nature portfolio

https://doi.org/10.1038/s41592-021-01282-5

Supplementary information

Single-cell chromatin state analysis with Signac

In the format provided by the authors and unedited

Supplementary Figures

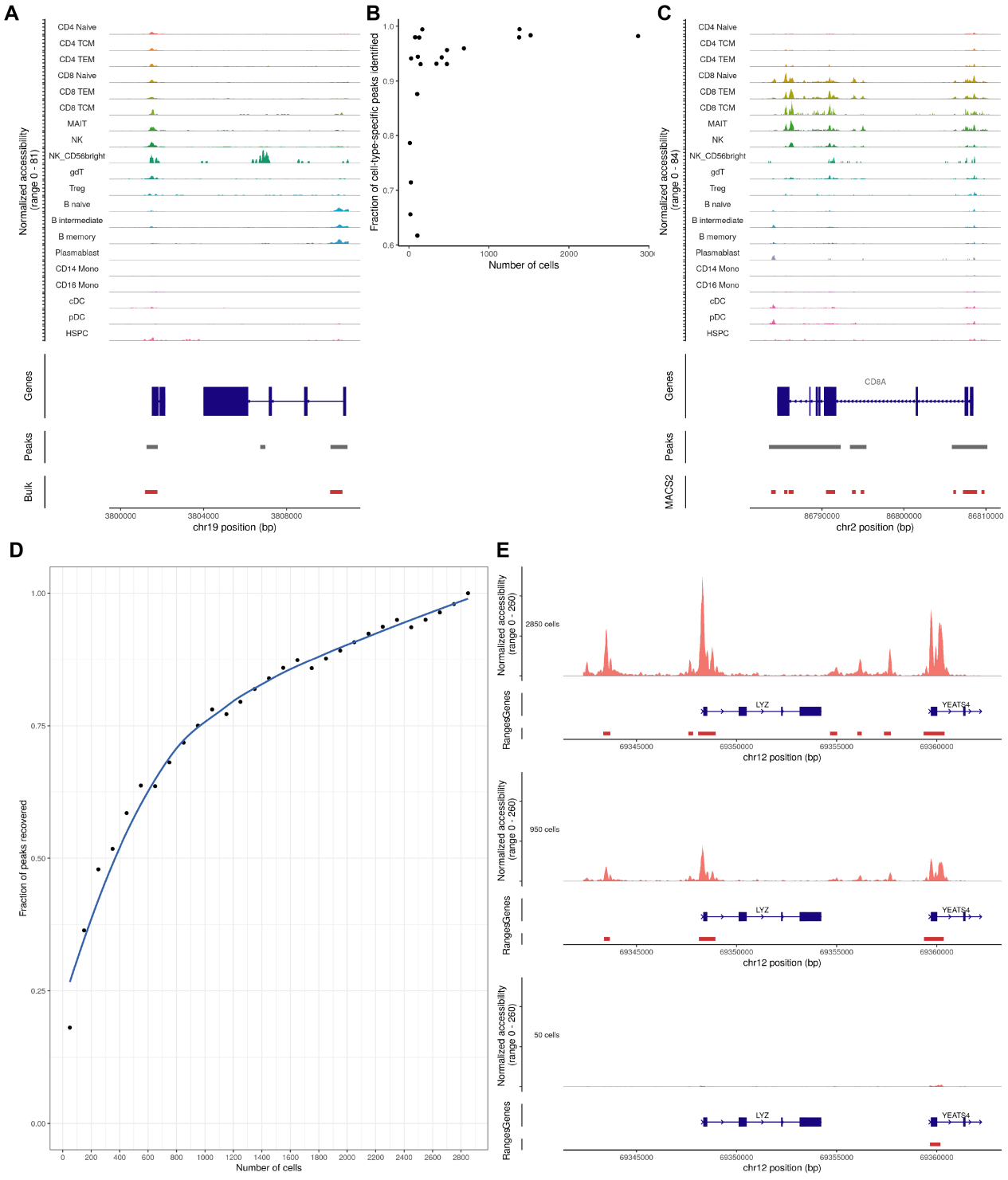

Figure S1. Comparison of peak calling methods for single-cell DNA accessibility data.
(A) Comparison of cell-type-specific peak calling (grey) and bulk peak calling (red) using MACS2. (B) Fraction of cell-type-specific peaks for each cell type that were recovered when calling peaks on the bulk cell population. (C) Comparison of MACS2 (red) and Cellranger (grey) peak calls. (D) Fraction of peaks identified when using different numbers of input cells. Peak calls for the 2,850-cell dataset were used as a baseline comparison group. (E) Representative genome browser views of DNA accessibility for different cell downsamplings, with associated MACS2 peak calls shown.

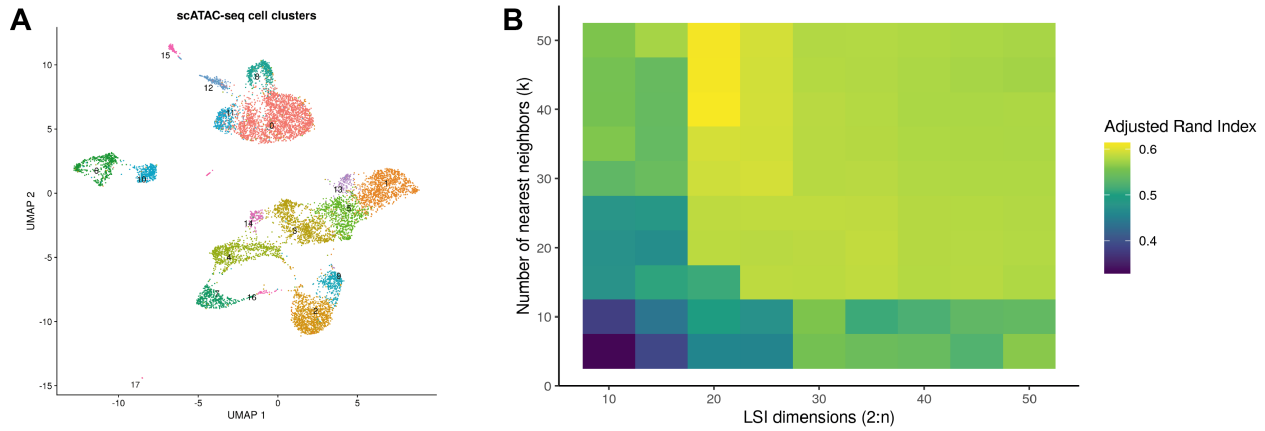

Figure S2. Clustering of single-cell DNA accessibility data.
(A) Two-dimensional UMAP visualization constructed from the PBMC multiomic DNA assay, with cells colored by cluster identity. (B) Effect of parameter choice on cell clustering results. Number of nearest neighbors (k) was varied from 5 to 50, and the maximum LSI component varied from 10 to 50 when constructing a nearest-neighbor graph for clustering. For each clustering result, the adjusted Rand index was computed using the cell type annotations derived from the RNA assay.

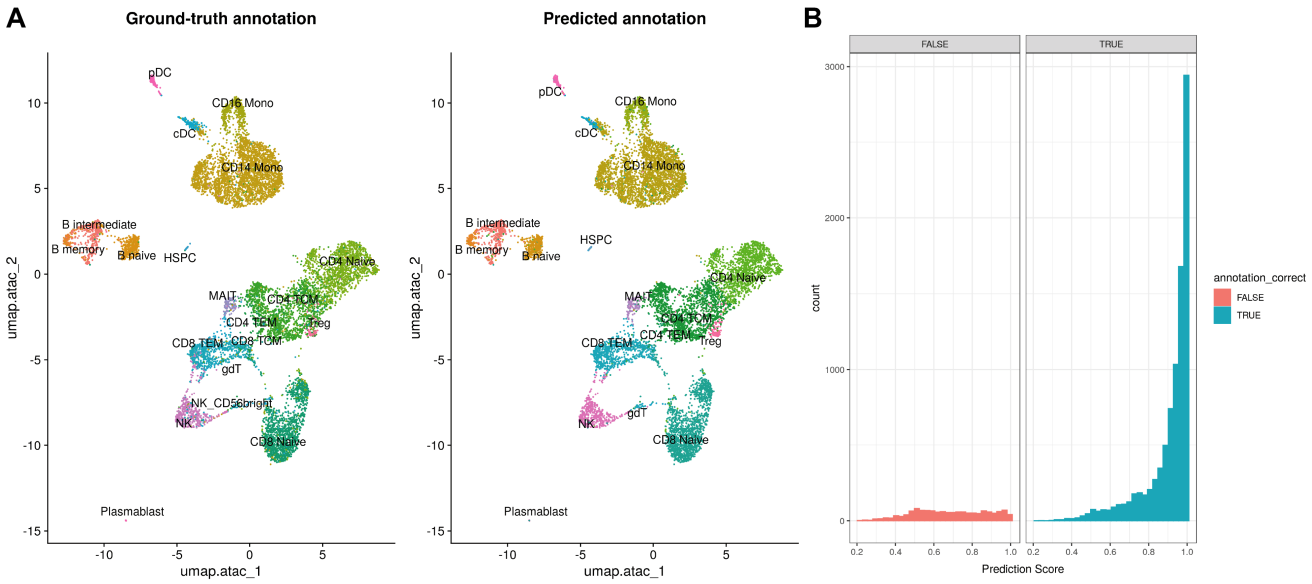

Figure S3. Multimodal label transfer accuracy
(A) UMAP representation of cells colored by ground-truth cell type annotation determined by the gene expression assay (left) and colored by the predicted cell type label via multimodal label transfer. (B) Distribution of label transfer prediction scores for correct and incorrect predicted cell labels.

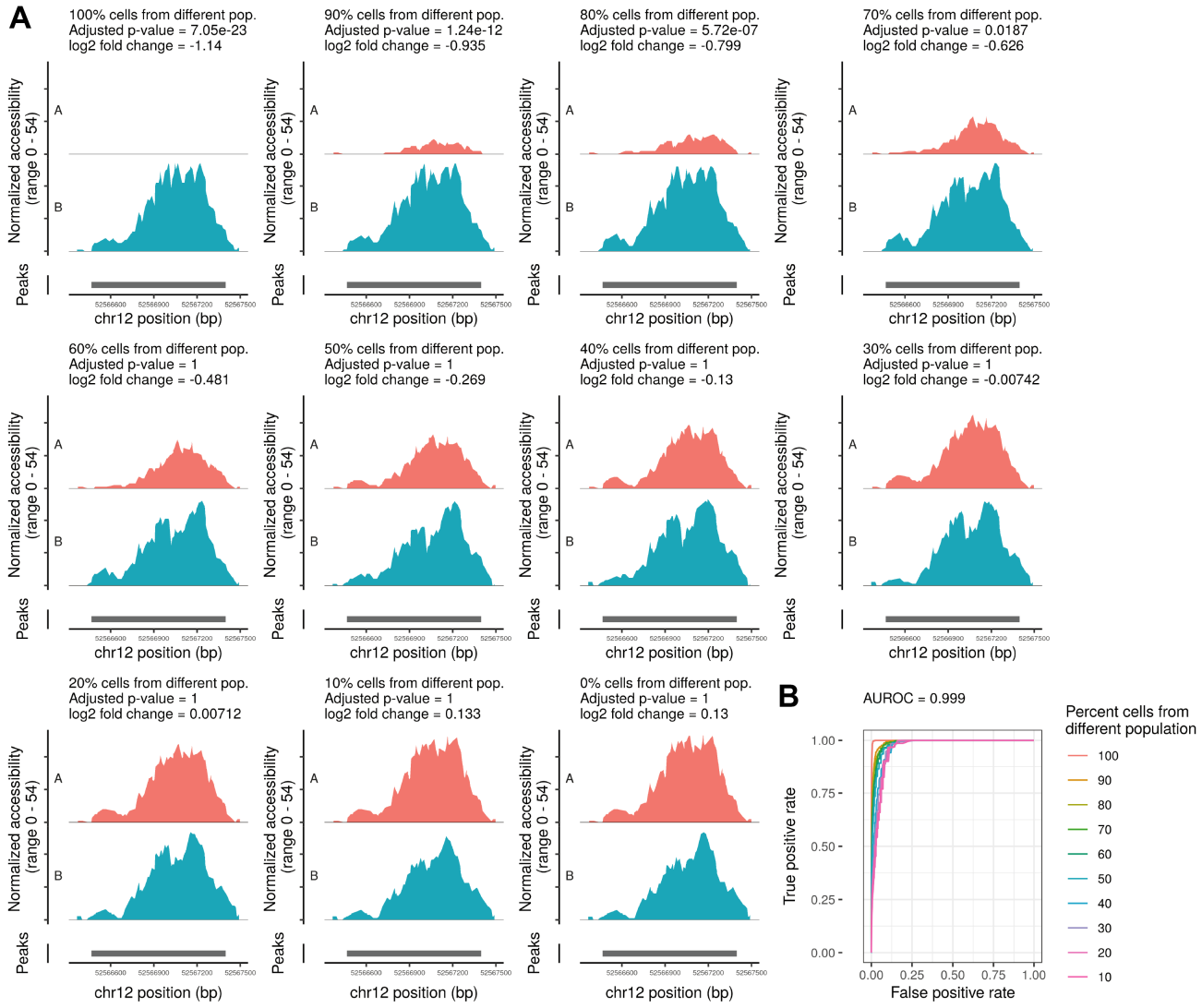

Figure S4. Differential accessibility sensitivity analysis.
(A) Fold-change and adjusted p-values for one representative peak. P-values were determined by a likelihood ratio test, and corrected for multiple testing using Bonferroni correction. (B) Receiver-operator characteristic (ROC) curves for different cell mixture populations. True-positive differentially accessible peaks were defined as peaks with a significant difference (adjusted p-value <0.01 and absolute log2 fold-change >0.4) when comparing the full CD4+ TCM and NK cell populations.

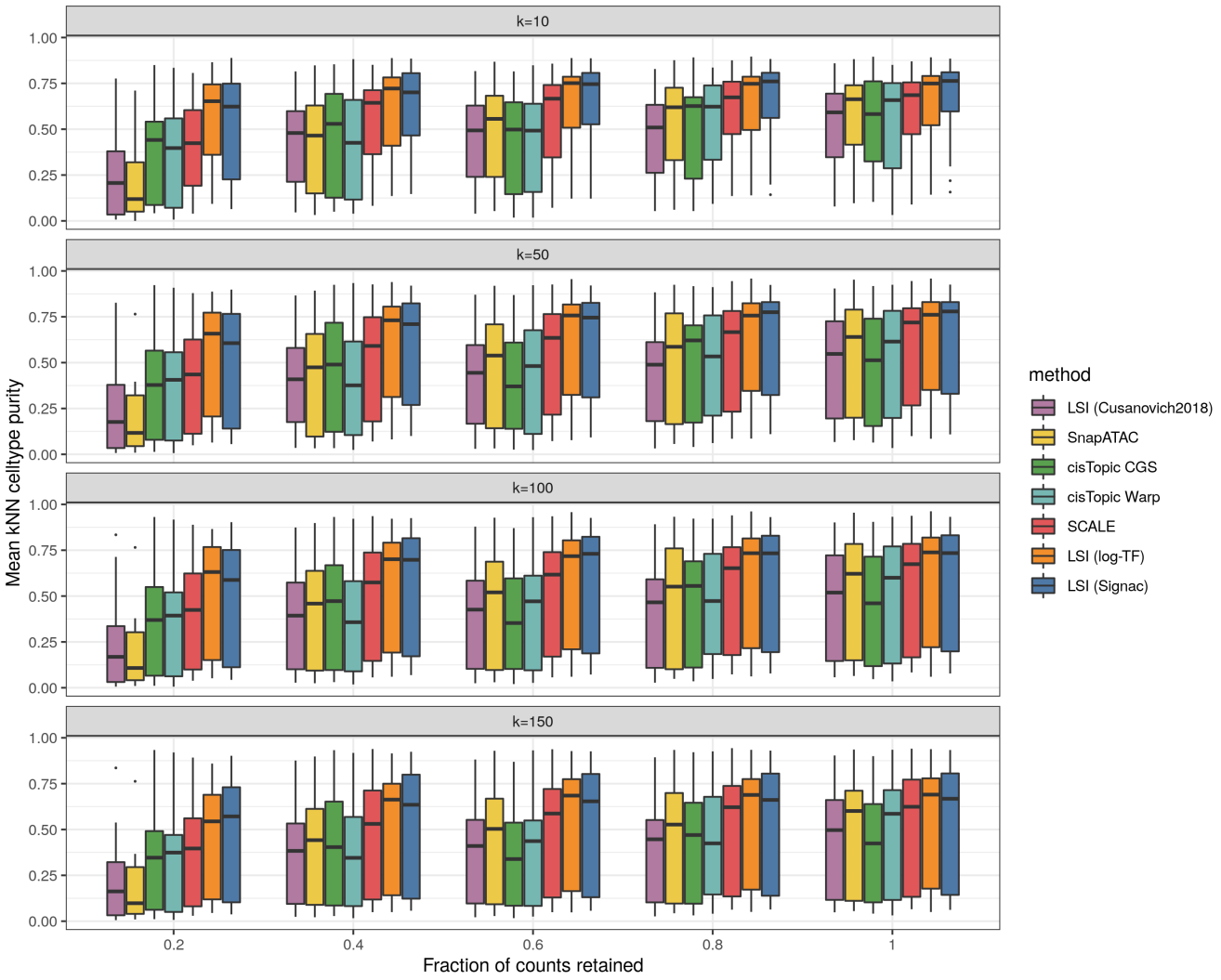

Figure S5. KNN purity metric for scATAC-seq dimension reduction.
Mean k-nearest neighbor celltype purity for each cell type in the PBMC multiome dataset, with k=10, k=50, k=100, and k=150.

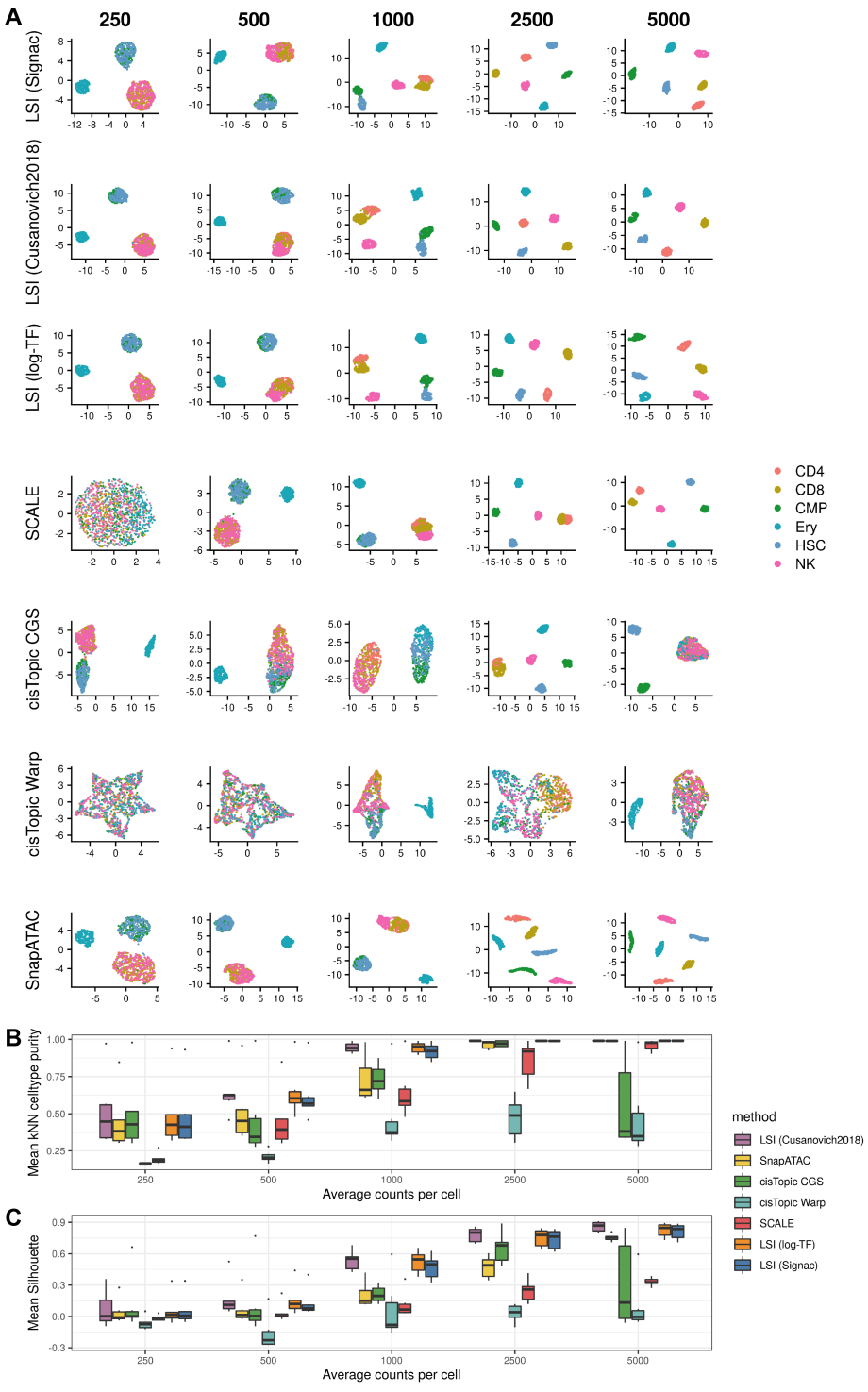

Figure S6. Evaluation of dimension reduction methods using synthetic scATAC-seq data
(A) UMAP representation of synthetic scATAC-seq dataset with mean of 250, 500, 1,000, 2,500, or 5,000 counts per cell. Cells are colored according to their known cell type (CD4 = CD4+ T cell, CD8 = CD8+ T cell, CMP = common myeloid progenitor, Ery = erythroid, HSC = hematopoietic stem cell, NK = natural killer. (B) Mean KNN purity score for each annotated cell type. (C) Mean silhouette score for each annotated cell type. For each boxplot, n=6 points (cell types). Boxplot lower and upper hinges represent first and third quartiles. Upper/lower whiskers extend to the largest/smallest value no further than 1.5x the inter-quartile range. Data beyond the whiskers are plotted as single points.