Robust Detection of Alternative Splicing in a Population of Single Cells

Joshua D. Welch[1], Yin Hu[2], Jan F. Prins[1]

[1]*Department of Computer Science, UNC Chapel Hill.* [2]*Sage Bionetworks.*

ABSTRACT

Single cell RNA-seq data promises to be an invaluable tool for characterizing cellular heterogeneity, but study of alternative splicing in single cells has been limited by the unique challenges of single cell data and lack of suitable analysis methods. We present SingleSplice, which is to our knowledge the first algorithm for identifying alternative splicing in a population of single cells. SingleSplice uses a statistical model trained on the technical noise profile of synthetic spike-in transcripts to identify genes exhibiting biological variation in isoform composition. We applied SingleSplice to data from 279 mouse embryonic stem cells and discovered genes that show significant alternative splicing across the set of cells. A subset of these genes are linked to cell cycle stage, suggesting a novel connection between alternative splicing and the cell cycle. Using SingleSplice, we also characterized the isoform usage heterogeneity of 466 adult and fetal human cortical cells. SingleSplice is directly applicable to the profusion of publicly available single cell RNA-seq datasets and thus immediately opens a number of interesting biological questions for investigation.

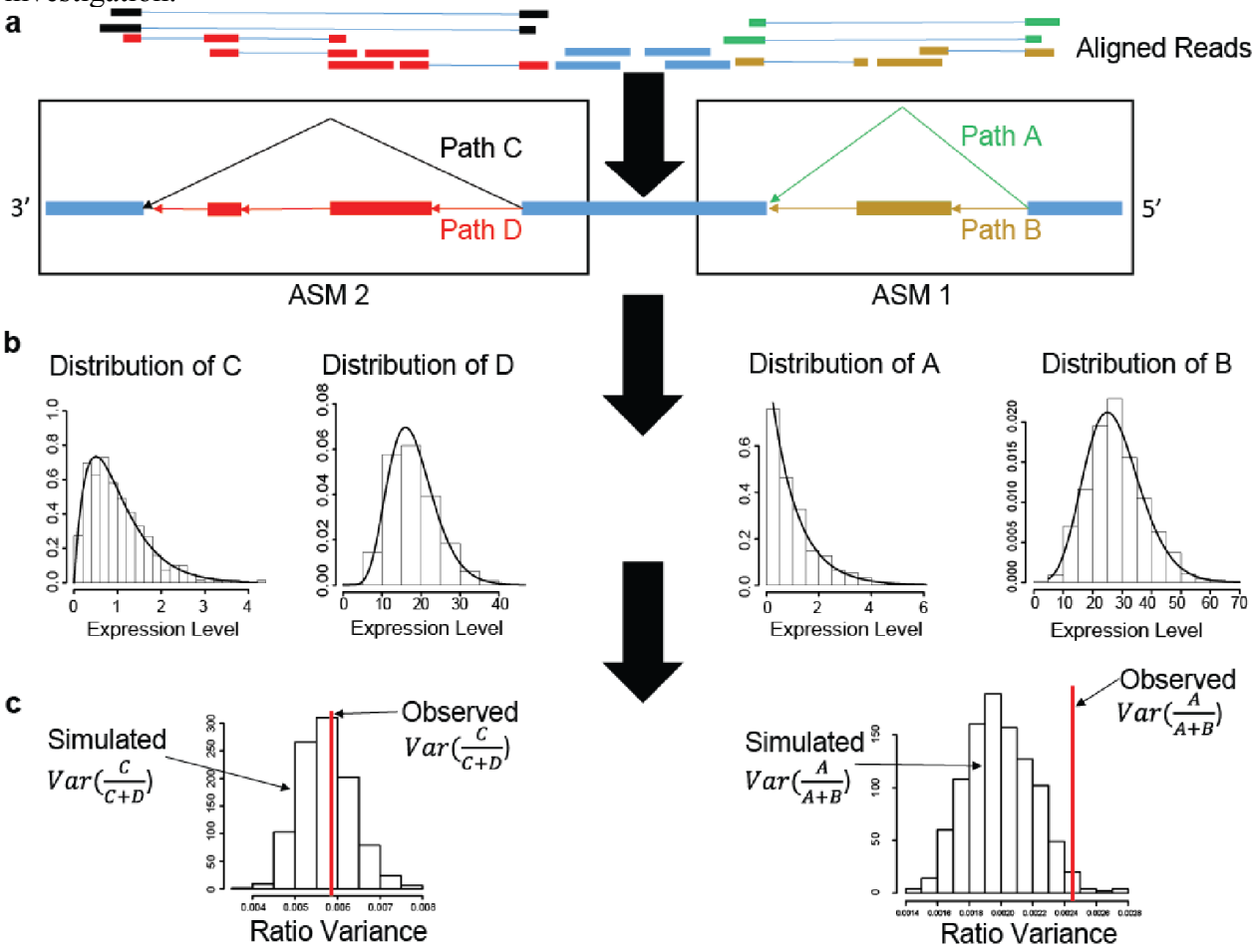

Figure 1: Overview of SingleSplice method (a) Alternative splicing modules (ASMs) are built directly from read alignments without annotation. The abundance of each ASM path is then estimated. (b) Distributions describing expected technical variation are fit for each ASM path. (c) Expected variation in isoform composition due to technical noise is then computed by parametric bootstrapping for each pair of ASM paths.