

A statistical simulator scDesign for rational scRNA-seq experimental design

Jingyi Jessica Li

University of California, Los Angeles

E-mail: jli@stat.ucla.edu

Abstract: Motivation: Single-cell RNA sequencing (scRNA-seq) has revolutionized biological sciences by revealing genome-wide gene expression levels within individual cells. However, a critical challenge faced by researchers is how to optimize the choices of sequencing platforms, sequencing depths and cell numbers in designing scRNA-seq experiments, so as to balance the exploration of the depth and breadth of transcriptome information.

Results: Here we present a flexible and robust simulator, scDesign, the first statistical framework for researchers to quantitatively assess practical scRNA-seq experimental design in the context of differential gene expression analysis. In addition to experimental design, scDesign also assists computational method development by generating high-quality synthetic scRNA-seq datasets under customized experimental settings. In an evaluation based on 17 cell types and 6 different protocols, scDesign outperformed four state-of-the-art scRNA-seq simulation methods and led to rational experimental design. In addition, scDesign demonstrates reproducibility across biological replicates and independent studies. We also discuss the performance of multiple differential expression and dimension reduction methods based on the protocol-dependent scRNA-seq data generated by scDesign. scDesign is expected to be an effective bioinformatic tool that assists rational scRNA-seq experimental design and comparison of scRNA-seq computational methods based on specific research goals.

Availability and implementation: We have implemented our method in the R package scDesign, which is freely available at <https://github.com/Vivianstats/scDesign>.