

Technical summary, PhD thesis

BIOINFORMATIC TOOLS FOR SINGLE-CELL DATA ANALYSIS IN CLINICAL STUDIES

Brinda Monian
Department of Chemical Engineering

Thesis committee: J. Christopher Love (adviser), Douglas Lauffenburger, Wayne Shreffler

Mechanistic understanding of disease has been dramatically enhanced by an explosion of new high-throughput experimental techniques for profiling biological samples, including RNA-Seq, mass spectrometry, and single-cell sequencing. However, the ability to gather exponentially more measurements comes with the pitfalls of increased Type I error and reduced interpretability. In theory, single-cell measurements can be helpful in combating this problem, since each sample of cells represents hundreds to thousands of observations. But thinking is still emerging on how best to utilize single-cell data to boost statistics and generate meaningful findings.

This thesis represents several parallel efforts to develop and apply new bioinformatic techniques to generate robust findings from single-cell data. The advances are especially pertinent for small clinical studies with low sample numbers. In the first part of the thesis, two classes of methods are introduced: gene module discovery in single-cell RNA sequencing data using sparse PCA, and probability-based metrics for evaluating the association between two modalities of single-cell data (in this case, single-cell RNA sequencing and paired TCR sequencing data). The methods are shown on two different human datasets, as proof-of-concept and examples of the biological findings capable of being unearthed.

In the second part of the thesis, these methods are applied to larger clinical datasets with questions surrounding acquired tolerance and clinical reactivity in food allergy. In the first study, T-helper cells from peanut-allergic patients undergoing oral immunotherapy were profiled to identify therapy-induced effects and baseline predictors of outcome. Two distinct subsets of expanded TH2 clonotypes were found to be functionally suppressed, but not deleted, by the therapy. In the second study, transcriptional correlates of clinical reactivity were evaluated in peanut-activated memory T-helper cells from peanut-allergic adults. Cells from more reactive patients had higher expression of TH1 and MHC I gene programs, suggesting a pathogenic role for auxiliary, non-TH2 cell types. In each of these studies, new single-cell analysis techniques were integrated to generate clinical findings with improved robustness and interpretability.