**Supplementary Material**

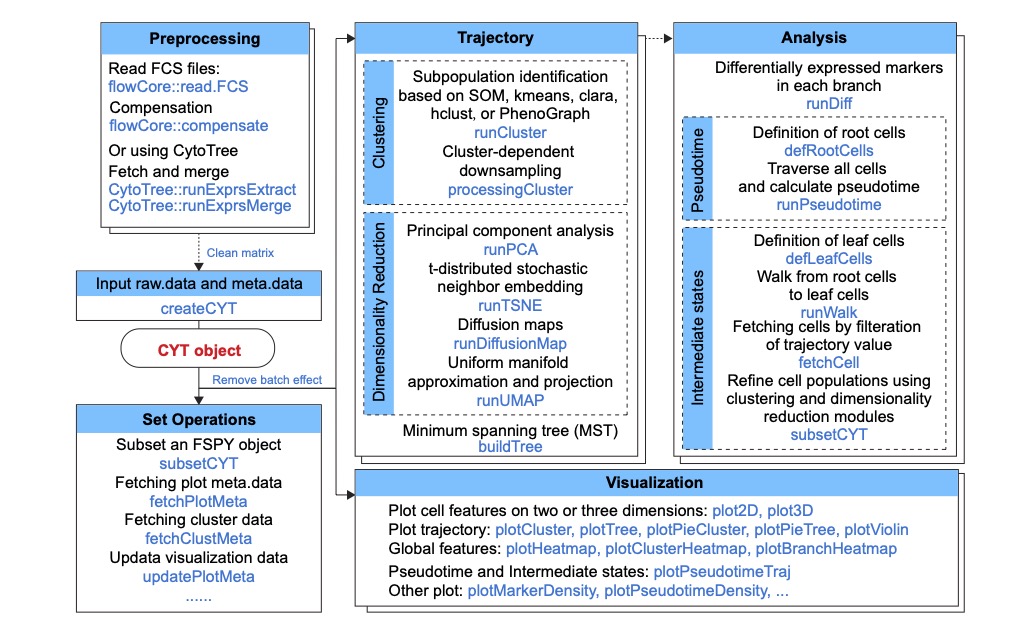
**CytoTree: an R/Bioconductor package for analysis and visualization of flow and mass cytometry data**

Yuting Dai1,†, Aining Xu1,†, Jianfeng Li1, Liang Wu1, Shanhe Yu1, Jun Chen2, Weili Zhao1,\*, Xiao-Jian Sun1,\* and Jinyan Huang1,\*

1. Shanghai Institute of Hematology, State Key Laboratory of Medical Genomics, National Research Center for Translational Medicine at Shanghai, Ruijin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 197 Ruijin Er Road, Shanghai 200025, China.
2. Division of Biomedical Statistics and Informatics, Department of Health Sciences Research and Center for Individualized Medicine, Mayo Clinic, 200 1st St SW, Rochester, MN 55905, USA

﻿† Equal contribution

\* Corresponding authors: E-mail: huangjy@sjtu.edu.cn (Jinyan H), xjsun@sibs.ac.cn (Xiao-Jian S), zhao.weili@yahoo.com (Weili Z).

****

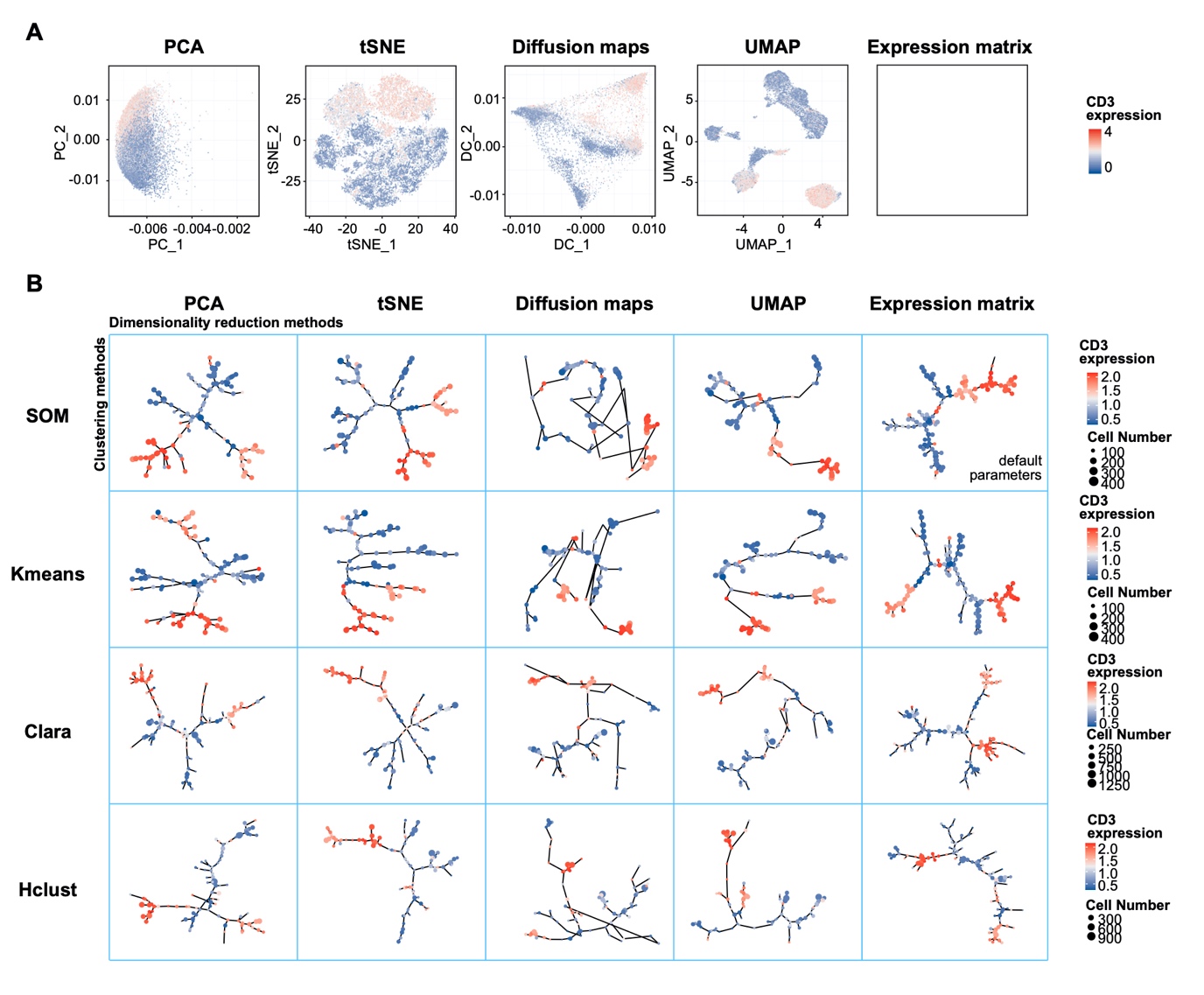
**Figure S1. Structure of CytoTree package**

The table headers describe the available modules: preprocessing, trajectory, analysis, visualization and set operations. A short description (black font) and the corresponding function (blue font) are provided for each module. The CytoTree workflow begins with the reading of the FCS data. Compensation, filtration, concatenation and normalization are included in the preprocessing module. A clean matrix after preprocessing is required to build an CYT object, and the analysis workflows of all other modules are all based on the CYT object. The trajectory module contains functions used to perform clustering and dimensionality reduction for cells. The analysis module is based on calculation results from the trajectory module. The visualization module includes functions to generate publication-quality plots from the CYT object. The set operations module includes a function for subsetting an CYT object based on user-defined cells or fetching meta information for clusters and cells during the analysis.

****

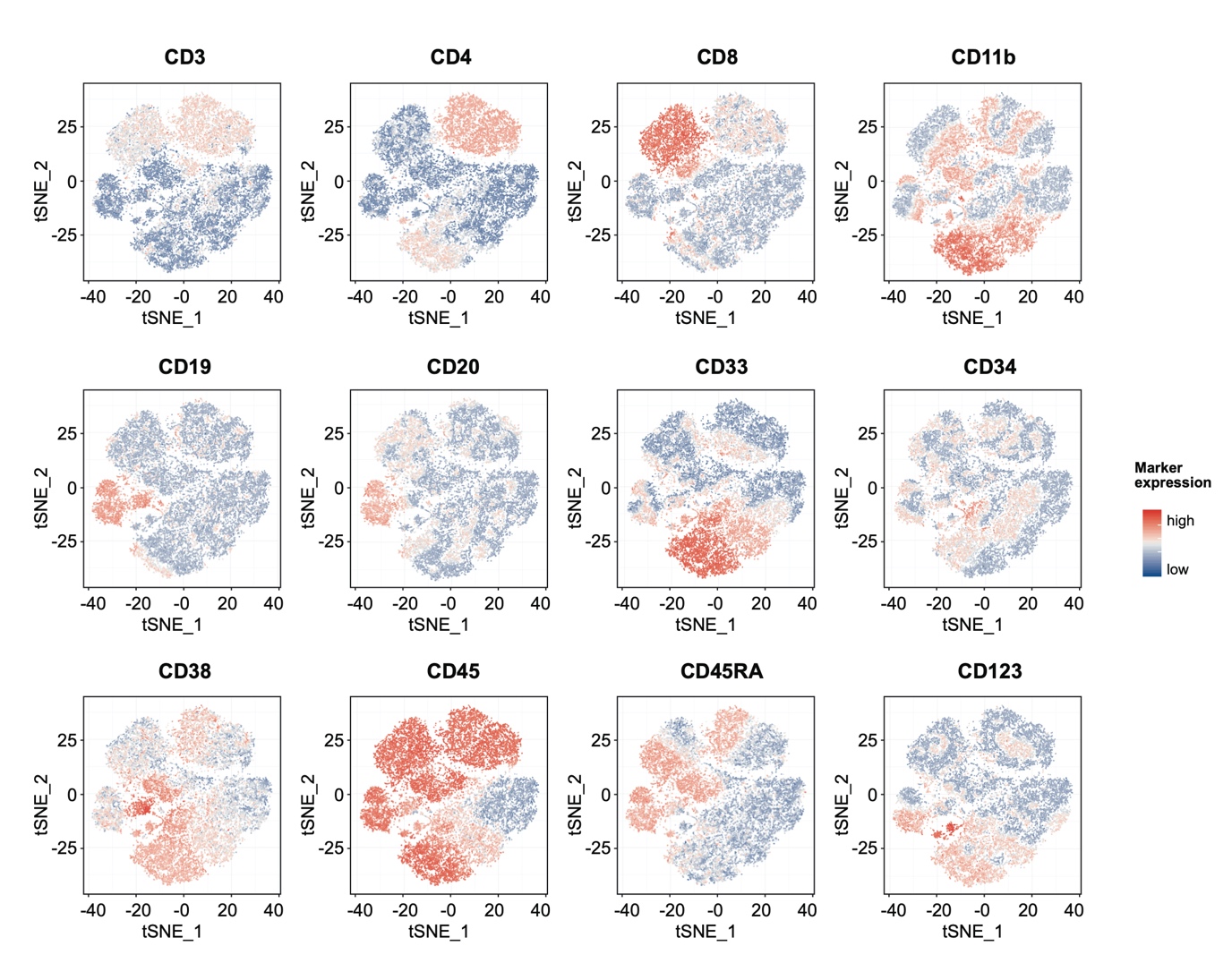
**Figure S2. Comparison of scatter plot of FlowJo and CytoTree**

**(A)** Scatter plot when opened raw FCS file using FlowJo. **(B)** Scatter plot when opened raw FCS file using CytoTree. **(C)** Scatter plot using FlowJo after manual gating. **(D)** Scatter plot using CytoTree after manual gating using CytoTree.

****

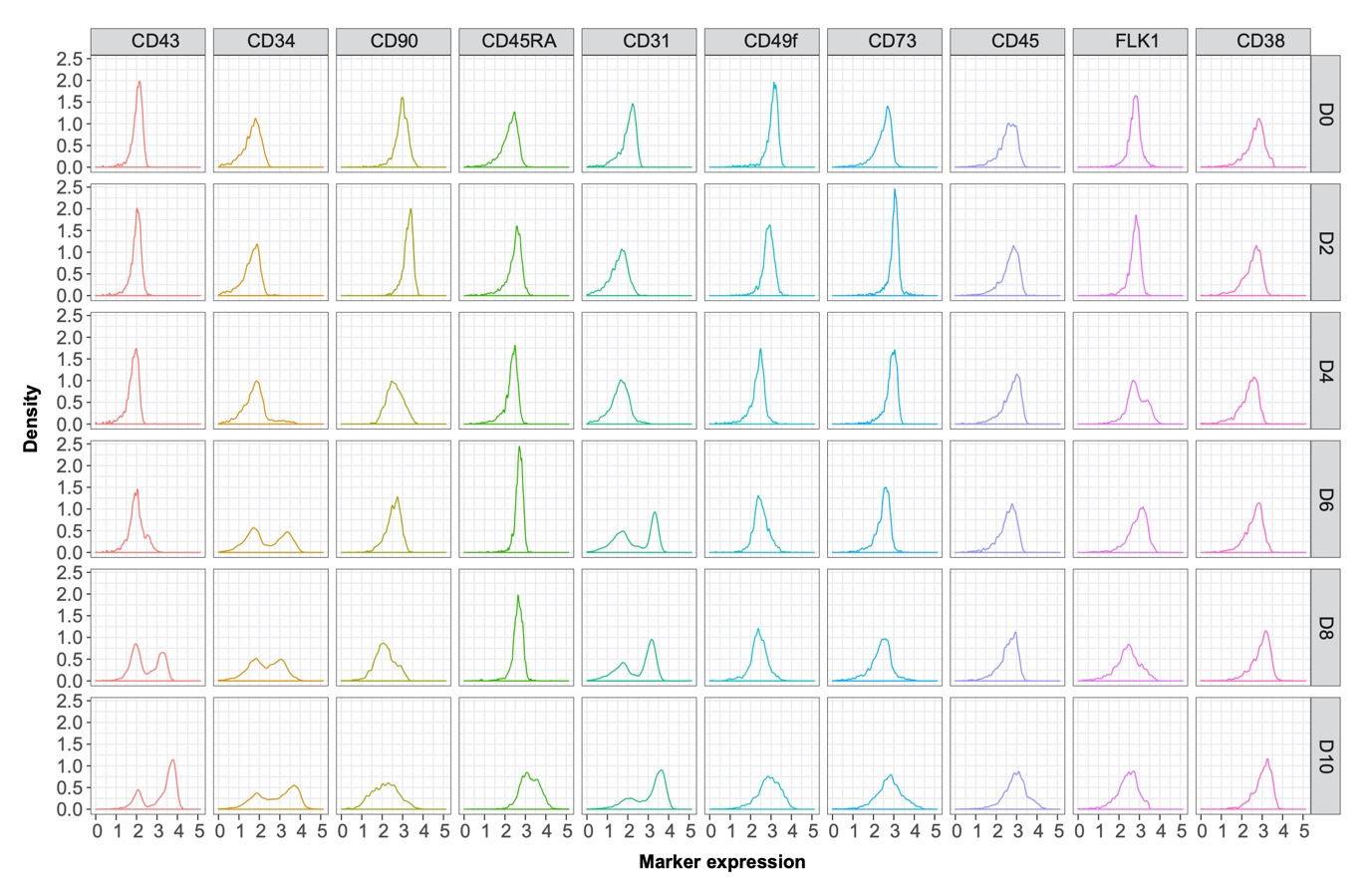
**Figure S3. Visualization of trajectory constructed based on different clustering methods and dimensionality reduction methods**

**(A)** The visuality of the four dimensionality reduction methods using the 13-panel mass cytometry data, namely PCA, tSNE, diffusion maps and UMAP. PC\_1 and PC\_2 represented the first two principal components calculated by PCA, tSNE\_1 and tSNE\_2 represented the two components calculated by tSNE, DC\_1 and DC\_2 represented the first two diffusion components calculated by diffusion maps, and UMAP\_1 and UMAP\_2 represented the two components calculated by UMAP. Color in each point revealed the expression level of CD3. Red color represented a higher expression. **(B)** The trajectory tree constructed using MST based on PCA, tSNE, diffusion maps and UMAP coordinates, and the raw expression matrix, separately. Cluster color was scaled to the mean intensity of the marker expression of the cells within each cluster. Cluster size was scaled based on the cell number of each cluster.

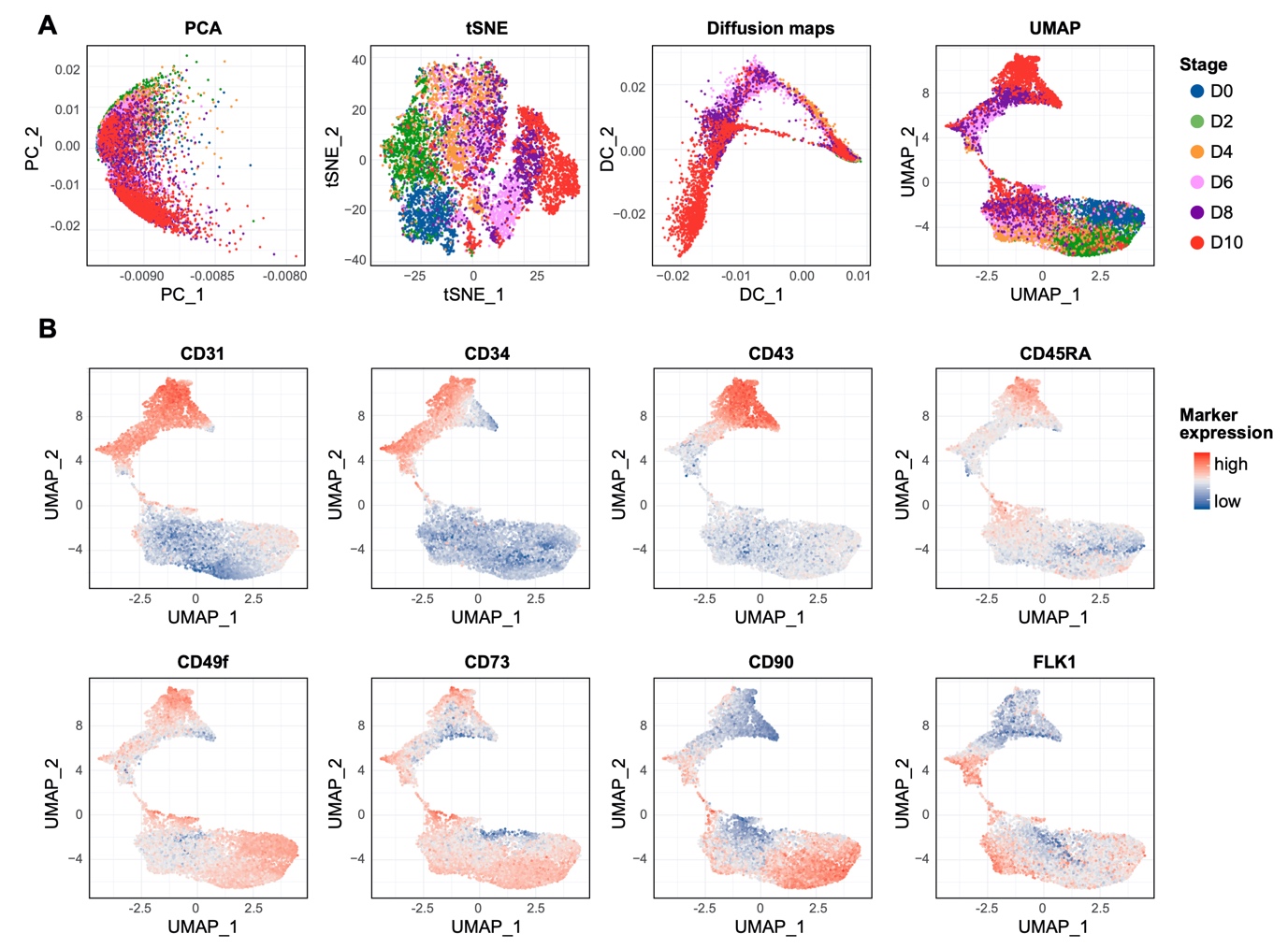
****

**Figure S4. tSNE visualization of marker expression of the 13-panel mass cytometry data**

tSNE plot of the expression level of CD3, CD4, CD8, CD11b, CD19, CD20, CD33, CD34, CD38, CD45, CD45RA and CD123. Color in each point revealed the expression level of each marker. Red color represented a higher expression.

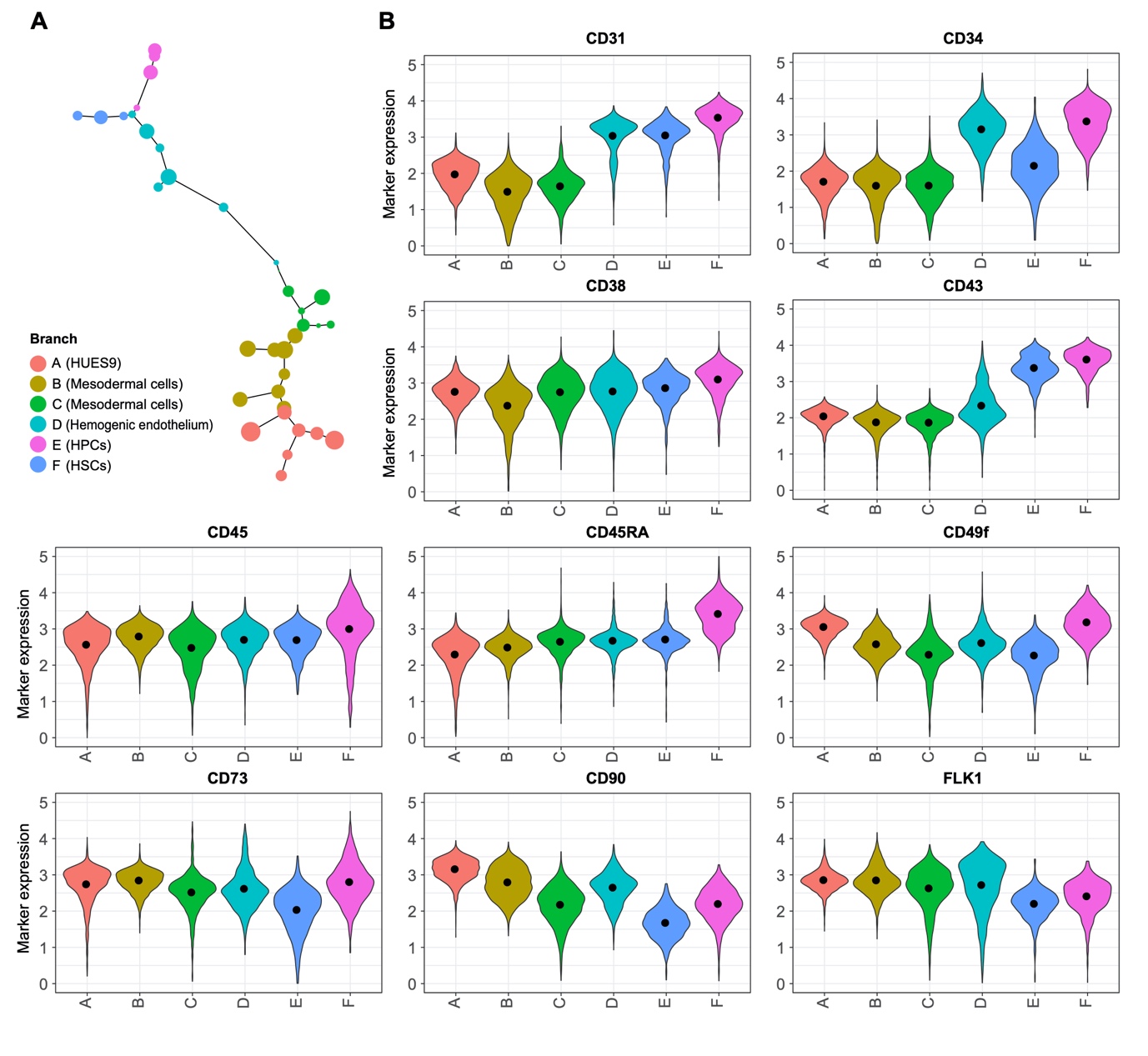
****

**Figure S5. Marker density plot of time-course flow cytometry data**

****

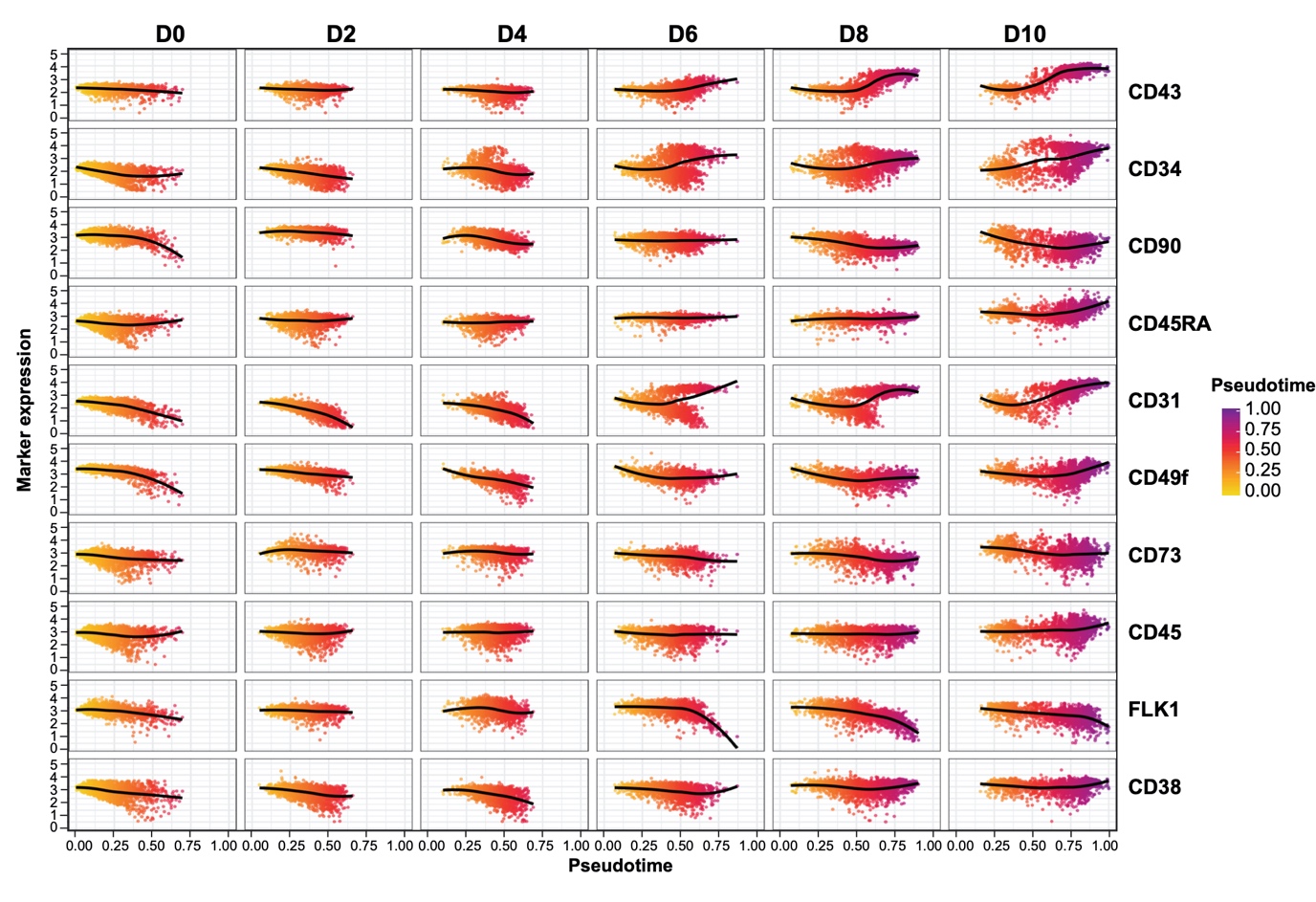
**Figure S6. Dimensionality reduction and UMAP visualization of marker expression of the time-course flow cytometry data**

**(A)** The visuality of the first two components calculated by the four dimensionality reduction methods using time-course flow cytometry data: PCA, tSNE, diffusion maps and UMAP. Cells were colored by time point from D0 to D10, separately. **(B)** UMAP plot of the expression level of CD31, CD34, CD43, CD45RA, CD49f, CD73, CD90 and FLK1.

****

**Figure S7. Branch analysis of differentiation trajectory using the time-course flow cytometry data**

**(A)** Branches identified by CytoTree and putative populations were annotated manually according to marker expression. **(B)** Violin plot of the expression level of CD31, CD34, CD38, CD43, CD45, CD45RA, CD49f, CD73, CD90 and FLK1. Each branch was colored different.

****

**Figure S8. Correlation between markers expression and pseudotime in different time points**