Flexible Experimental Designs for Valid Single-cell RNA-sequencing Experiments Allowing Batch Effects Correction

Fangda Song[1]; Ga Ming Angus Chan[1]; Yingying Wei[1]

[1] The Chinese University of Hong Kong

Abstract:
Despite their widespread applications, single-cell RNA-sequencing (scRNA-seq) experiments are still plagued by batch effects and dropout events. Although the completely randomized experimental design has frequently been advocated to control for batch effects, it is rarely implemented in real applications due to time and budget constraints. Here, we mathematically prove that under two more flexible and realistic experimental designs—the "reference panel"and the "chain-type" designs—true biological variability can also be separated from batch effects. We develop Batch effects correction with Unknown Subtypes for scRNA-seq data (BUSseq), which is an interpretable Bayesian hierarchical model that closely follows the data-generating mechanism of scRNA-seq experiments. BUSseq can simultaneously correct batch effects, cluster cell types, impute missing data caused by dropout events, and detect differentially expressed genes without requiring a preliminary normalization step. We demonstrate that BUSseq outperforms existing methods with simulated and real data.

Keywords:
Single-cell sequencing; Experimental design; Bayesian hierarchical model; Model identifiability; Missing not at random