Assessing changes in cell composition between biological conditions in single-cell data

Alemu Takele Assefa, Bie Verbist, **Koen Van den Berge**

Janssen Research and Development

Assessing differences in cellular composition between conditions and disease states is of principal interest in biology, helping in unraveling diseases and informing drug development. The data for such analyses typically consist of a count matrix, where each element of the matrix denotes the number of cells observed for a particular cell identity (be it cell type or state) in a sample. These data are compositional, as the count for each sample is constrained by an arbitrary total, making the counts negatively correlated between cell identities within a sample. While recent methods respect the compositional nature, often a naive analysis is performed where this is ignored, resulting in decreased performances of the analysis. In addition, modern methods accounting for compositionality using compositional transformations often ignore the heteroscedastic nature of the original count data. In this work, we evaluate currently available approaches for testing differential cell type composition. We introduce a workflow that combines building blocks from state-of-the-art methodology by applying a compositional transformation to the cell count data while adressing the mean-variance association that remains after the transformation. A bias correction is necessary as a result of the compositional transformation, and we discuss statistical inference taking into account the uncertainty of the bias correction. The benefits are demonstrated using simulations and real data.

Bio: Koen Van den Berge

Koen Van den Berge is a Statistician, with a PhD in Statistical Genomics from Ghent University. He performed postdoctoral studies at the University of California, Berkeley and Ghent University, developing statistical methods to analyze biological high-throughput sequencing data, e.g. (single-cell) RNA-seq data.

In his role as Statistician at Janssen R&D, he works on statistical analysis of single-cell datasets with the ultimate goal of improving therapies for important diseases. We analyze datasets and develop statistical methodology for robust population-scale analysis of large single-cell datasets, focussing on single-cell RNA-sequencing.