SCITUNA: A Network Alignment Approach for Integrating Multiple Single-Cell RNA-Seq Datasets

Onur   Dogan onur.dogan@std.antalya.edu.tr Turkey Antalya Bilim University
Burak Onur   Erten   Turkey Antalya Bilim University
Cesim   Erten   Turkey Antalya Bilim University
Aissa   Houdjedj   Turkey Antalya Bilim University
Hilal   Kazan hilalkazan@gmail.com Turkey Antalya Bilim University
Mohamed   Krichen   Turkey Antalya Bilim University
Yacine Marouf   Turkey Antalya Bilim University

The throughput and cost of single-cell RNA sequencing (scRNA-seq) are in continuous improvement, and so is the demand for larger-scale scRNA-seq data, which could require integrating multiple datasets from different sequencing experiments. The integration of different scRNA-seq datasets could be challenging due to batch effect, a phenomenon that could occur when the experiments are run in different laboratories, at different time periods, or when using different instruments and technologies. Batch effect correction is a necessary process to prevent misleading results in downstream analysis on the integrated data. The challenge in scRNA-seq integration is mainly to merge the datasets while keeping the cell populations separate and maintaining the local structure of the datasets. We introduce SciTuna, a Single-Cell RNA-seq datasets Integration Tool Using Network Alignment with batch effect correction. Our method finds matching cells between the batches and uses an iterative approach to refine the integration of each cell based on the nearest neighboring cells. We show that our method outperforms other integration methods such as Seurat, Batman, and scAlign using simulated, semi-real, and real data based on different metrics. SciTuna also shows a reliable performance integrating datasets with semi-overlapping population compositions. Lastly, comparative differential expression analysis was carried out on the integrated datasets to demonstrate the batch effect correction and the robustness of the integration method.