

Single-cell RNA-seq analysis report

10x Genomics PBMC 10k v3 (public-data example)

Project DEMO-SC-01

Data source 10x Genomics
public datasets

Analysis date 2026-10-01

Issued by HST GENOMICS
bioinformatics

Project overview

This report is a sample deliverable built on public data to show how HST GENOMICS delivers a standard single-cell RNA-seq analysis. Data: the 10x Genomics public dataset "PBMC 10k v3" (healthy donor peripheral blood mononuclear cells, Chromium v3 chemistry), starting from the Cell Ranger filtered feature-barcode matrix.

11,037

cells after QC

22

clusters

1,950

median genes / cell

9.0%

median mito %

Methods

- Software: scanpy 1.12.4, anndata 0.13.4, Leiden via igraph 1.0.0.
- QC: cells with ≥ 200 genes, $< 6,000$ genes and $< 20\%$ mitochondrial reads kept; genes detected in ≥ 3 cells kept. 11,769 cells in, 11,037 retained.
- Normalisation: counts per cell scaled to $1e4$ then $\log_1p$; 2,000 highly variable genes (seurat_v3); scaling and PCA (50 components).
- Clustering and embedding: kNN graph ($k=15$) on 30 PCs, Leiden clustering (resolution 0.8), UMAP.
- Markers: Wilcoxon rank-sum test of each cluster versus the rest. Annotation: the top 30 significantly up-regulated markers of each cluster ($\log_2FC > 0.5$, adjusted $P < 0.05$) were matched against 11 canonical marker sets; a type is assigned only with at least two matching

genes. Clusters matching two of the T/NK, B and myeloid lineages are labelled doublet-like. A cluster is never labelled CD4 T or CD8 T when CD4 or CD8A is significantly down-regulated in it. Clusters with no matching marker among their top 30 are named only if a marker-set mean score is at least 0.5 (shown as score in the Basis column of Table 1); otherwise they stay Unassigned and keep their number. Doublets: scored with scrublet, flagged but not removed.

Quality control

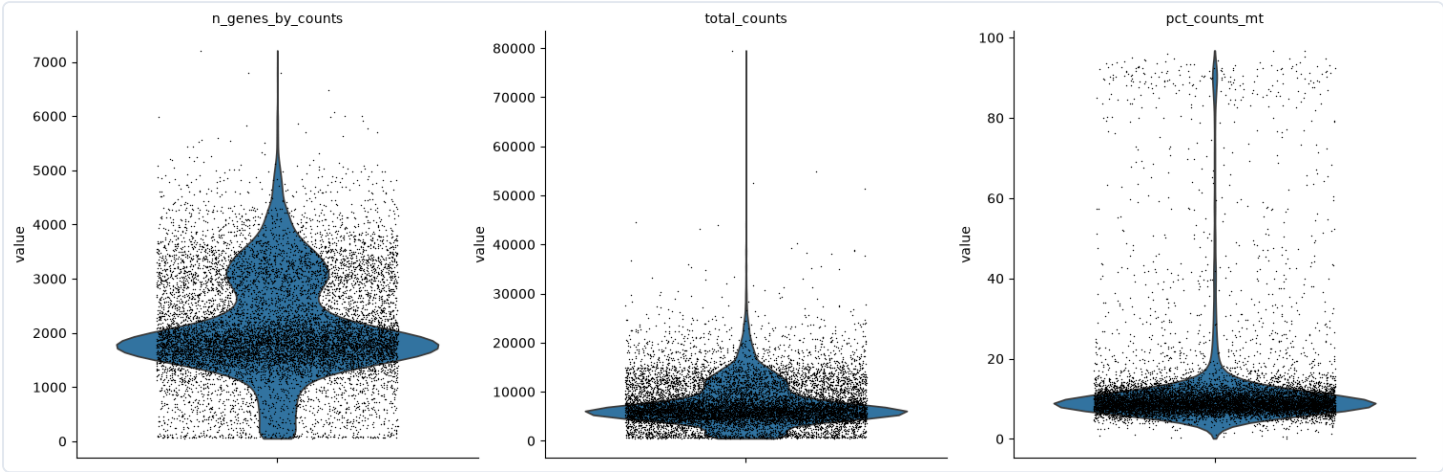


Figure 1. Genes per cell, UMIs per cell and mitochondrial fraction before filtering.

After filtering: median 1,950 genes and 6,729 UMIs per cell; median mitochondrial fraction 9.0%.

Results

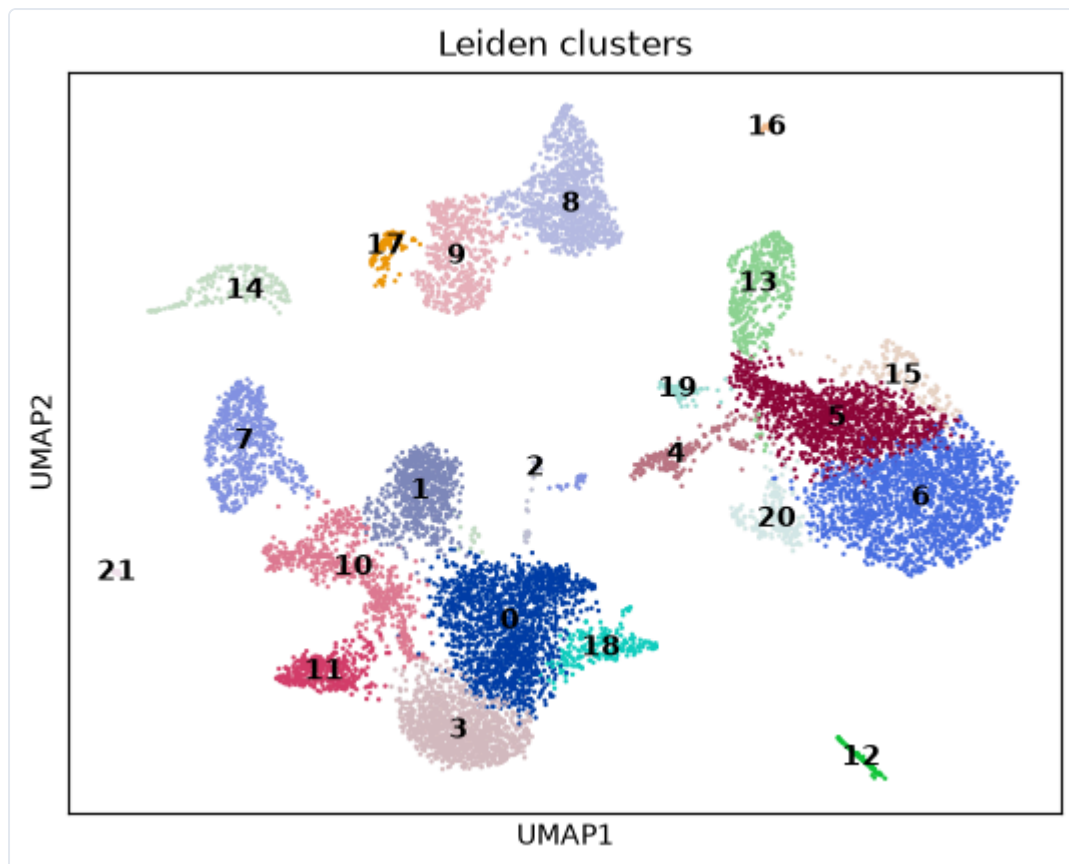


Figure 2. UMAP coloured by Leiden cluster (22 clusters).

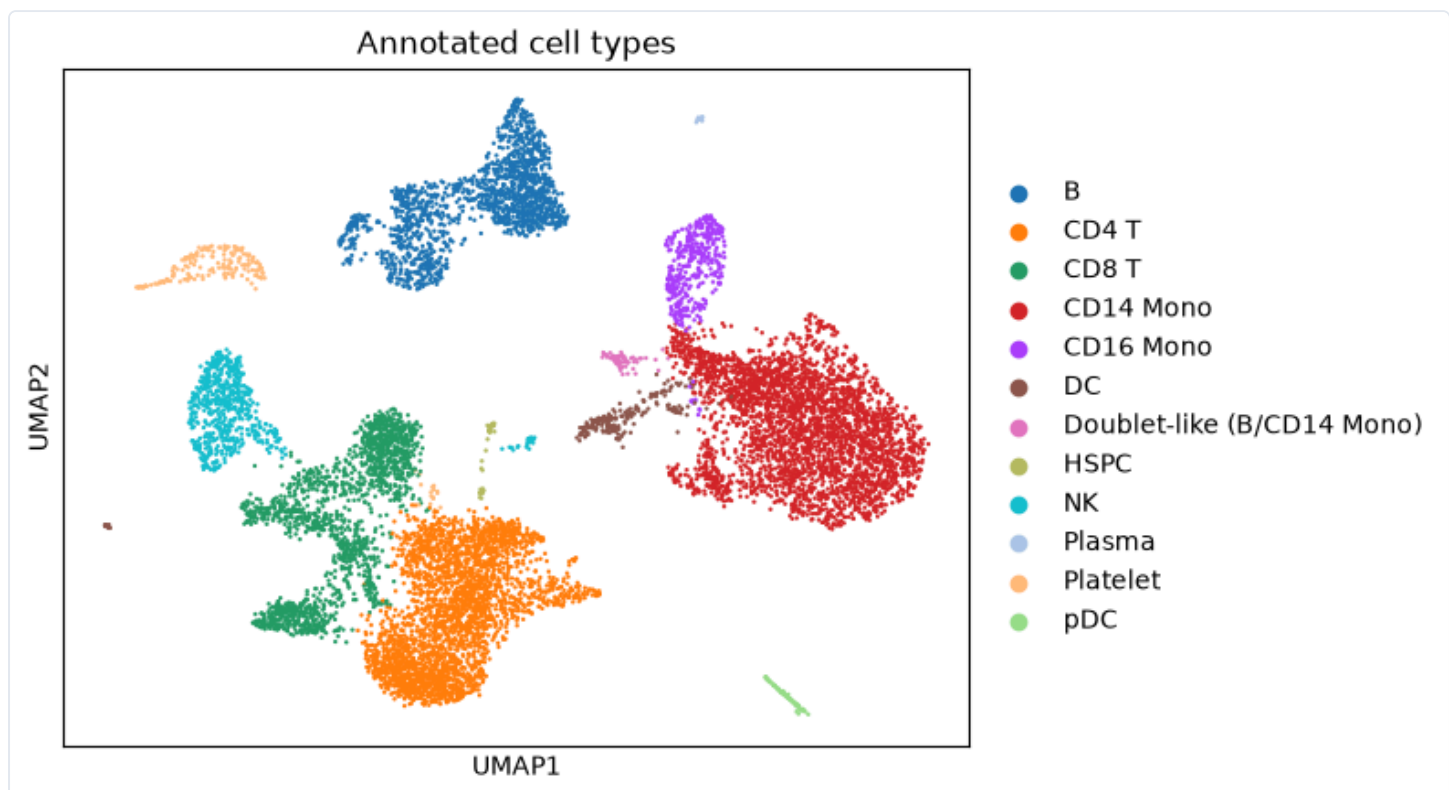


Figure 3. UMAP coloured by annotated cell type.

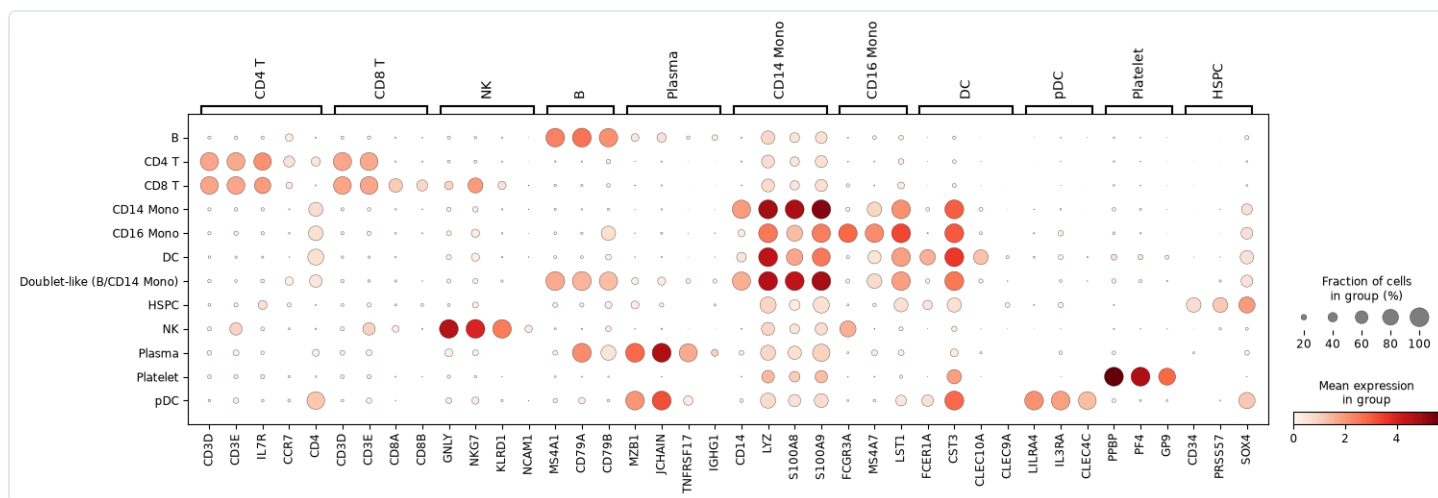


Figure 4. Canonical marker expression by cell type (dot size = fraction expressing, colour = mean expression).

Cluster	Annotation	Cells	Predicted doublets	Basis	Top 5 markers
0	CD4 T	1620	1%	CD3D,CD3E,IL7R	IL7R, LDHB, TRAC, TPT1, RPS29
1	CD8 T	684	3%	CD3D,CD3E	KLRB1, GZMK, KLRG1, NKG7, GZMA
2	HSPC	38	0%	CD34,PRSS57,SOX4	RPS24, RPL7A, HNRNPA1, SOX4, ITM2C
3	CD4 T	1118	0%	score 0.77	RPL30, RPL32, RPS3A, RPS14, RPS12
4	DC	193	11%	FCER1A,CST3,CLEC10A	HLA-DPA1, CST3, HLA-DPB1, HLA-DRA, HLA-DRB1
5	CD14 Mono	1270	0%	score 2.65	LGALS2, CPVL, PSAP, NEAT1, FGL2
6	CD14 Mono	1671	1%	CD14,LYZ,S100A8,S100A9	S100A8, S100A12, S100A9, MND4, VCAN
7	NK	608	4%	GNLY,NKG7,KLRD1	GNLY, NKG7, PRF1, KLRD1, CTSW
8	B	987	0%	MS4A1,CD79A,CD79B	IGHM, IGHD, CD79A, CD37, TCL1A
9	B	481	5%	MS4A1,CD79A,CD79B	MS4A1, BANK1, CD79A, CD37, CD74
10	CD8 T	595	2%	CD3D,CD3E,CD8A,CD8B	CCL5, NKG7, IL32, CTSW, GZMM

11	CD8 T	372	1%	CD8A,CD8B	CD8B, RPS12, RPS3A, RPL32, RPS6
12	pDC	83	16%	LILRA4,IL3RA	UGCG, CCDC50, PLD4, IRF8, TCF4
13	CD16 Mono	394	11%	FCGR3A,MS4A7,LST1	LST1, FCGR3A, AIF1, SMIM25, COTL1
14	Platelet	179	5%	PPBP,PF4,GP9	PF4, PPBP, NRGN, CAVIN2, TUBB1
15	CD14 Mono	138	1%	CD14,LYZ,S100A8,S100A9	TYMP, MX1, TNFSF13B, MX2, IFI44
16	Plasma	19	5%	MZB1,JCHAIN,TNFRSF17	MZB1, SUB1, JCHAIN, SSR4, PPIB
17	B	95	99%	MS4A1,CD79A,CD79B	MS4A1, IGHM, CD79A, IGHD, LINC00926
18	CD4 T	202	0%	CD3D,CD3E	IL32, TRAC, B2M, HLA-A, CD3D
19	Doublet-like (B/CD14 Mono)	69	97%	B: MS4A1,CD79A,CD79B; CD14 Mono: CD14,S100A8,S100A9	MS4A1, CD79A, IGHM, RALGPS2, CD79B
20	CD14 Mono	205	94%	CD14,LYZ,S100A8,S100A9	S100A12, S100A9, S100A8, CD14, VCAN
21	DC	16	0%	CST3,CLEC9A	CLEC9A, IDO1, C1orf54, HLA-DPA1, HLA-DPB1

Table 1. Cluster annotation and markers (full table: tables/cluster_markers_wilcoxon.csv).

Doublets: scrublet flags 5.0% of cells. The following clusters have more than 50% predicted doublets but markers of a single cell type, so their labels are kept and counted; in a real project they would be reviewed against sample information: cluster 17 (B, 99%), cluster 20 (CD14 Mono, 94%).

Cells per annotated type

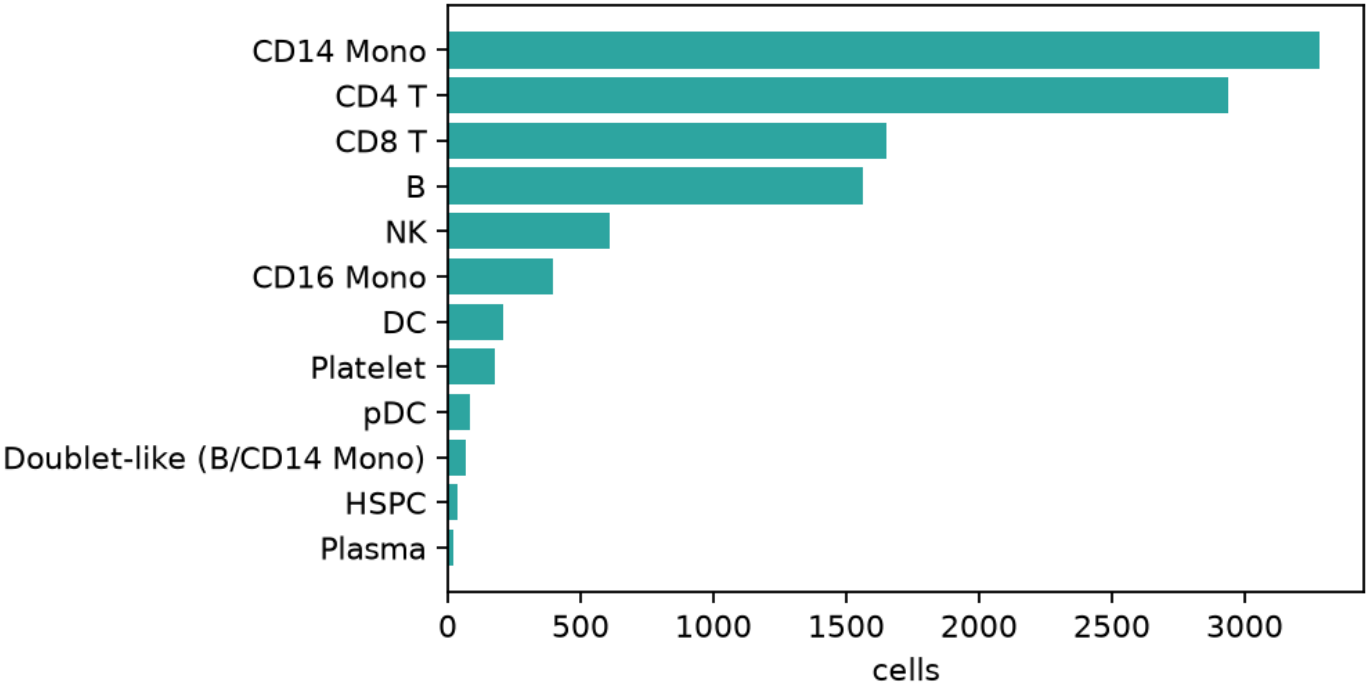


Figure 5. Cells per annotated type.

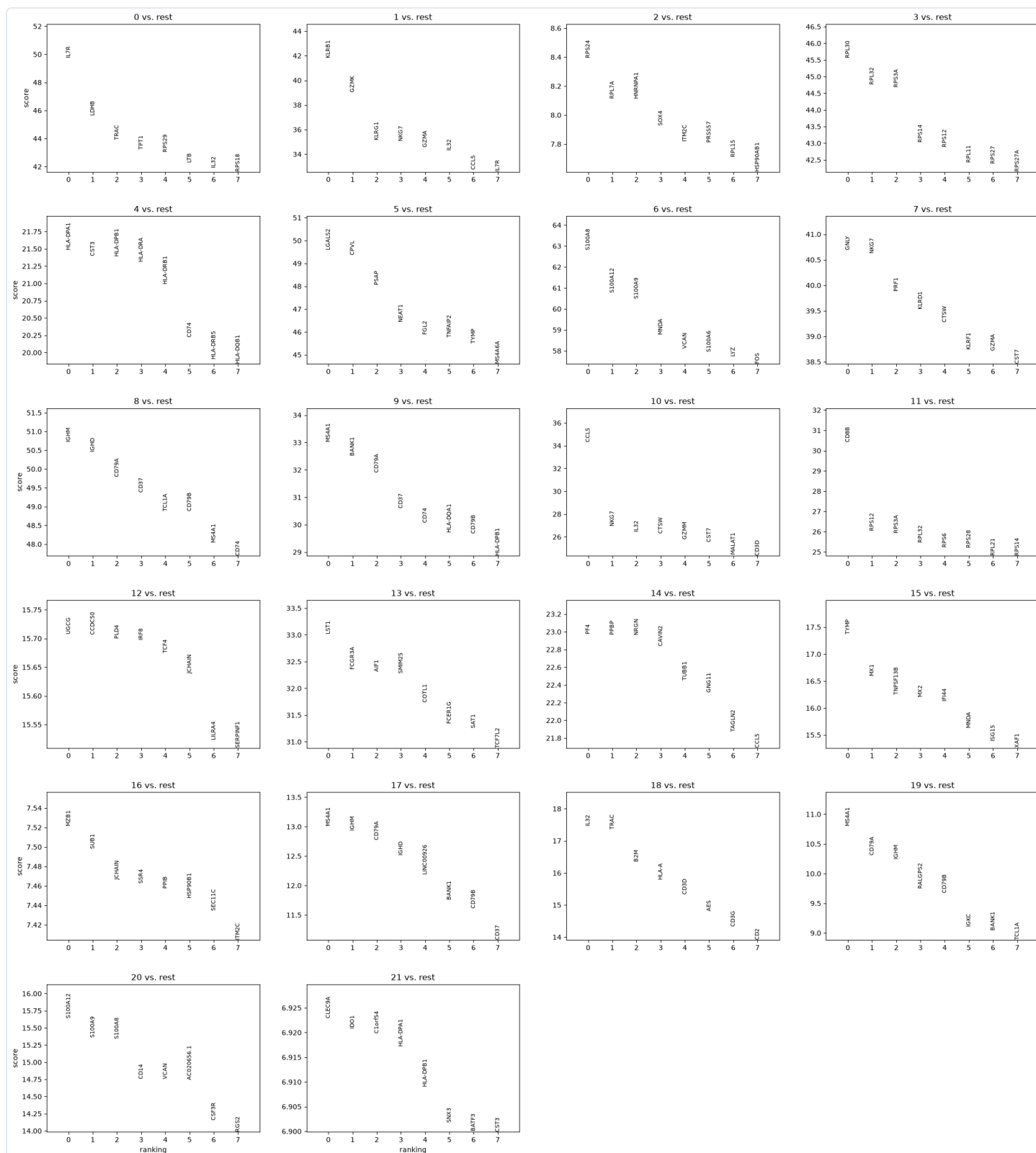


Figure 6. Top 8 markers per cluster by test score.

Deliverables

- tables/cell_metadata_umap.csv: per-cell QC metrics, cluster, cell type and UMAP coordinates
- tables/cluster_markers_wilcoxon.csv: full marker table per cluster

- tables/cell_type_proportions.csv, tables/cluster_marker_scores.csv: proportions and annotation scores
- run_scrna.py, report_lib.py, requirements.txt, README.txt: complete re-runnable code, parameters and instructions
- figures/*.png: publication figures; processed h5ad object available on request
- run_scrna.py: complete, re-runnable analysis code with parameters

Methods paragraph (for the manuscript)

The Cell Ranger filtered matrix was processed with scanpy 1.12.4. Cells with ≥ 200 and $< 6,000$ det

This example uses public data; every number and figure is recomputed by the attached code. Real projects follow the same structure with content defined by the written plan.