TMF: Cell-type specific inference from bulk RNA-sequencing data by integrating single cell reference profiles via EPIC-unmix

Principal Investigator: Quan Sun, PhD

Project Title: Cell-type specific inference from bulk RNA-sequencing data by integrating single cell reference profiles via EPIC-unmix

Abstract: Gene expression levels are usually measured from tissue samples by RNA-sequencing, which are commonly referred to as bulk RNA-seq data. Although such bulk RNA-seq data can reflect disease etiology to some degree, they cannot capture the functional heterogeneity across cell types because bulk tissue samples contain a mixture of different cell types. To gain celltypespecific (CTS) insights, we need single-cell (sc) or single-nuclei (sn) RNA-seq data in population samples who also have disease related phenotypes measured. However, the vast majority of sc/snRNA-seq data are derived from <10 individuals, with very few largest ones from 200-300 individuals 1,2. Before sc/snRNA-seq data become available in large samples of individuals, computational methods are in pressing needs to estimate CTS gene expression directly from bulk RNA-seq data. To achieve this goal, we propose EPIC-unmix, EmPirical Bayesian cell type specifIC unmixing of bulk expression profiles, a novel method that infers CTS profiles for every sample with bulk RNA-seq data. EPIC-unmix is a two-step empirical Bayesian method that integrates sc/snRNA-seq reference data and bulk RNA-seq data from target samples to update prior parameters, improving accuracy of CTS expression inference in target samples. The specific aims of this study include: (1) develop and refine the EPIC-unmix method; (2) compare EPIC-unmix with alternative methods, specifically TCA 3 and bMIND 4 through simulation and real-data analysis; (3) apply EPIC-unmix to real bulk RNA-seq data and perform CTS eQTL analysis.

PI:	Quan Sun
Award Type:	TOPMed Fellowship
Start Year:	2022