# 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
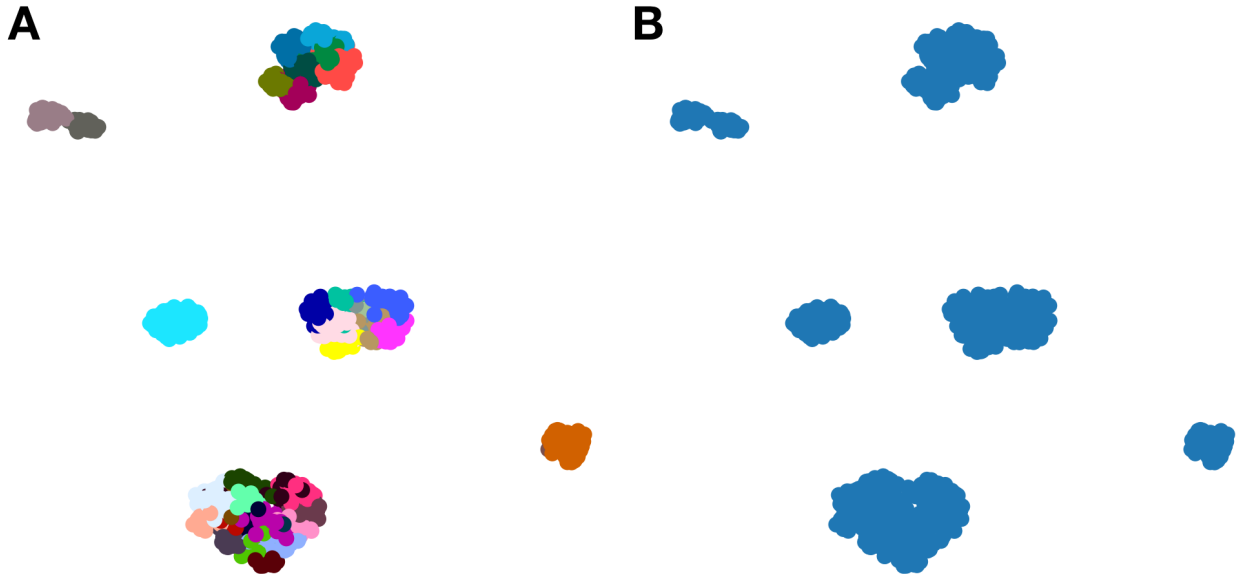
April 22, 2021

# Supplementary figures



FIGURE S1: Analysis of a mixture of 5 known cell lines. (A) UMAP embedding of cell lines from the sc-mixology experiment colored by the level 0 of the hierarchy proposed by the nested Stochastic Block Model (B) UMAP embedding colored by the classification made by SCCAF when partitions in (A) are used.
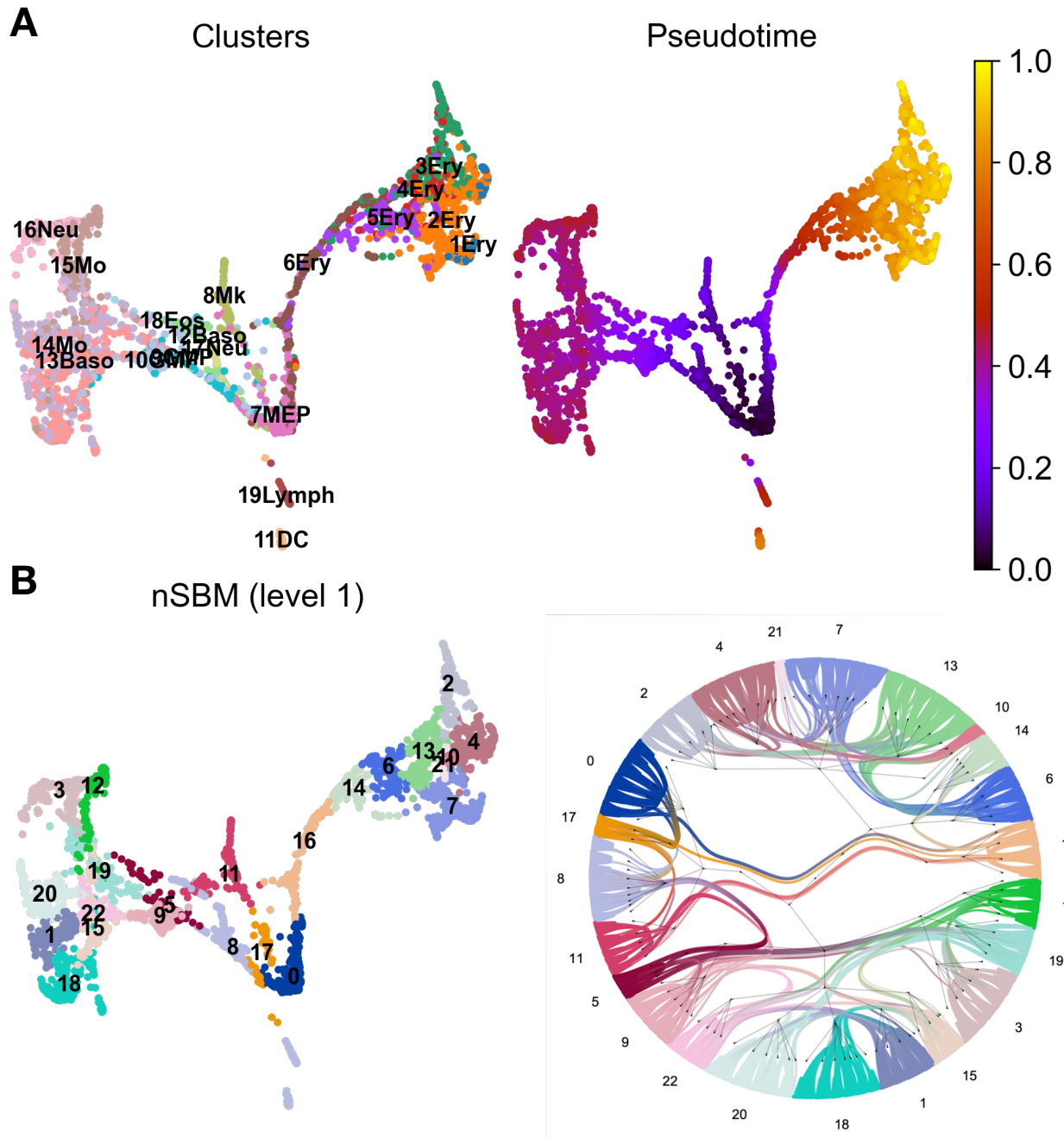
FIGURE S2: Analysis of hematopoietic differentiation. (A) Low dimensional embedding of single cells colored by original cell type and pseudotime. (B) Cells are colored according to the nSBM grouping at level 3 of the hierarchy, next to a radial tree representation of the same model.
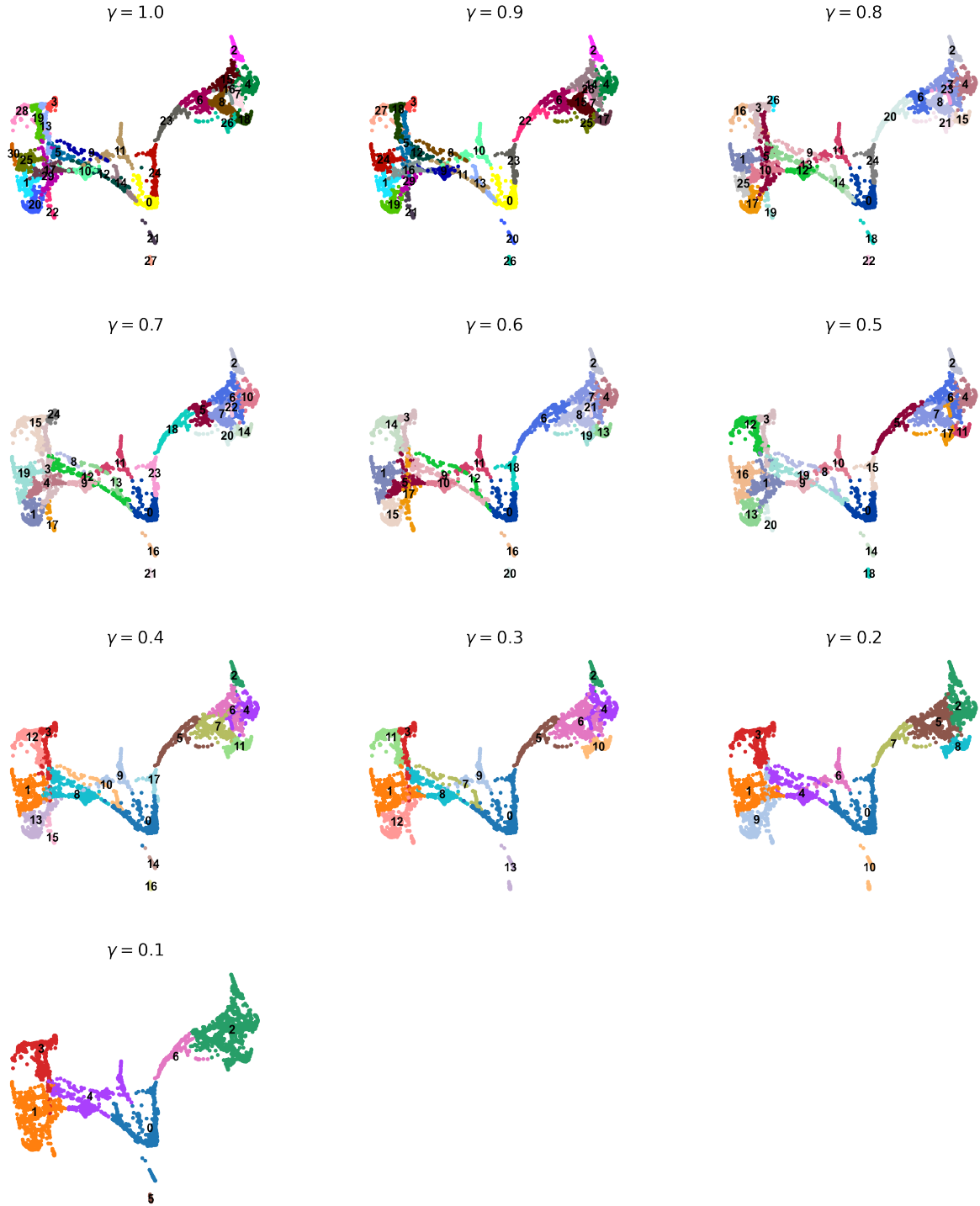
FIGURE S3: Low dimension embedding of single cells for hematopoietic differentiation colored according to Leiden clustering at decreasing resolution, from 1.0 to 0.1. Lowering the distribution does not grant that cells are grouped in a hierarchical way, *e.g.* groups 6 and 8 in the Erythroid branch at resolution $\gamma = 1$ are merged or split at coarser resolutions ($\gamma = 0.6$ and $\gamma = 0.3$)
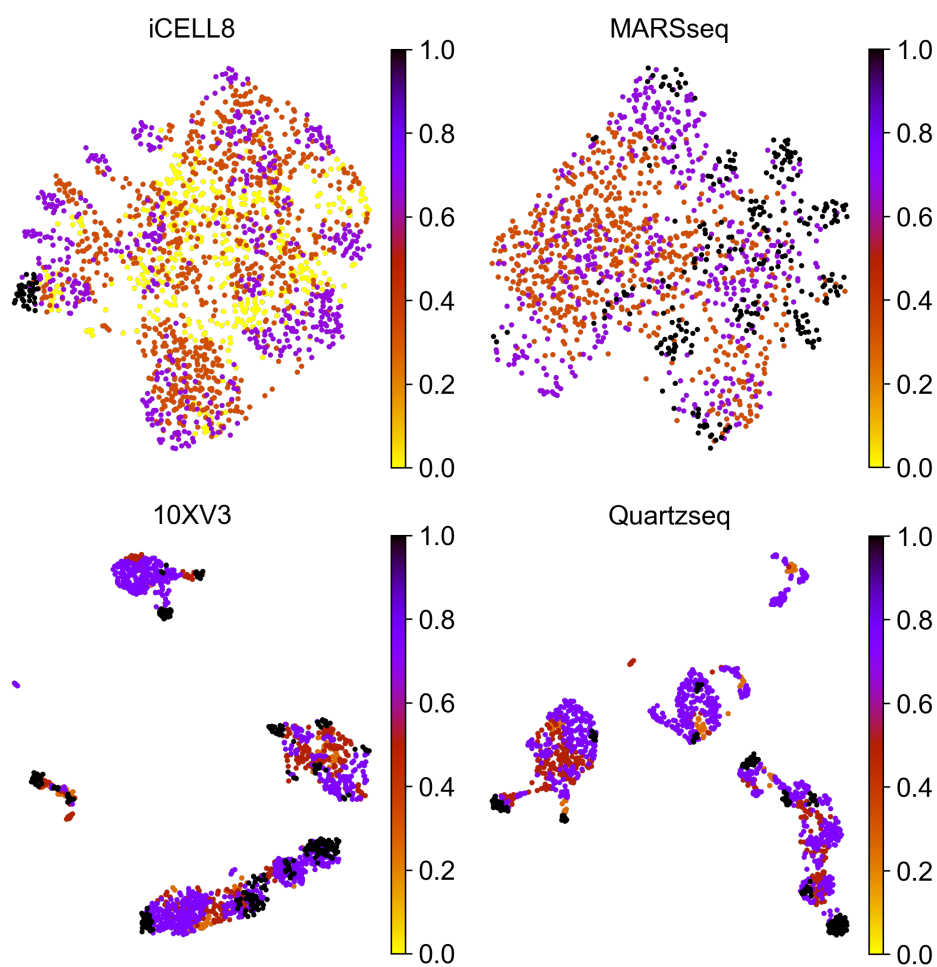
FIGURE S4: Cell Stability. UMAP embeddings of PBMC data profiled with four different technologies ranked by their quality. Cells are colored by the Cell Stability metric.)
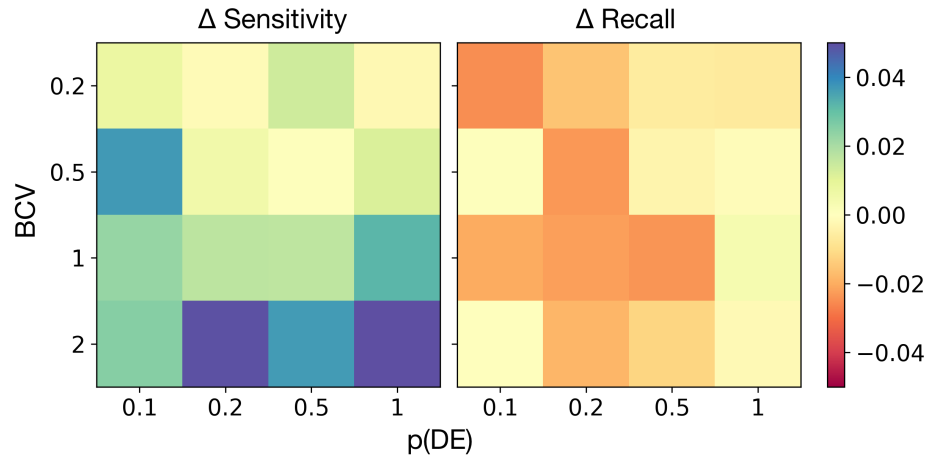
FIGURE S5: Results of differential gene expression with or without marginals. The heatmaps show the difference in sensitivity (left) or recall (right) of statistical testing for differential genes among cell groups when marginals are included or not in the model. Tests have been conducted on simulated data with different levels of noise (BCV) or probability of differential expression (p(DE))

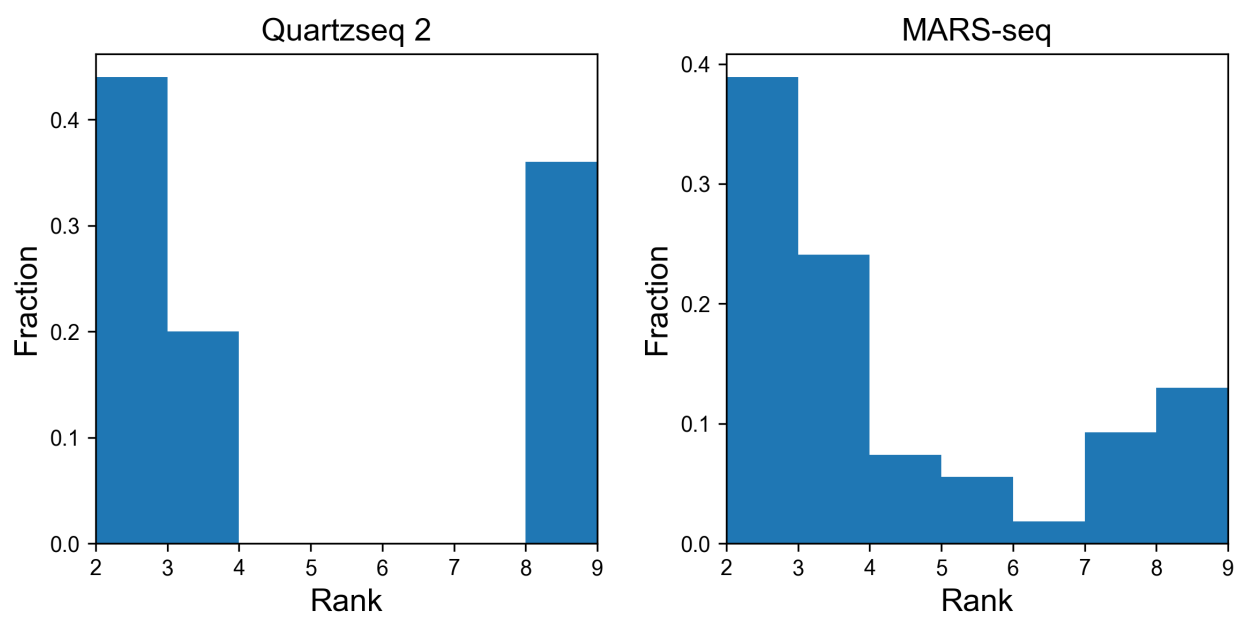FIGURE S6: TO BE WRITTEN