Nested Stochastic Block Models Applied to the Analysis of Single Cell Data

Leonardo Morelli[1,2], Valentina Giansanti[1,3], and Davide Cittaro[1]

[1]Center for Omics Sciences, IRCCS San Raffaele Institute, Milan, Italy
[2]Università Vita-Salute San Raffaele, Milan, Italy
[3]Department of Informatics, Systems and Communication, University of Milano-Bicocca, Milan, Italy

April 15, 2021

Supplementary figures

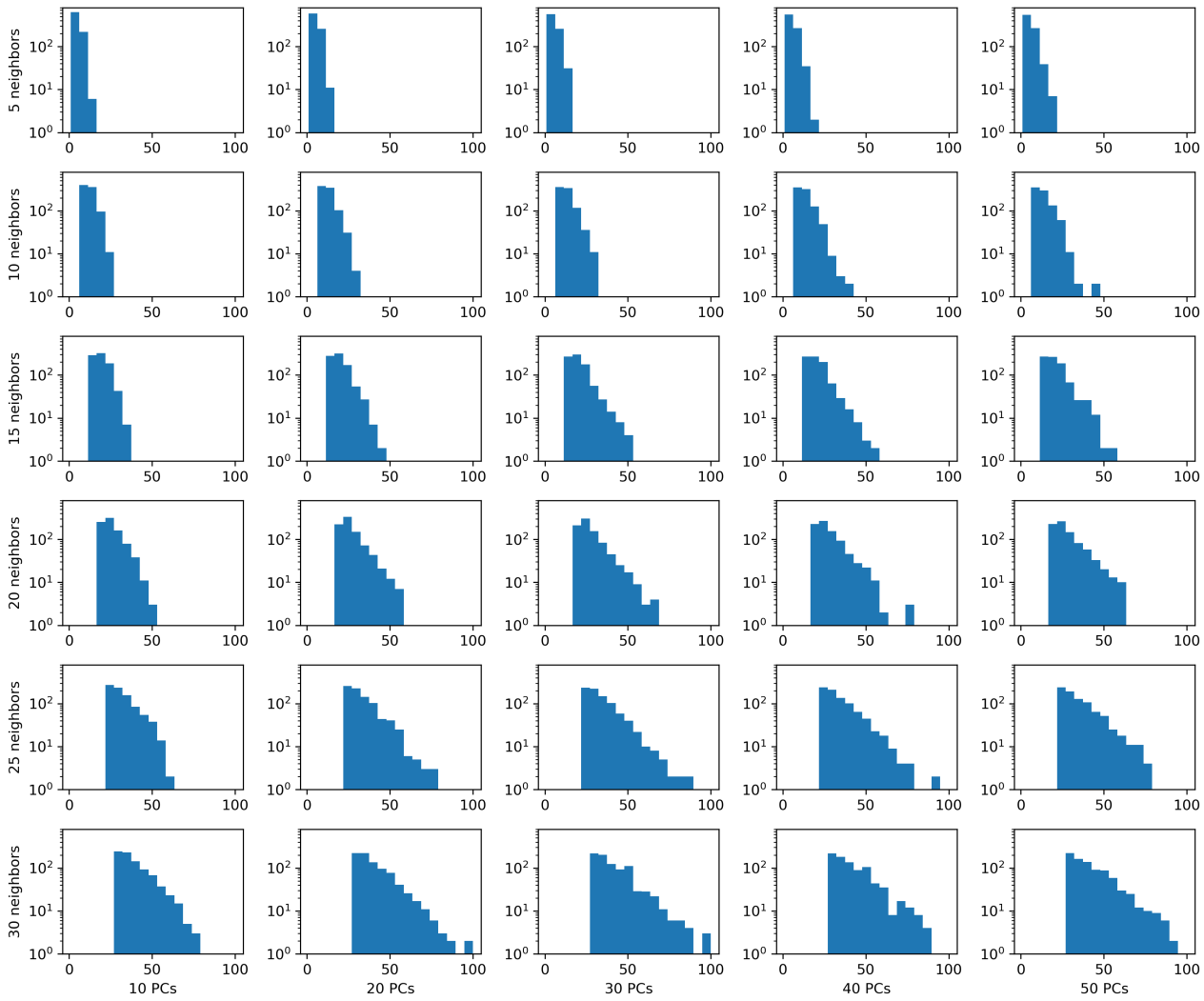

FIGURE S1: Degree distribution of multiple kNN graphs derived from scRNA-seq mixology datasets using variable number of Principal Components or number of neighbors. Each histogram shows the number of nodes (on y axis) within a specific degree bin (on x axis). Both the parameters influence the sparseness of the graph.

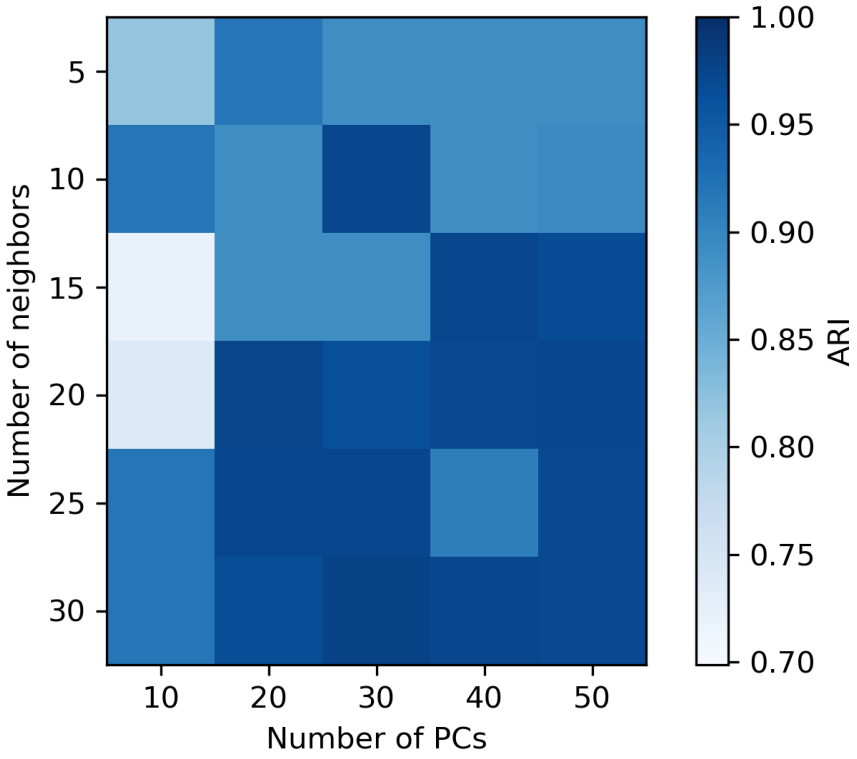

FIGURE S2: Adjusted Rand Index for different kNN graphs after MCMC run. Maximal ARI over all hierarchy level is shown. Darker color indicates higher concordance with the ground truth

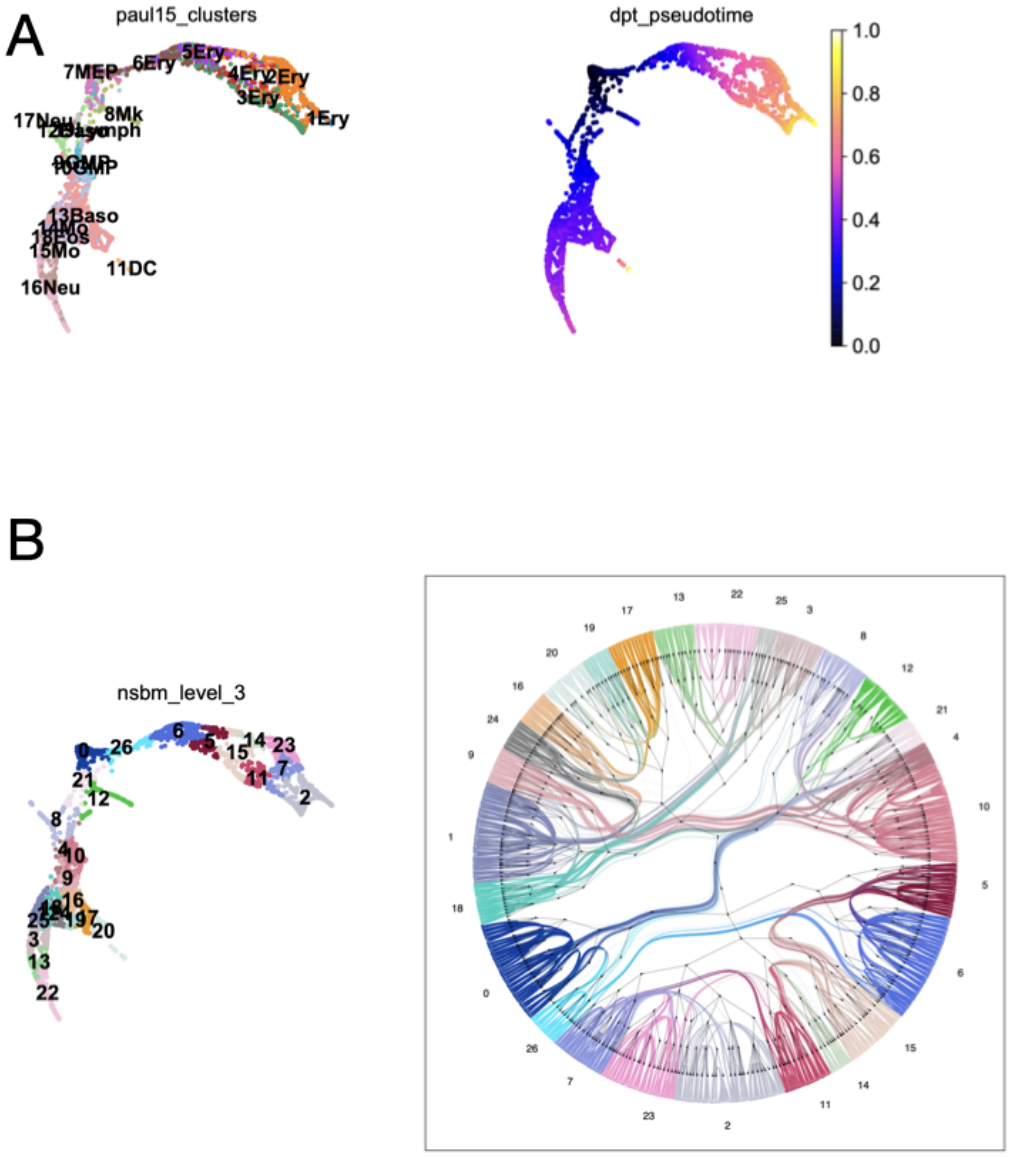

FIGURE S3: Analysis of hematopoietic differentiation. (A) Low dimensional embedding of single cells colored by original cell type and pseudotime. (B) Cells are colored according to the nSBM grouping at level 3 of the hierarchy, next to a radial tree representation of the same model.

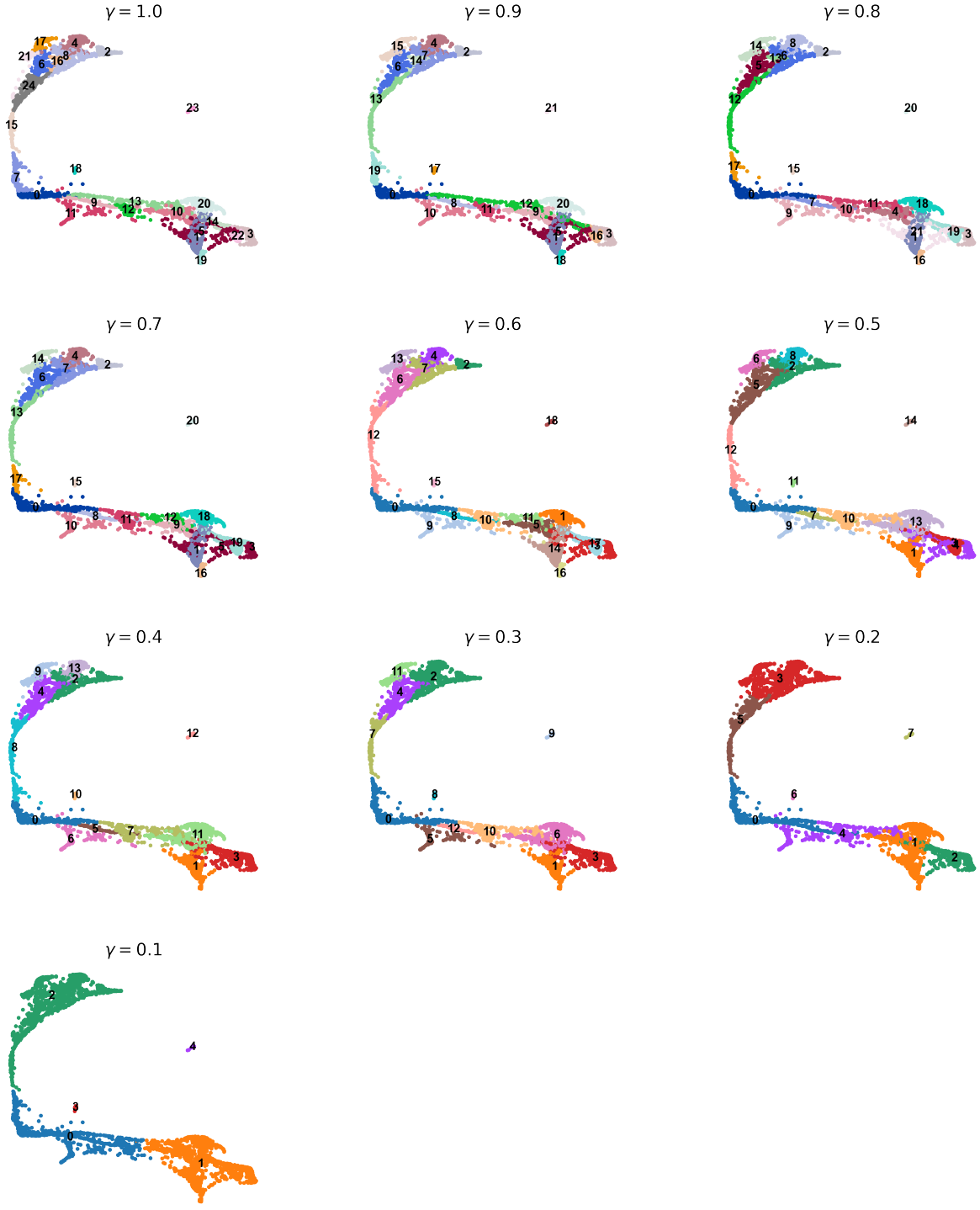

FIGURE S4: Low dimension embedding of single cells for hematopoietic differentiation colored according to Leiden clustering at decreasing resolution, from 1.0 to 0.1. Lowering the distribution does not grant that cells are grouped in a hierarchical way, *e.g.* in the Erythroid branch groups 7 and 15 at resolution r=1 are merged or split at coarser resolutions ($\gamma = 0.6$ and $\gamma = 0.3$)

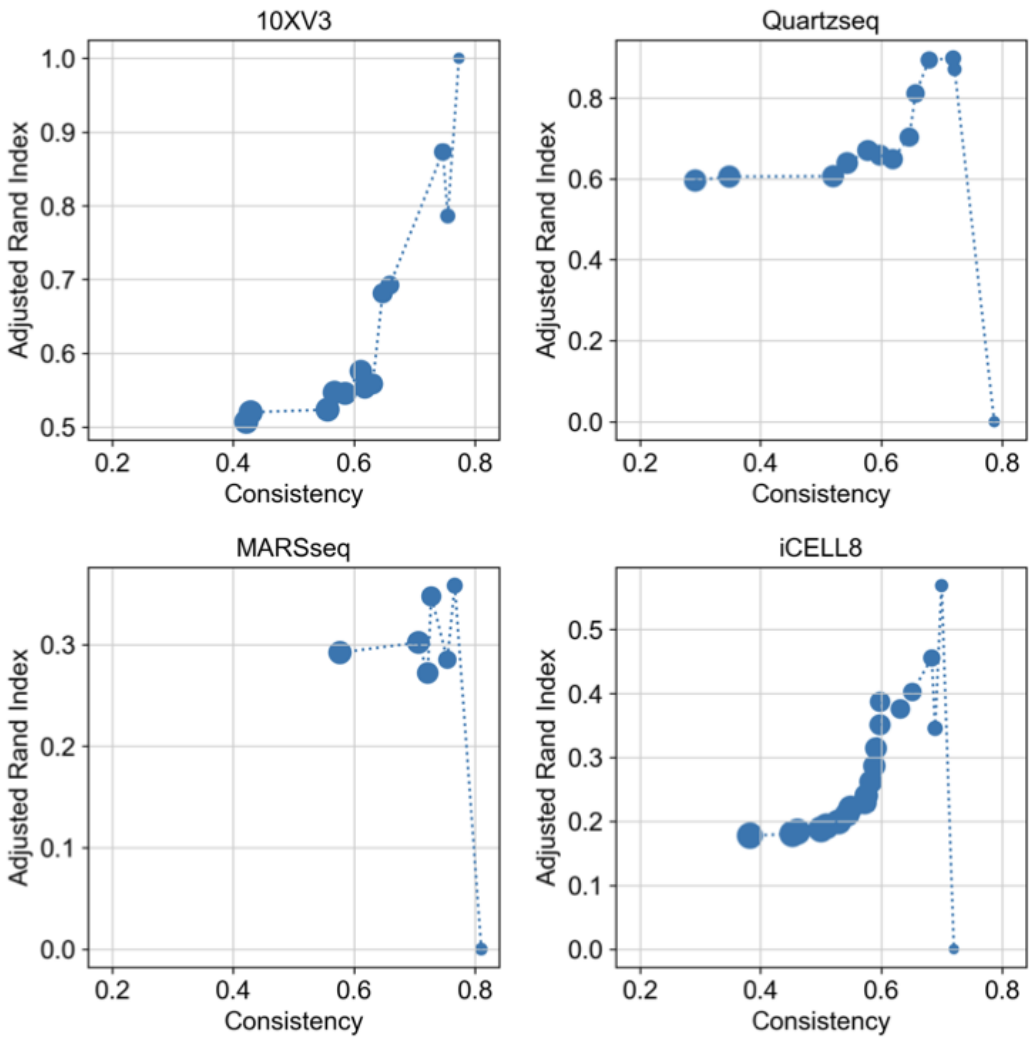

FIGURE S5: Adjusted Rand Index between cell clusters and cell type annotation filtering data at different cutoffs of consistency. Dot size is proportional to the number of cells remaining after filtering.