

Dimension reduction and visualization of single cell RNA sequencing data by deep variational autoencoder

Jin Gu

Tsinghua University
E-mail: jgu@tsinghua.edu.cn

Abstract: Single cell RNA sequencing (scRNA-seq) is a powerful technique to analyze the transcriptomic heterogeneities in single cell level. It is an important step for studying cell sub-populations and lineages based on scRNA-seq data by finding an effective low-dimensional representation and visualization of the original data. The scRNA-seq data are much noisier than traditional bulk RNA-Seq: in the single cell level, the transcriptional fluctuations are much larger than the average of a cell population and the low amount of RNA transcripts will increase the rate of technical dropout events. In this study, we proposed VASC (deep Variational Autoencoder for scRNA-seq data), a deep multi-layer generative model, for the unsupervised dimension reduction and visualization of scRNA-seq data. It can explicitly model the dropout events and find the nonlinear hierarchical feature representations of the original data. Tested on twenty datasets, VASC shows superior performances in most cases and broader dataset compatibility compared with four state-of-the-art dimension reduction and visualization methods. Also, VASC makes better representations for very rare cell populations in the 2D visualization. Then, for a case study of pre-implantation embryos, VASC successfully re-establishes the cell dynamics and identifies several candidate marker genes associated with the early embryo development.