

单细胞转录组分析报告

10x Genomics PBMC 10k v3 (公开数据示例)

项目 DEMO-SC-01

数据来源 10x Genomics 公开数据
集

分析日期 2026-10-01

出具 HST GENOMICS 生物信息分
析

项目概况

本报告是公开数据示例交付物，展示 HST GENOMICS 单细胞转录组标准分析的交付形式。数据为 10x Genomics 官方公开数据集「PBMC 10k v3」（健康人外周血单个核细胞，Chromium v3 化学），输入为 Cell Ranger 输出的过滤后基因-细胞矩阵。

11,037

质控后细胞

22

聚类数

1,950

每细胞中位基因数

9.0%

线粒体比例中位数

分析方法

- 软件：scanpy 1.12.4、anndata 0.13.4、leiden (igraph) 1.0.0。
- 质控：保留基因数 ≥ 200 、基因数 $< 6,000$ 、线粒体比例 $< 20\%$ 的细胞；保留在 ≥ 3 个细胞中检出的基因。输入 11,769 个细胞，保留 11,037 个。
- 标准化：每细胞总计数归一到 $1e4$ 后 $\log_1p$ ；以 seurat_v3 方法选 2,000 个高变基因；缩放后做 PCA（50 个主成分）。
- 聚类与降维：前 30 个主成分构建 kNN 图 ($k=15$)，Leiden 聚类（分辨率 0.8），UMAP 可视化。
- 标志基因：每簇对其余细胞做 Wilcoxon 秩和检验。细胞类型注释：取每簇显著上调 ($\log_2FC>0.5$ 、校正 $P<0.05$) 的前 30 个标志基因，与 11 组经典标志基因集比对，至少 2 个重合才命名；T/NK、B、髓系中两个谱系同时满足的簇标为疑似双细胞；谱系定义基因 (CD4、CD8A) 在簇内显著下调时不注为对应 T 细胞类型；前 30 个标志基因无重合的簇，若某一标志基因集平均

打分 ≥ 0.5 则按该集命名（表 1「注释依据」列标为 score），否则标 Unassigned 并保留编号。双细胞：scrublet 打分，只标记不删除。

质量控制

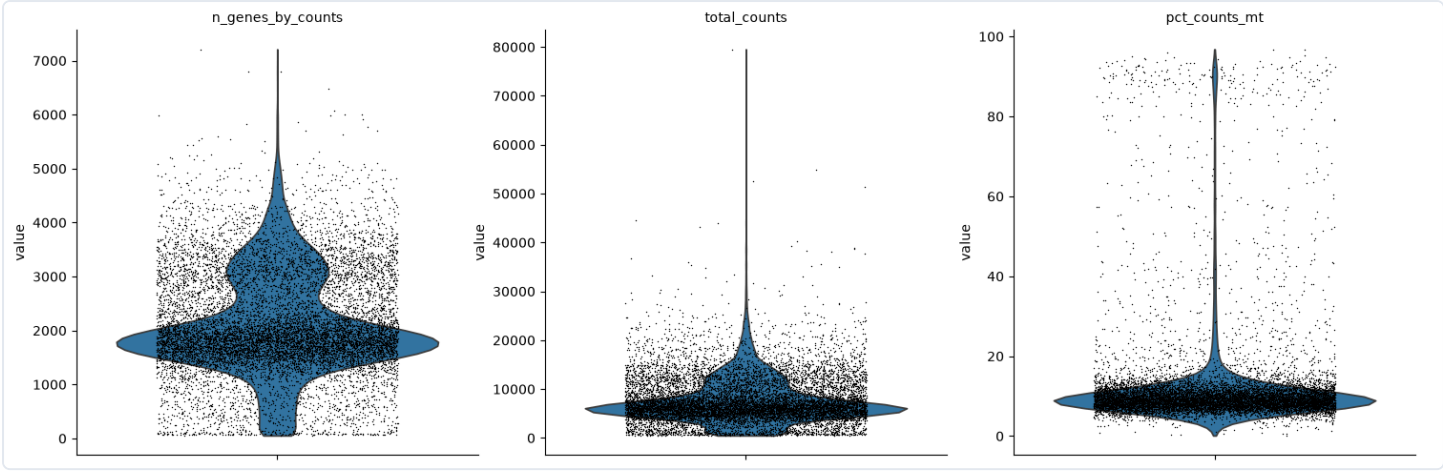


图 1 过滤前每细胞基因数、UMI 数与线粒体比例分布。

过滤后每细胞中位基因数 1,950，中位 UMI 6,729，线粒体比例中位数 9.0%。

结果

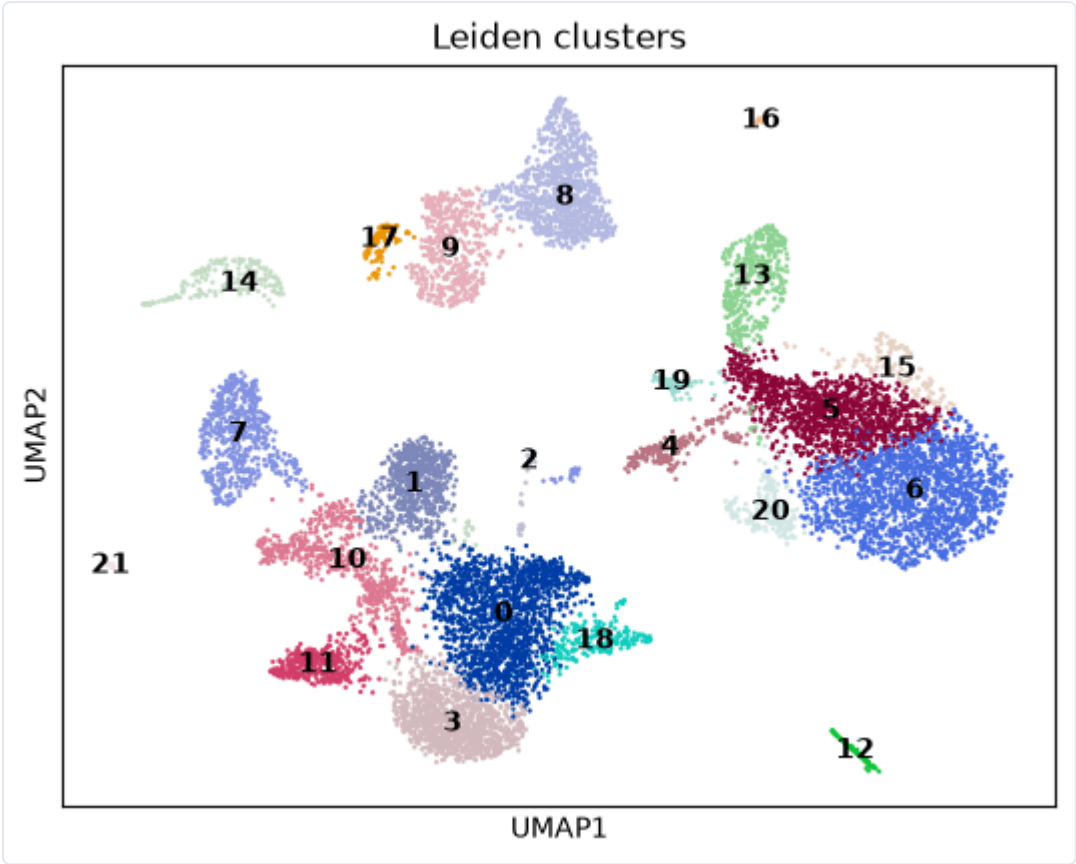


图 2 UMAP，按 Leiden 聚类着色（22 个簇）。

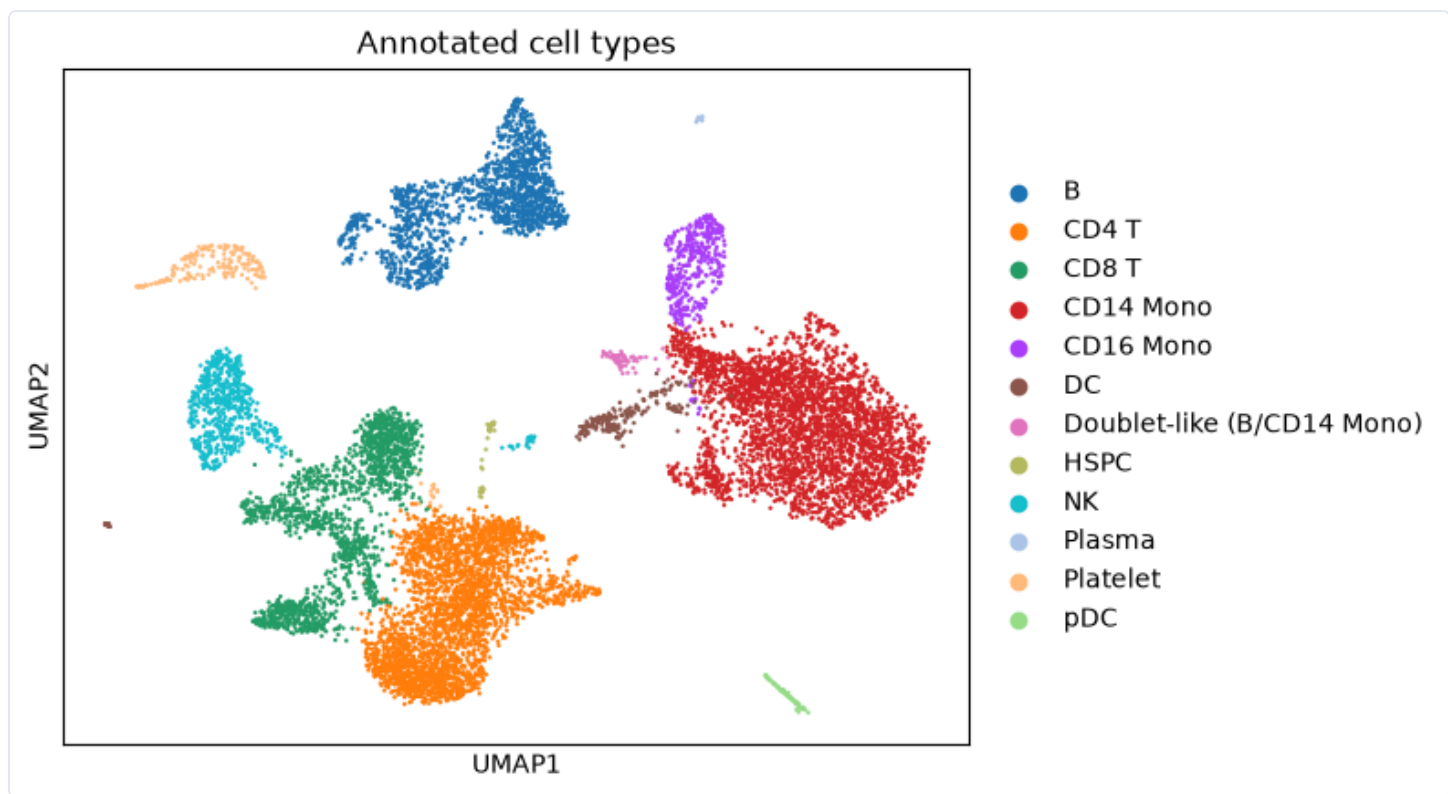


图 3 UMAP，按注释的细胞类型着色。

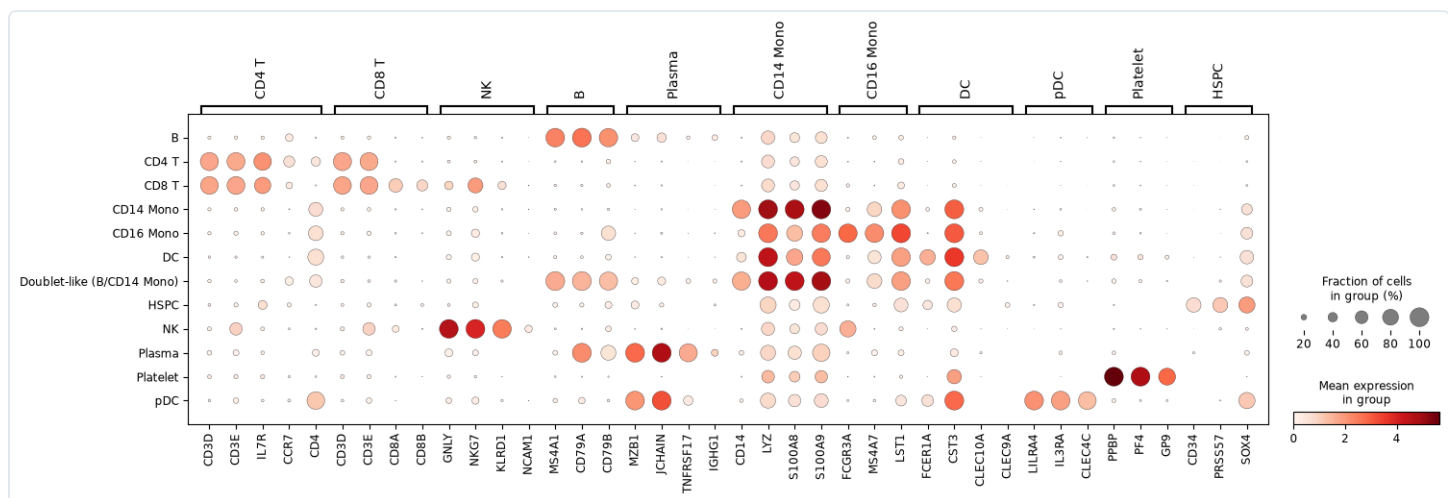


图 4 经典标志基因在各细胞类型中的表达（点大小 = 表达细胞比例，颜色 = 平均表达）。

簇	注释	细胞数	疑似双细胞	注释依据	前 5 个标志基因
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, MND A, 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

表 1 各簇的注释与标志基因（完整表见 tables/cluster_markers_wilcoxon.csv）。

双细胞提示：scrublet 判为双细胞的细胞占 5.0%。以下簇的预测双细胞比例超过 50%，但标志基因只指向单一细胞类型，因此保留原注释、计入比例，建议真实项目结合样本情况复核：簇 17（B，99%）、簇 20（CD14 Mono，94%）。

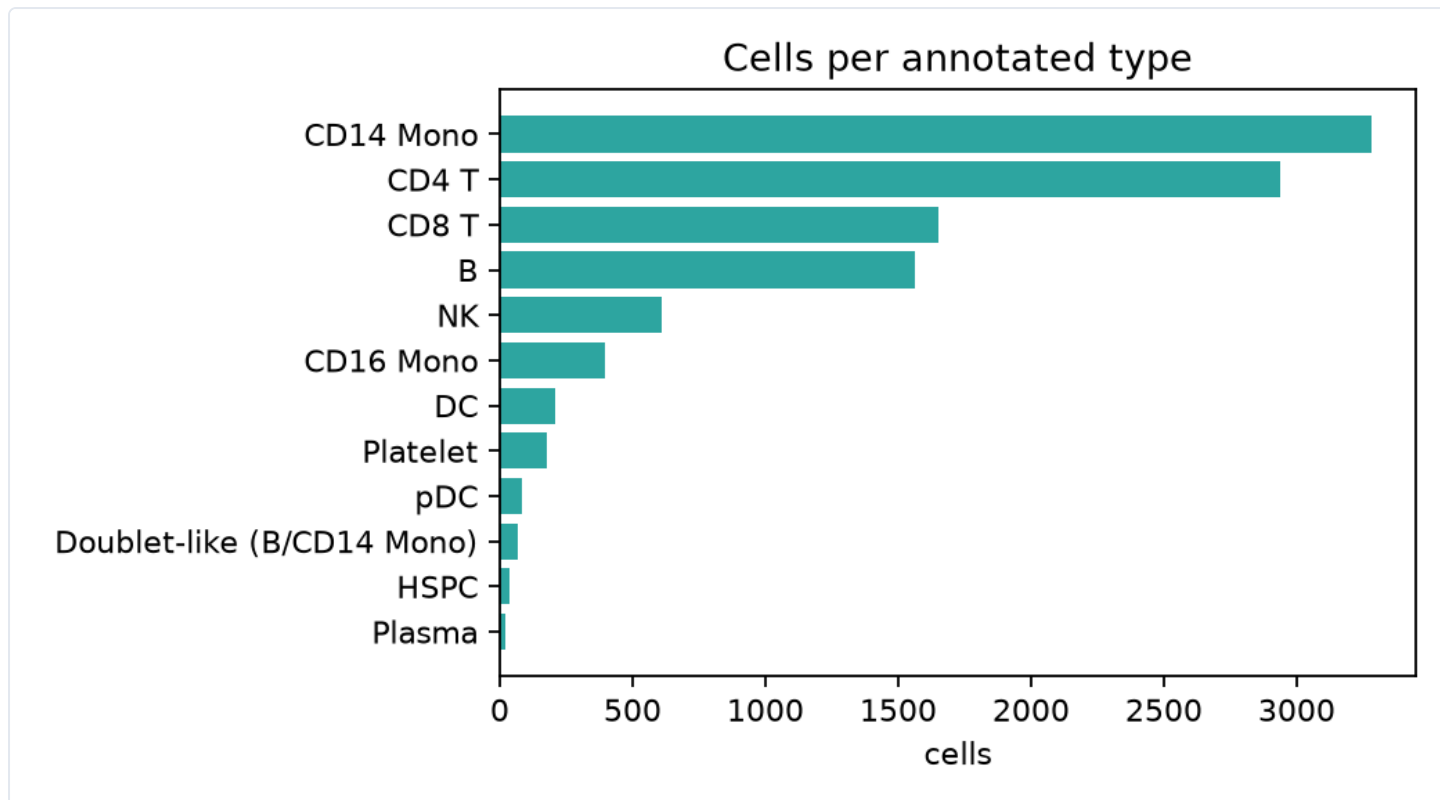


图 5 各细胞类型的细胞数。

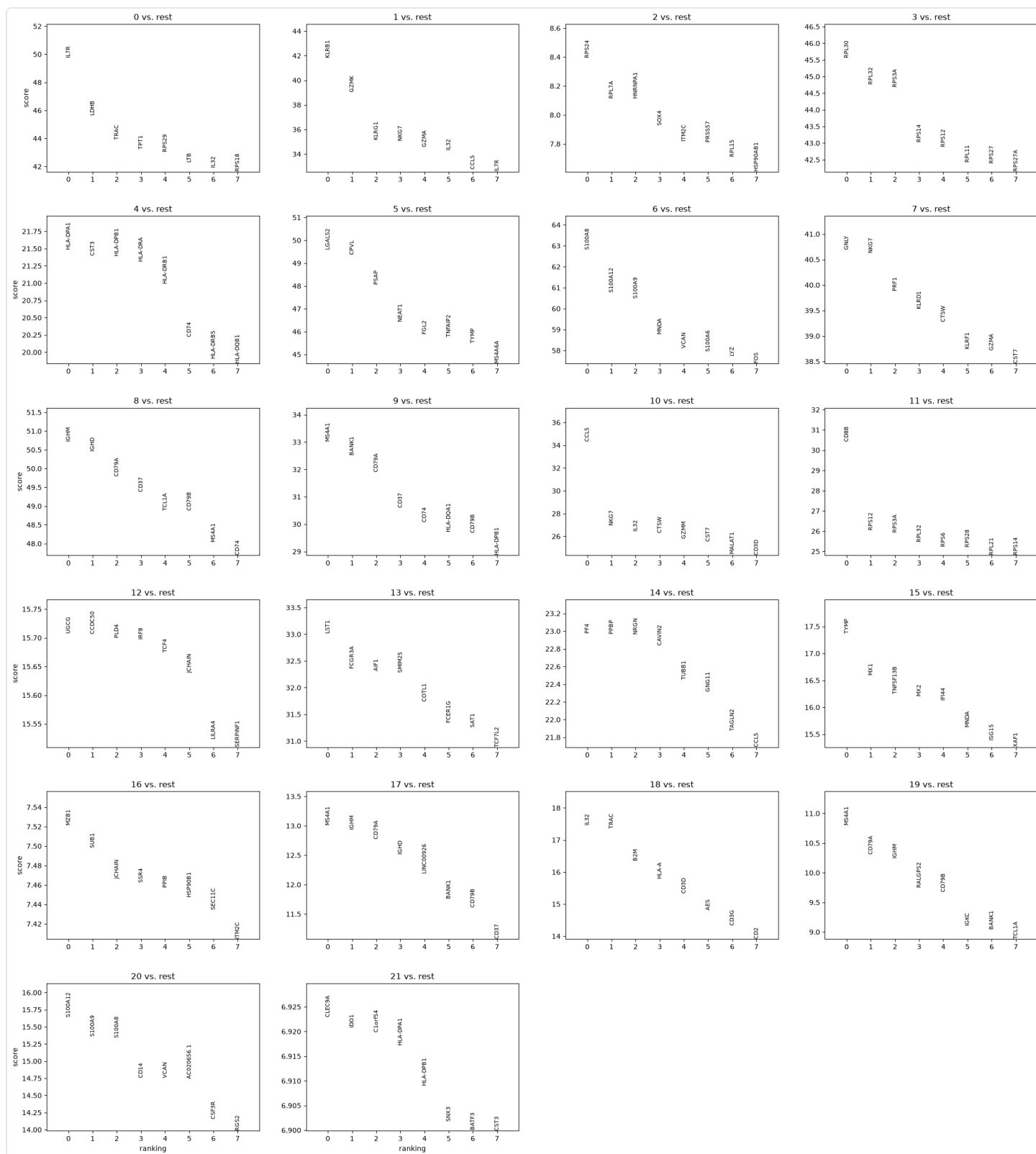


图 6 每簇前 8 个标志基因的检验得分。

交付物清单

- tables/cell_metadata_umap.csv: 每个细胞的质控指标、簇、细胞类型与 UMAP 坐标
- tables/cluster_markers_wilcoxon.csv: 各簇标志基因全表

- tables/cell_type_proportions.csv、tables/cluster_marker_scores.csv：细胞类型比例与注释打分
- run_scrna.py、report_lib.py、requirements.txt、README.txt：可复现的完整分析代码、参数与运行说明
- figures/*.png：出版级图；处理后的 h5ad 对象可按需提供
- run_scrna.py：可复现的完整分析代码与参数

方法学段落（可直接改写进论文）

使用 scanpy 1.12.4 处理 Cell Ranger 过滤矩阵。保留检出基因数 ≥ 200 、 $< 6,000$ 且线粒体读段比例 $< 20\%$ 的细胞，以

本示例使用公开数据，所有数字与图均由随附代码重新计算得到；真实项目交付结构与此一致，内容按书面方案定制。