widely used to dissect these gene expression and network progressions from static, single-cell measurements that are now becoming a standard technique to tackle complex cell state processes.