Identifying Systematic Variation at the Single-Cell Level by Leveraging Low-Resolution Population-Level Data

Elior Rahmani[1(✉)], Michael I. Jordan[1,2], and Nir Yosef[1,3,4,5(✉)]

[1] Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA, USA
{erahmani,niryosef}@berkeley.edu, jordan@cs.berkeley.edu
[2] Department of Statistics, University of California, Berkeley, Berkeley, CA, USA
[3] Center for Computational Biology, University of California, Berkeley, Berkeley, CA, USA
[4] Chan-Zuckerberg Biohub, San Francisco, CA, USA
[5] Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA

Abstract. A major limitation in single-cell genomics is a lack of ability to conduct cost-effective population-level studies. As a result, much of the current research in single-cell genomics focuses on biological processes that are broadly conserved across individuals, such as cellular organization and tissue development. This limitation prevents us from studying the etiology of experimental or clinical conditions that may be inconsistent across individuals owing to molecular variation and a wide range of effects in the population. In order to address this gap, we developed "kernel of integrated single cells" (Keris), a novel model-based framework to inform the analysis of single-cell gene expression data with population-level effects of a condition of interest. By inferring cell-type-specific moments and their variation across conditions using large tissue-level bulk data representing a population, Keris allows us to generate testable hypotheses at the single-cell level that would otherwise require collecting single-cell data from a large number of donors. Within the Keris framework, we show how the combination of low-resolution, large bulk data with small but high-resolution single-cell data enables the identification of changes in cell-subtype compositions and the characterization of subpopulations of cells that are affected by a condition of interest. Using Keris we estimate linear and non-linear age-associated changes in cell-type expression in large bulk peripheral blood mononuclear cells (PBMC) data. Combining with three independent single-cell PBMC datasets, we demonstrate that Keris can identify changes in cell-subtype composition with age and capture cell-type-specific subpopulations of senescent cells. This demonstrates the promise of enhancing single-cell data with population-level information to study compositional changes and to profile condition-affected subpopulations of cells, and provides a potential resource of targets for future clinical interventions.

A preprint of the full paper is available at https://www.biorxiv.org/content/10.1101/2022.01.27.478115v1.

I. Pe'er (Ed.): RECOMB 2022, LNBI 13278, p. 371, 2022.
https://doi.org/10.1007/978-3-031-04749-7_32