

Bayesian approach to single-cell differential expression analysis

Peter V Kharchenko^{1–3}, Lev Silberstein^{3–5} & David T Scadden^{3–5}

Single-cell data provide a means to dissect the composition of complex tissues and specialized cellular environments. However, the analysis of such measurements is complicated by high levels of technical noise and intrinsic biological variability. We describe a probabilistic model of expression-magnitude distortions typical of single-cell RNA-sequencing measurements, which enables detection of differential expression signatures and identification of subpopulations of cells in a way that is more tolerant of noise.

standard RNA-seq analysis methods^{13,14}, and the reported sets of top differentially expressed genes can include high-magnitude outliers or dropout events, showing poor consistency within each cell population (Fig. 1b). The abundance of dropout events has been previously noted in single-cell quantitative PCR data and accommodated with zero-inflated distributions¹⁵.

Two prominent characteristics of dropout events make them informative in further analysis of expression state. First, the overall dropout rates are consistently higher in some single-cell samples than in others (Supplementary Figs. 1 and 2), indicating that the contribution of an individual sample to the downstream cumulative analysis should be weighted accordingly. Second, the dropout rate for a given cell depends on the average expression magnitude of a gene in a population, with dropouts being more frequent for genes with lower expression magnitude. Quantification of such dependency provides evidence about the true expression magnitude. For instance, dropout of a gene observed at very high expression magnitude in other cells is more likely to be indicative of true expression differences than of stochastic variability.

We modeled the measurement of each cell as a mixture of two probabilistic processes – one in which the transcript is amplified

¹Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, USA. ²Hematology/Oncology Program, Children's Hospital, Boston, Massachusetts, USA. ³Harvard Stem Cell Institute, Cambridge, Massachusetts, USA. ⁴Center for Regenerative Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA. ⁵Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts, USA. Correspondence should be addressed to P.V.K. (peter.kharchenko@post.harvard.edu).

RECEIVED 2 OCTOBER 2013; ACCEPTED 28 MARCH 2014; PUBLISHED ONLINE 18 MAY 2014; DOI:10.1038/NMETH.2967