

1154-VC-2544 **Jonathan Fischer*** (jrfischer@berkeley.edu), jrfischer@berkeley.edu, and **Dan D. Erdmann-Pham** and **Yun S. Song**. *A generative model-based deconvolution of bulk expression data using single-cell references*. Preliminary report.

Bulk gene expression experiments estimate thousands of transcript levels averaged over myriad cells. Unfortunately, direct comparison of different bulk expression profiles is complicated by the mixtures of distinct cell types in each sample, obscuring whether perceived differences are actually due to changes in expression or cell type composition. Recent advances make it possible to measure gene expression in individual cells, achieving higher resolution in exchange for increased noise. If carefully incorporated, such data can be used as references for the supervised deconvolution of bulk samples to yield cell type proportions. This permits us to disentangle the effects of differential expression and cell type mixtures, both of which are relevant to our understanding of aging and disease. Previous solutions prioritize the inference of cell type proportions at the expense of error quantification or the ability to identify when cell types are missing from the reference. We thus propose a generative model which uses asymptotic theory and a robust estimation procedure to supplement cell type proportion estimates with confidence intervals and a hypothesis test for reference cell type missingness. We demonstrate the effectiveness of our approach on real and simulated data. (Received September 17, 2019)