From scRNA-seq count matrices to phylogenetic inference: rethinking the problem of lineage reconstruction.

Center for Integrative Bioinformatics Vienna (CIBIV), Max Perutz Labs,
University of Vienna and Medical University of Vienna, Vienna BioCenter, Vienna, Austria

In recent decades, there has been increased interest in the reconstruction of high-resolution developmental lineages based on single-cell RNA sequencing (scRNA-seq) data. On the bioinformatic side, numerous tools have been developed to understand trajectories based on a cell's gene expression profile. Such tools share a common idea: they aim to adapt lineage construction algorithms to the format of pre-processed and often already annotated scRNA-seq experiments. However, the assessment of such methods is unclear as long as scRNA-seq analysis pipelines are not sufficiently standardized and discussions about, e.g., correct count normalization and embedding of expression profiles remain unresolved. Even when resulting trajectories seemingly show expected biological dynamics, methods lack subsequent statistical inference to critically assess the quality or stability of the results.

Therefore, we would like to motivate rethinking lineage reconstruction by not adapting algorithms but, on the contrary, reformulate the underlying problem into a phylogenetic framework and use already well-established phylogenomic tools, e.g., IQ-TREE. Exemplary reconstructed trees, leaf-labeled with single cells, already resemble clustering annotations attained in scRNA-seq downstream analysis. Moreover, in contrast to the common practice of assigning cells to an indefinite number of clusters, the tree is capable of visualizing transient cluster boundaries as well as the extent of variability between and within clusters. Apart from illustration, this approach opens an endless collection of phylogenetic analysis such as, to only name one, bootstrap information allowing for thorough statistical assessment of cluster stability.