Genome Biology

REVIEW

Open Access

Challenges and opportunities to computationally deconvolve heterogeneous tissue with varying cell sizes using single-cell RNA-sequencing datasets

Sean K. Maden[1], Sang Ho Kwon[2,3], Louise A. Huuki-Myers[2], Leonardo Collado-Torres[1,2], Stephanie C. Hicks[1,4,5,6*†] and Kristen R. Maynard[2,3,7*†]

†Stephanie C. Hicks and Kristen R. Maynard share equal contributions and are co-corresponding senior authors.

*Correspondence: shicks19@jhu.edu; kristen.maynard@libd.org

[1] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[2] Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA
[3] The Solomon H. Snyder Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, USA
[4] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA
[5] Center for Computational Biology, Johns Hopkins University, Baltimore, MD, USA
[6] Malone Center for Engineering in Healthcare, Johns Hopkins University, Baltimore, MD, USA
[7] Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA

Abstract

Deconvolution of cell mixtures in "bulk" transcriptomic samples from homogenate human tissue is important for understanding disease pathologies. However, several experimental and computational challenges impede transcriptomics-based deconvolution approaches using single-cell/nucleus RNA-seq reference atlases. Cells from the brain and blood have substantially different sizes, total mRNA, and transcriptional activities, and existing approaches may quantify total mRNA instead of cell type proportions. Further, standards are lacking for the use of cell reference atlases and integrative analyses of single-cell and spatial transcriptomics data. We discuss how to approach these key challenges with orthogonal "gold standard" datasets for evaluating deconvolution methods.

Keywords: Deconvolution, Single-cell RNA-sequencing, Single-nucleus RNA-sequencing, Cell sizes

Introduction

An important challenge in the analysis of gene expression data from complex tissue homogenates measured with RNA-sequencing (bulk RNA-seq) is to reconcile cellular heterogeneity or unique gene expression profiles of distinct cell types in the sample. A prime example is bulk RNA-seq data from human brain tissue, which consists of two major categories of cell types, neurons and glia, both of which have distinct morphologies, cell sizes, and functions across brain regions and sub-regions [1–3]. Failing to account for biases driven by molecular and biological characteristics of distinct cell types can lead to inaccurate cell type proportion estimates from deconvolution of complex tissue such as the brain [3].