

Peter Kharchenko, Harvard

Title:

Comparative analysis of disease-oriented single-cell transcriptional dataset collections

Abstract:

Single-cell RNA-seq assays are being increasingly applied in complex study designs, which involve measurements of many samples, commonly spanning multiple individuals, conditions, or tissue compartments. Joint analysis of such extensive, and often heterogeneous, sample collections requires a way of identifying and tracking recurrent cell subpopulations across the entire collection, and an effective way of exploring contrasts between samples. We develop comparative approaches for analysis of case-control study designs, commonly used to study the impact of disease or a drug on a particular tissue. The analysis starts by establishing probabilistic mapping – a joint graph – connecting all cells within the collection. The graph can then be used to propagate information between samples and to identify cell communities that show consistent grouping across broad subsets of the collected samples. The contrast between conditions is then formulated in terms of i) compositional shifts between different cell populations, ii) transcriptional shifts within the distinct cell populations, and iii) within-group expression state variability. The compositional data analysis techniques are applied in the context of cell types hierarchies to suggest most parsimonious explanation of the changes. The differential expression analysis is used to identify most affected cell types, and characterize the likely functional interpretation of the cell type-specific changes. We illustrate the application of these methods in the context of different studies of human disease, including cancer and diseases affecting the brain.