ieCS: Interactive Explorer of Single Cell Cluster Similarity

Ling Hai[1], Matthias Schlesner[1]

[1]Bioinformatics and Omics Data Analytics, German Cancer Research Center (DKFZ), Heidelberg, Germany

l.hai@dkfz.de

GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION

Research for a Life without Cancer

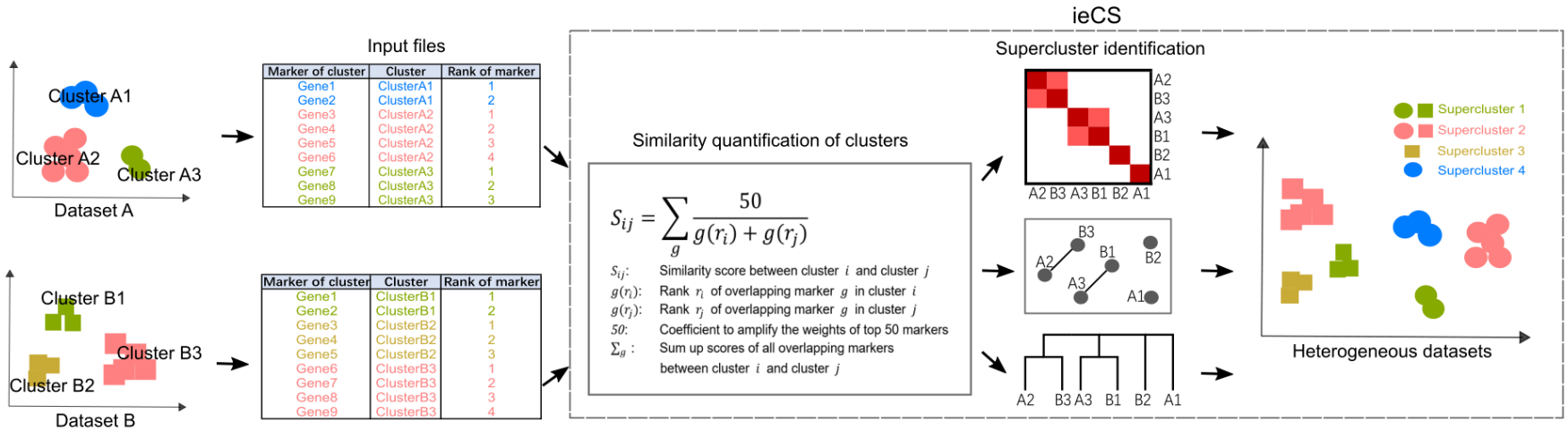
Motivation

Single-cell transcriptomics has a great potential to characterize disease-related cell subpopulations. However, it remains a challenge to compare cell subpopulation across multiple individuals and different conditions. We created the interactive explorer of single cell cluster similarity (ieCS), an R package with an interactive graphical user interface that aids to explore similarity of cell clusters among heterogeneous datasets.

Features

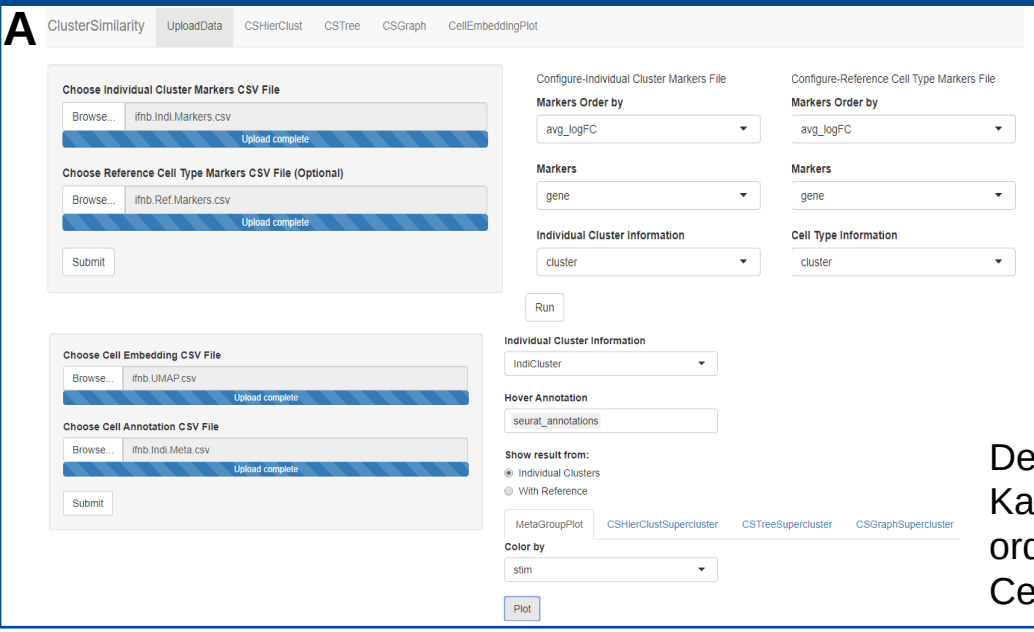
Shiny Application
Quantify similarity of cell clusters
Identify superclusters across datasets
Visualize dynamic supercluster composition interactively
Annotate cell clusters with custom reference

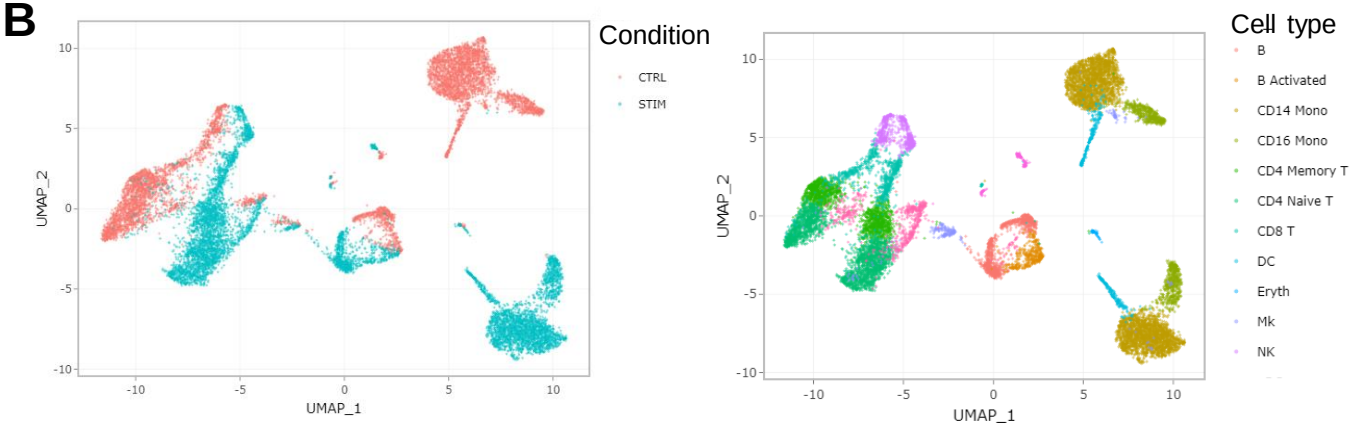
Workflow



ieCS requires cluster specific marker genes in individual datasets as input and quantifies similarity in pairwise clusters based on the ranks of overlapping marker genes. Based on the obtained similarity scores, ieCS then provides three methods to identify superclusters of similar cell clusters within a sample and also across different samples: hierarchical clustering of cell clusters, the Louvain community detection on a similarity network of cell clusters, and a greedy method to aggregate cell clusters into a tree structure. ieCS allows users to interactively explore superclusters and to visualize dynamic supercluster composition. In addition, ieCS accepts custom markers of cell type to annotate cell clusters.

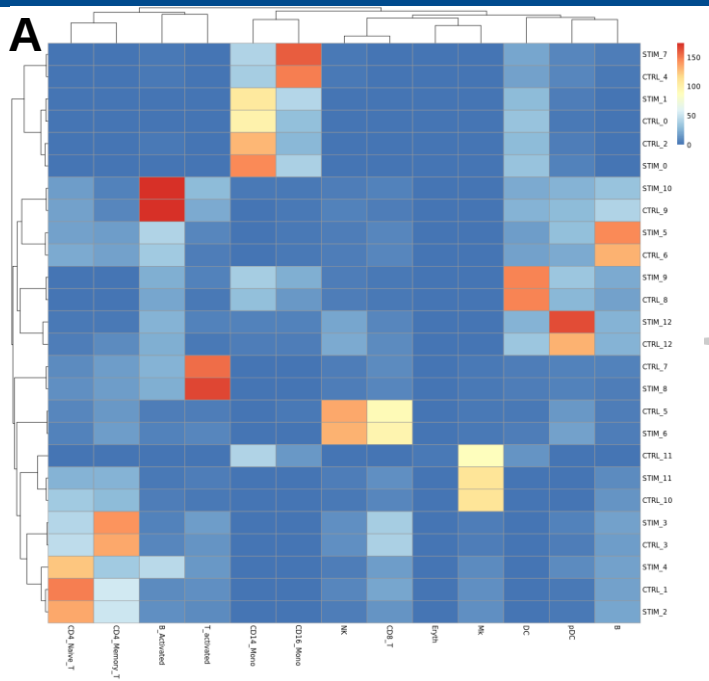
Demonstration dataset

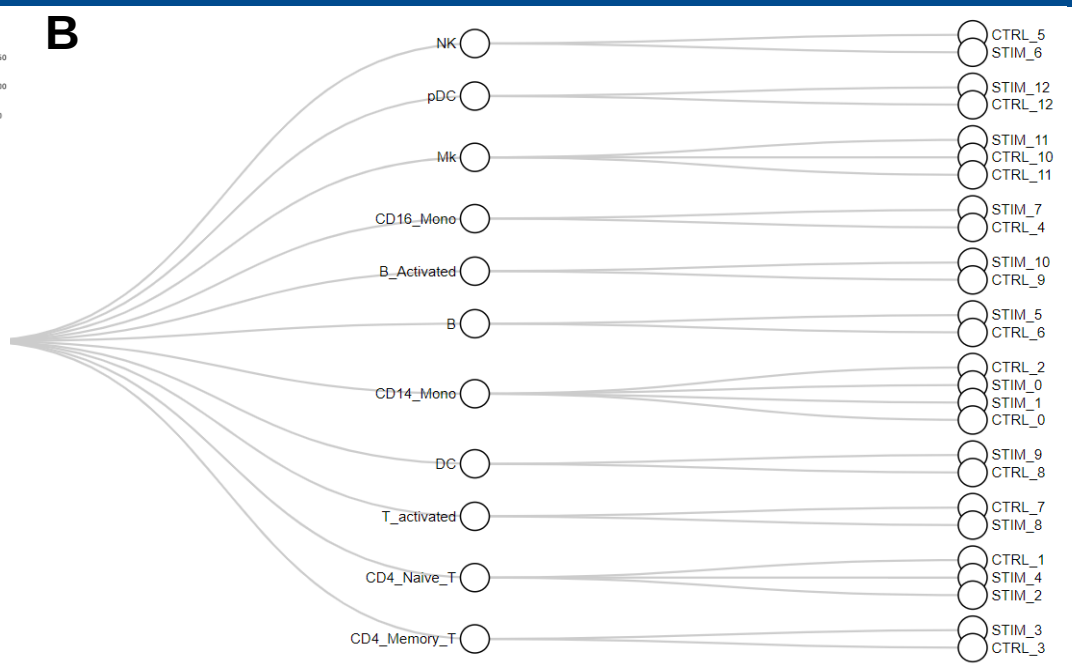




Demonstration with dataset containing two conditions, IFN-β stimulated and control, from Kang et al, Nature Biotech (2017), **A** Input panels of ieCS. Cluster-specific marker genes, the order of genes as well as cell embedding and annotation files for visualization are required. **B** Cells color by stimulation condition and cell types in UMAP.

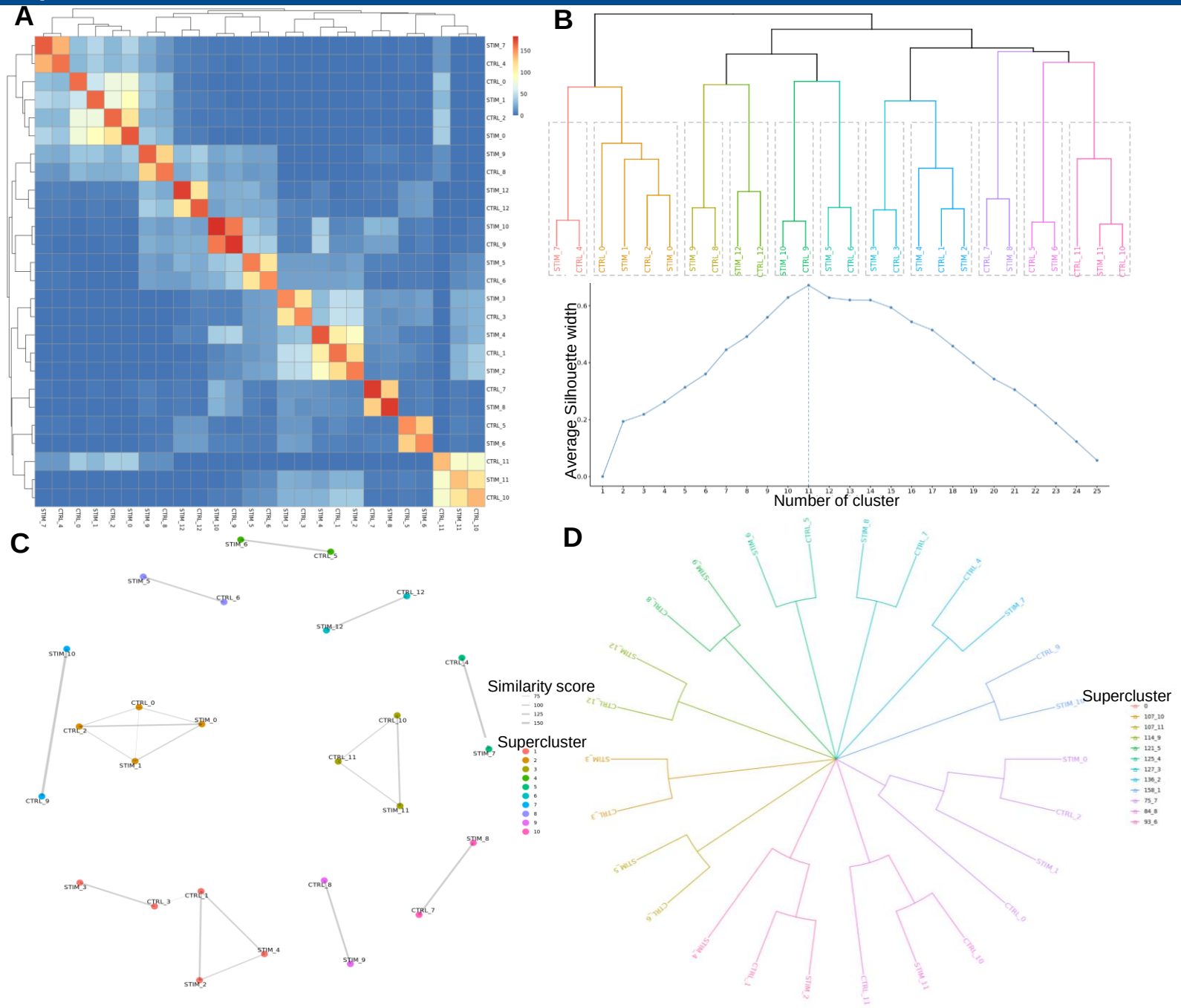
Cluster annotation with markers of cell types





ieCS accepts custom markers of cell types to annotate cell clusters. **A** Heatmap of cluster similarity between cell types and IFN-β dataset. **B** Assignment of IFN-β clusters to cell types based on similarity scores.

Supercluster identification



A Heatmap of pairwise similarity scores between clusters from both conditions. **B** Hierarchical clustering of pairwise similarity scores with optimal number of superclusters detected by Silhouette method. **C** Similarity network of pairwise similarity scores. Supercluster identified by Louvain community detection. **D** Aggregate tree of cell clusters. Greedy searching for most similar clusters to form superclusters based on gradient descent of similarity scores. **E** Cell embedding of complete dataset with colors indicating superclusters of aggregate tree. ieCS can accurately identify superclusters which match the cell types across conditions.