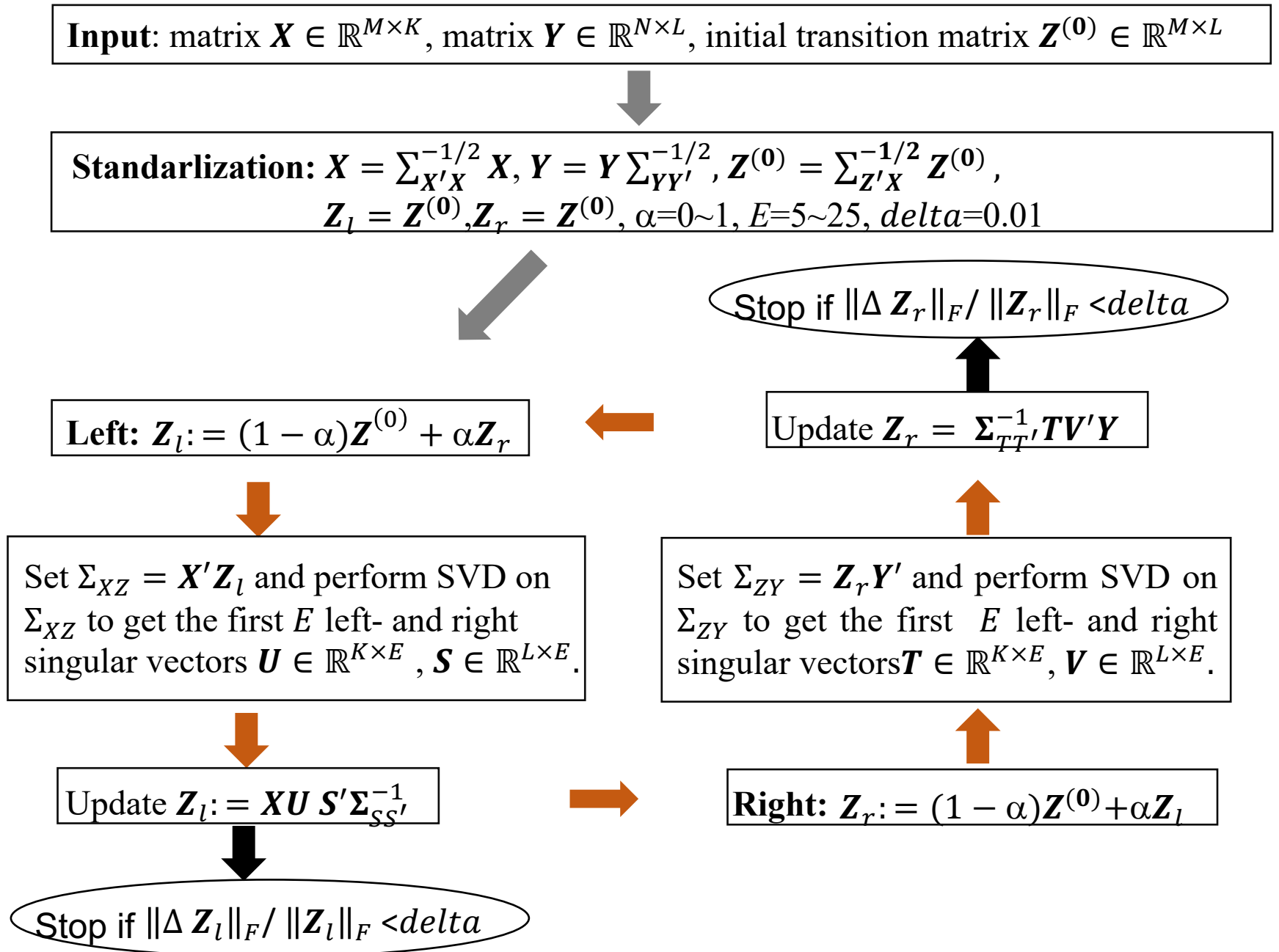


Supplementary Fig. 1

a

Bi-CCA iteration procedure



b

Divide-and-conquer SVD implementation for large-scale matrices

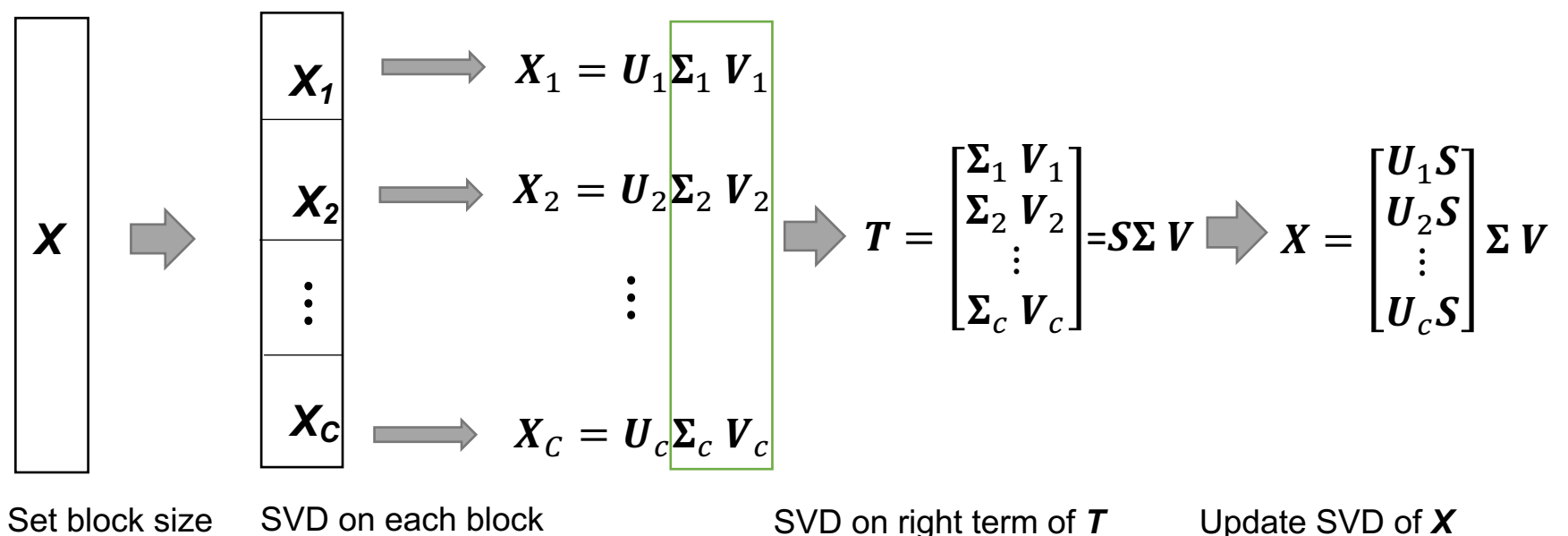


Fig. S1. Implementation of bindSC. Bi-CCA iteration procedure (a). Implementation of divide-and-conquer SVD in bi-CCA for large matrix SVD decomposition (b).

Supplementary Fig. 2

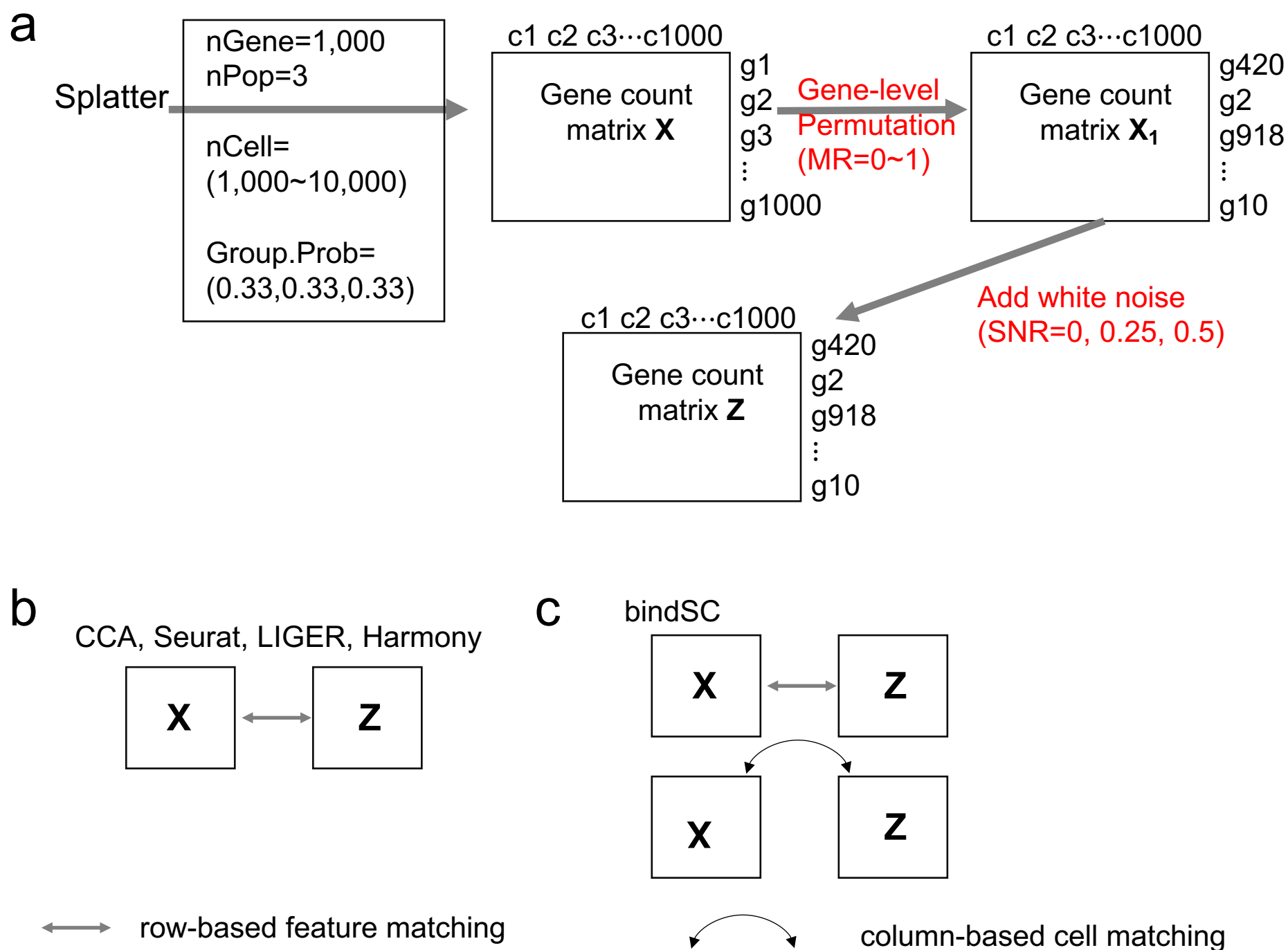


Fig. S2. Simulation settings. Simulation of gene score matrix **Z** (**a**). Each row in **X** denotes a gene (feature) and each column a cell. MR: misalignment rate; SNR: Signal-Noise-Ratio. Previous methods including CCA, Seurat, LIGER, and Harmony take **X** and **Z** as input assuming that cell alignment is unknown (**b**). bindSC takes two parts as input: 1) **X** and **Z** with cell alignment unknown; 2) **X** and **Z** with feature alignment unknown (**c**).

Supplementary Fig. 3

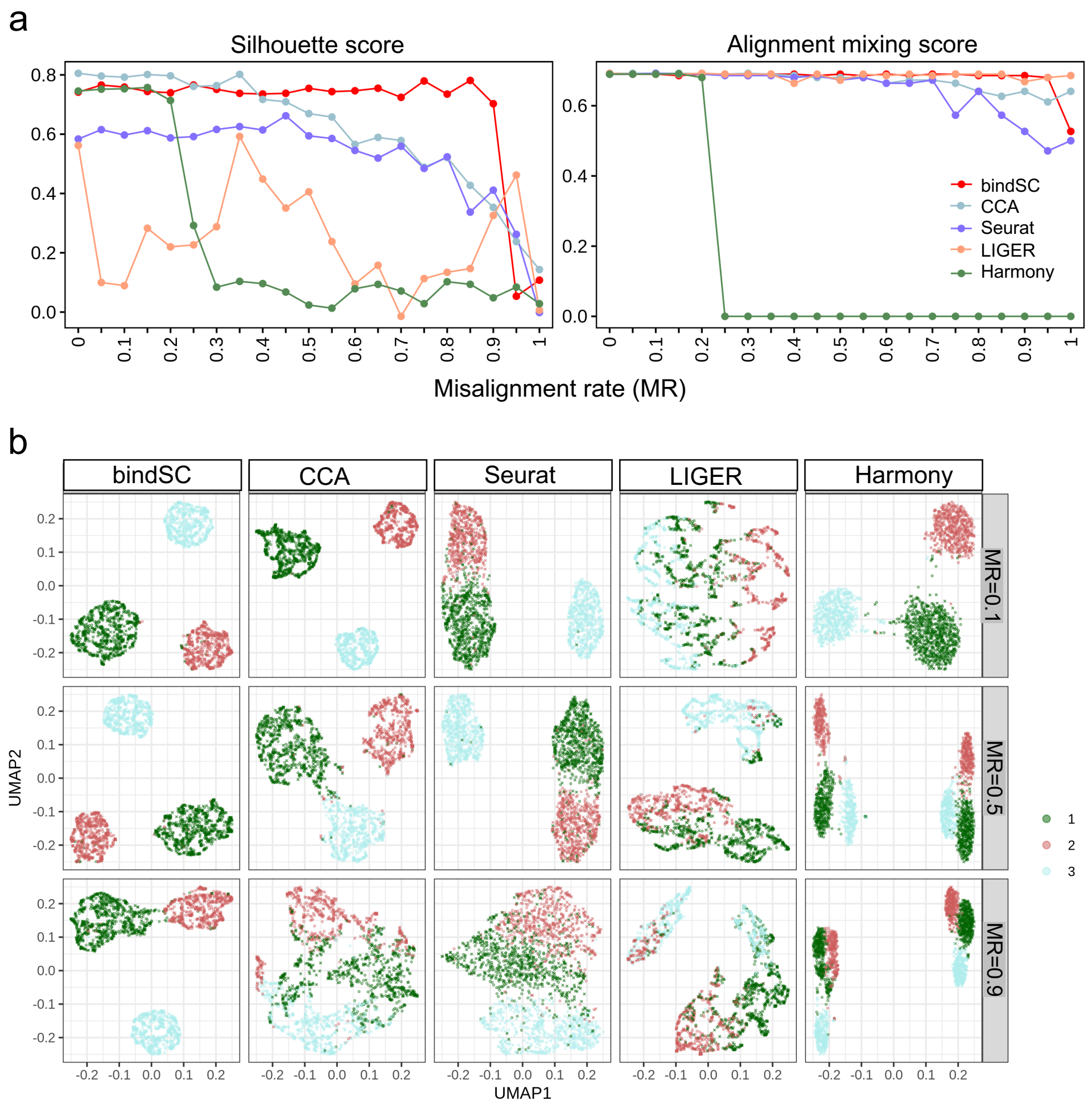


Fig. S3 Benchmarking bindSC performance on simulation datasets. Comparison of bindSC to CCA, Seurat, LIGER, and Harmony based on Silhouette score and alignment mixing score **(a)**. The dataset contains 1,000 genes and 1,000 cells equally distributed in 3 cell types. Signal-to-noise ratio (SNR) was set at 0. X-axes denote the misalignment rates (MR) between features in the two datasets, which ranges from 0 to 1. The features between two datasets have perfect match if MR = 0 and are unrelated if MR = 1. UMAP views of the co-embeddings generated by bindSC, CCA, Seurat, LIGER, and Harmony **(b)**. From top to bottom are results with MR = 0.1, 0.5, and 0.9, respectively. Each point denotes one cell that is colored based on its true cell type label (red, green, or cyan).

Supplementary Fig. 4

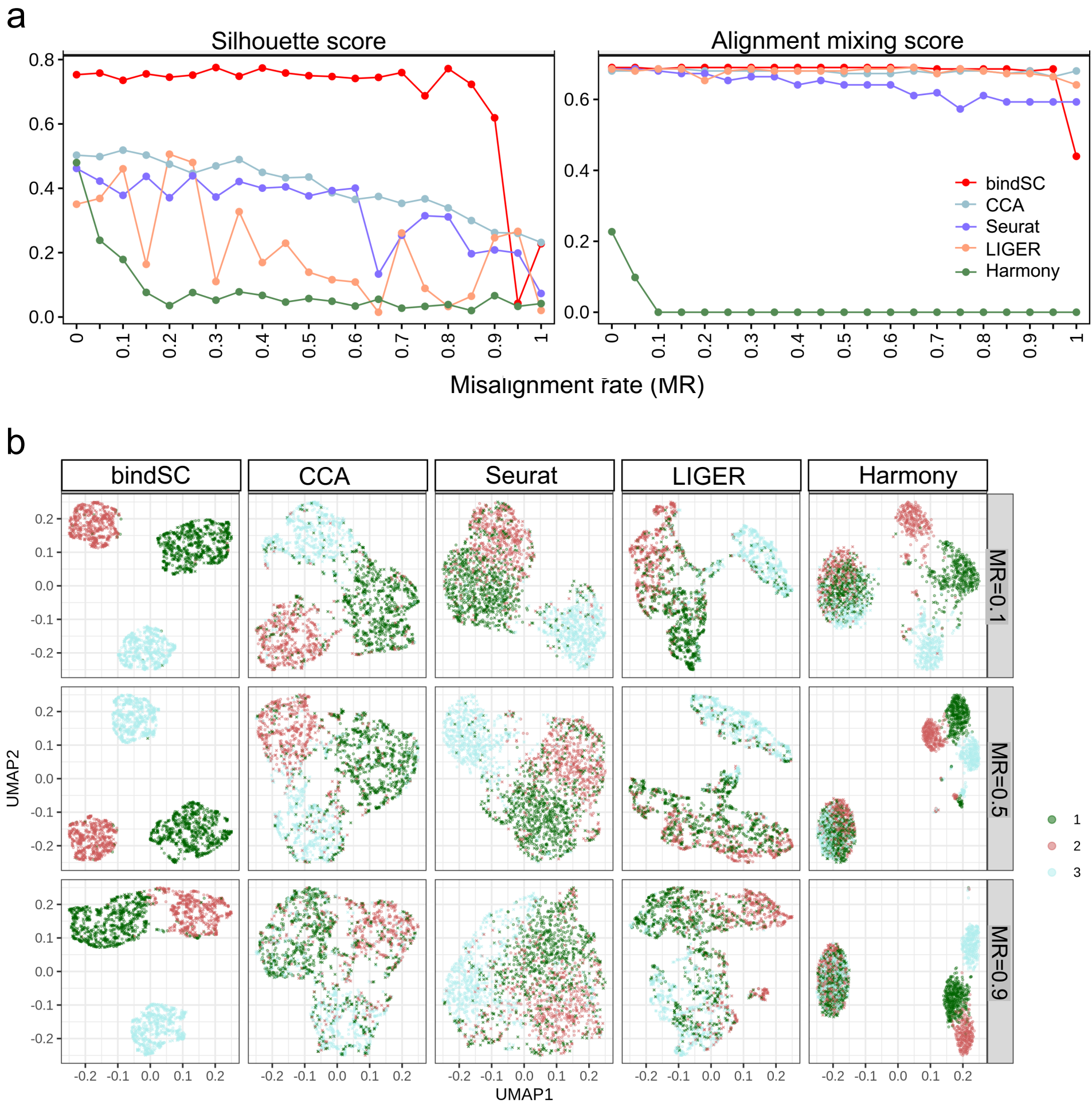


Fig. S4 Benchmarking bindSC performance on simulation datasets. Comparison of bindSC to CCA, Seurat, LIGER, and Harmony based on Silhouette score and alignment mixing score **(a)**. The dataset contains 1,000 genes and 1,000 cells equally distributed in 3 cell types. Signal-to-noise ratio (SNR) was set at 0. X-axes denote the misalignment rates (MR) between features in the two datasets, which ranges from 0 to 1. The features between two datasets have perfect match if $MR = 0$ and are unrelated if $MR = 1$. UMAP views of the co-embeddings generated by bindSC, CCA, Seurat, LIGER, and Harmony **(b)**. From top to bottom are results with $MR = 0.1$, 0.5 , and 0.9 , respectively. Each point denotes one cell that is colored based on its true cell type label (red, green, or cyan).

Supplementary Fig. 5

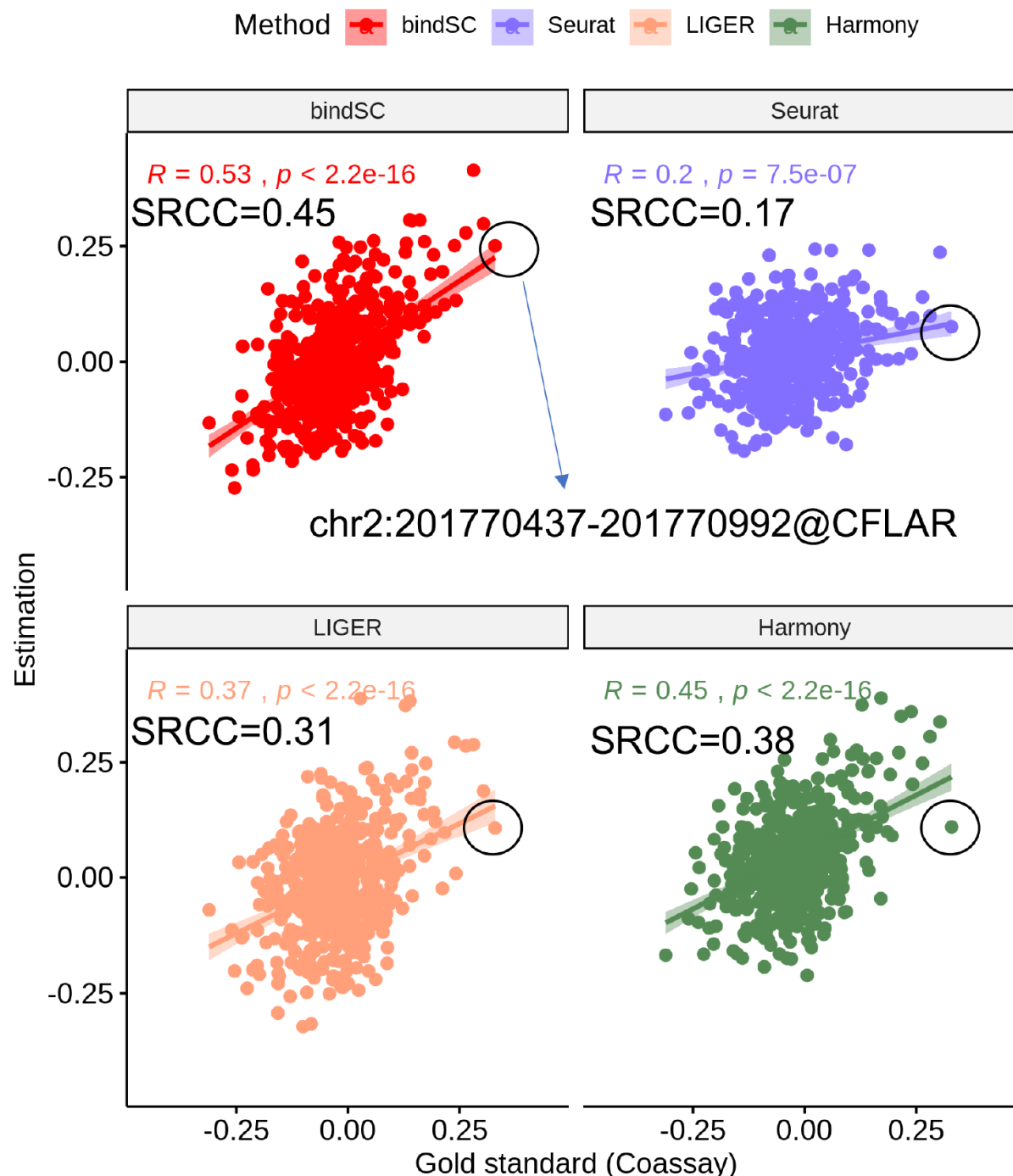


Fig. S5 Estimation of trans peak-gene regulatory elements supported by the Hi-C data. We selected the top 200 *NR3C1* target genes based on co-assayed cell profiles and identified 585 trans peak-gene regulatory elements that were supported by the published Hi-C data ⁵. X-axes are the SRCCs of peak-gene pairs estimated from the co-assayed cells, which serve as the gold standard, while Y-axes are the SRCCs estimated from pseudo-cells generated by each method. The overall concordance between X and Y are further quantified using a single SRCC shown on the up-left corner of each subfigure. Also, the peak-gene pair CFLAR@chr2:217,704,437-201,770,992 is highlighted in each subfigure. Pearson's correlation was performed to produce the coefficients (R) and the P values.

Supplementary Fig. 6

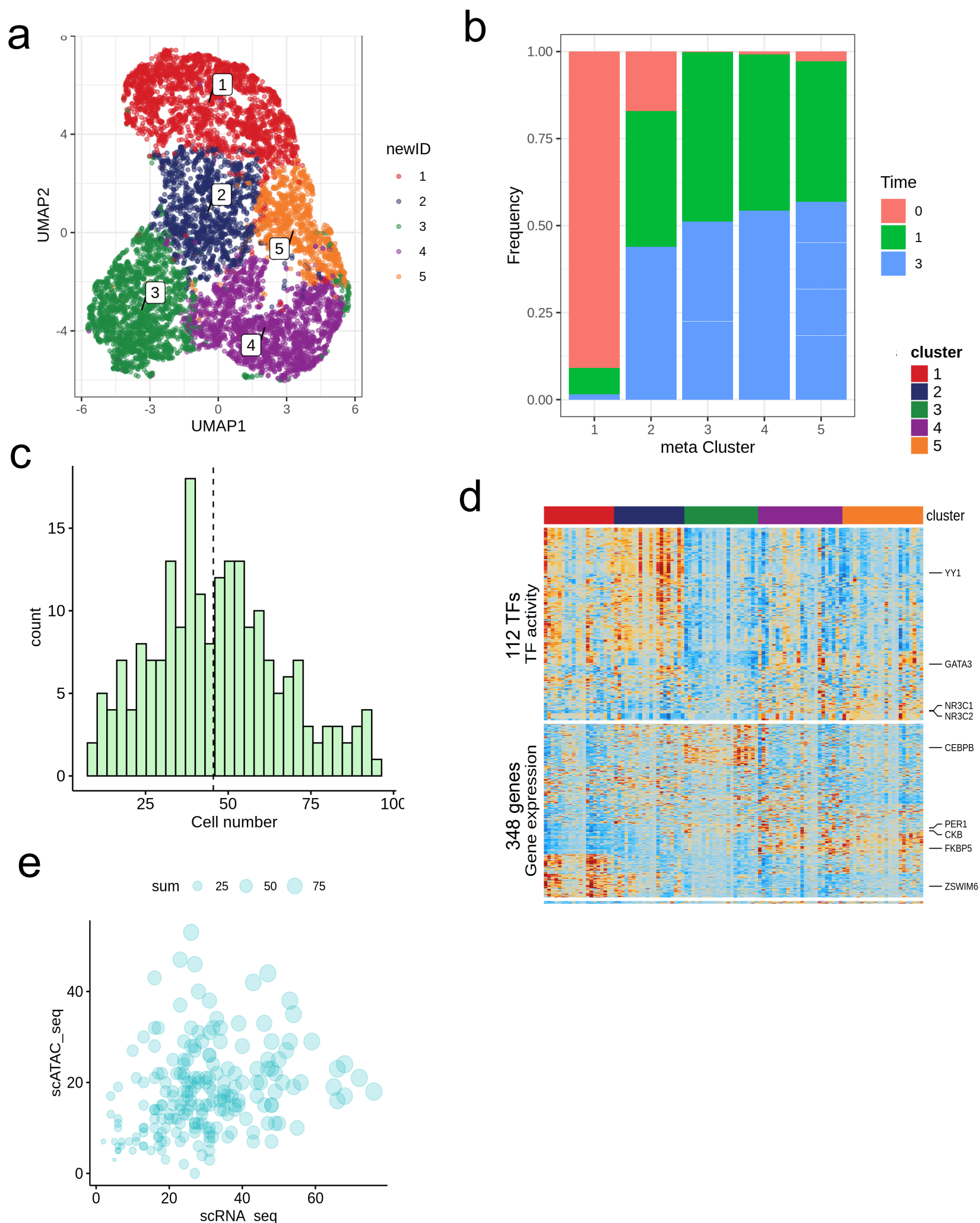


Fig. S6. Joint profiling of gene expression and chromatin accessibility data at the pseudo-cell resolution on the A549 lung cancer cell-line . UMAPs of cells coloring by cluster IDs obtained from unsupervised clustering (meta-cluster) in the bindSC co-embedding **(a)**. Proportion of cells from the 3 treatment times in each of the meta-cluster **(b)**. Histogram showing the number of cells in each pseudo-cell **(c)**. Heatmap showing known genes and TFs associated with glucocorticoid receptor (GR) activation process **(d)**. Each row is one gene/TF and each column is one pseudo cell, grouped/colored by cluster ID. Scatter plot showing the number of cells derived from the scRNA-seq and the scATAC-seq data for each pseudo-cell **(e)**. Each dot denotes one pseudo-cell and the dot size denotes number of cells included in it.

Supplementary Fig. 7

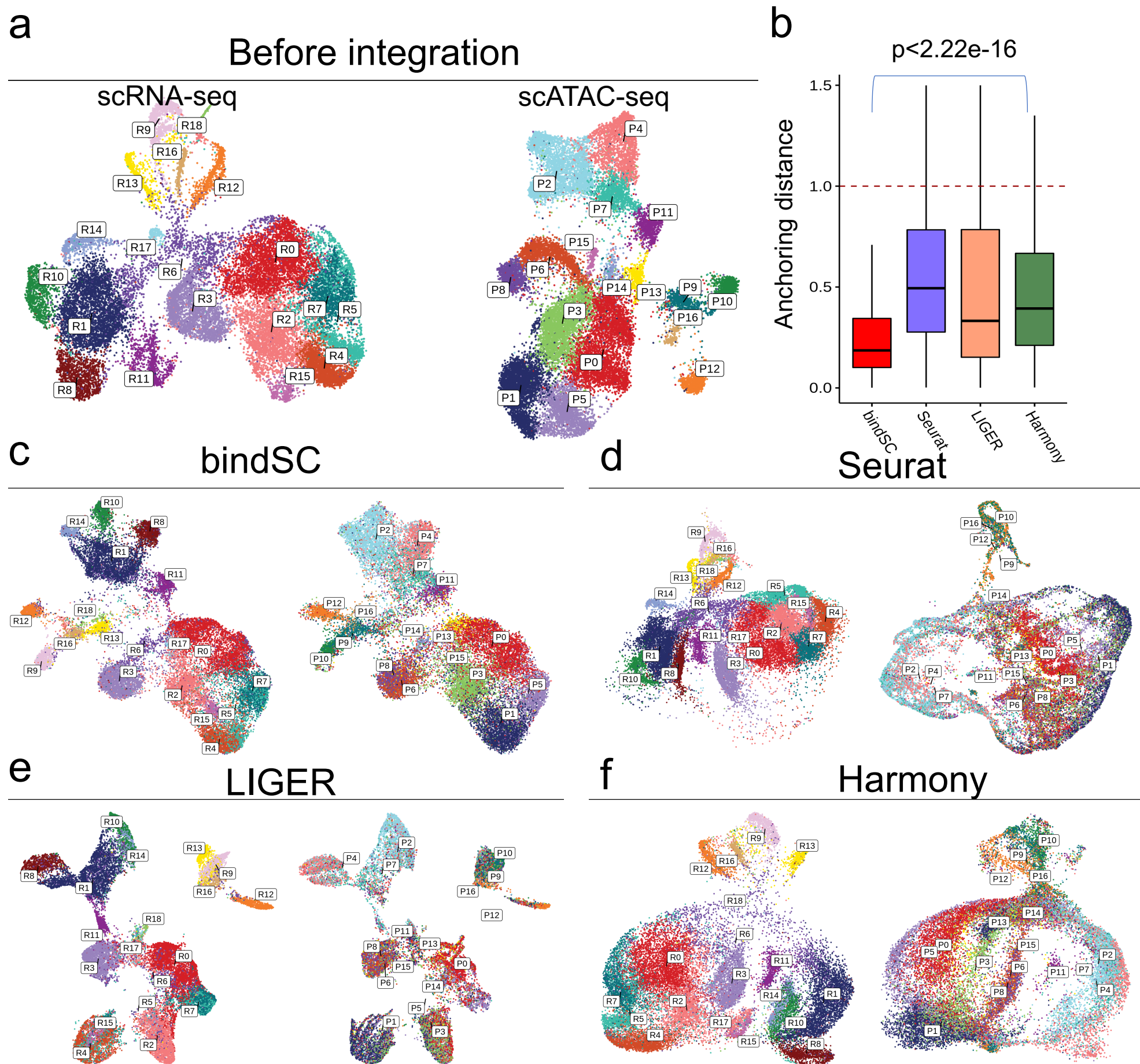


Fig. S7. Integrating single-cell RNA-seq and ATAC-seq data of a mouse skin cell atlas. UMAP of the mouse skin cells before performing integration, colored by clusters deriving from unsupervised clustering of the RNA data and the ATAC data, respectively **(a)**. Anchoring distances achieved by bindSC, Seurat, LIGER and Harmony **(b)**. UMAP of cells in the multiomics co-embeddings generated by bindSC **(c)**, Seurat **(d)**, LIGER **(e)**, and Harmony **(f)**, respectively. For each method, the left panel shows cells from the RNA-seq data and the right panel shows cells from the ATAC-seq data.

Supplementary Fig. 8

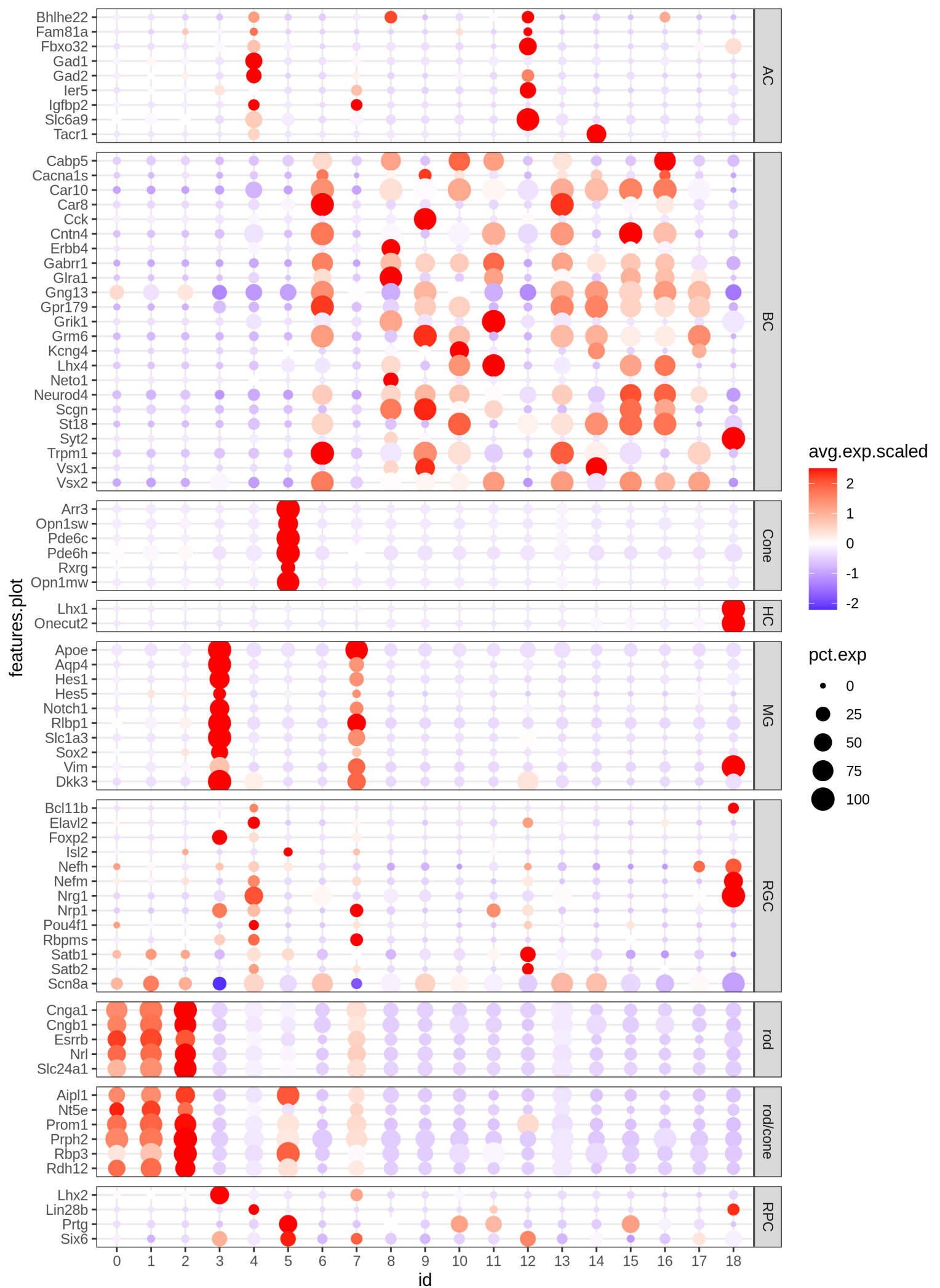


Fig. S8. Cell type annotation for cells in the mouse retina cell atlas. In the heatmap, X-axes denotes cluster IDs in the RNA clusters, while Y-axes denotes known retinal cell-type-specific marker genes for the AC, BC, cone, HC, RGC, rod, and RPC cells, respectively. The color gradient in each dot denotes the expression level and the dot size denotes percentage of cells that express the gene.

Supplementary Fig. 9

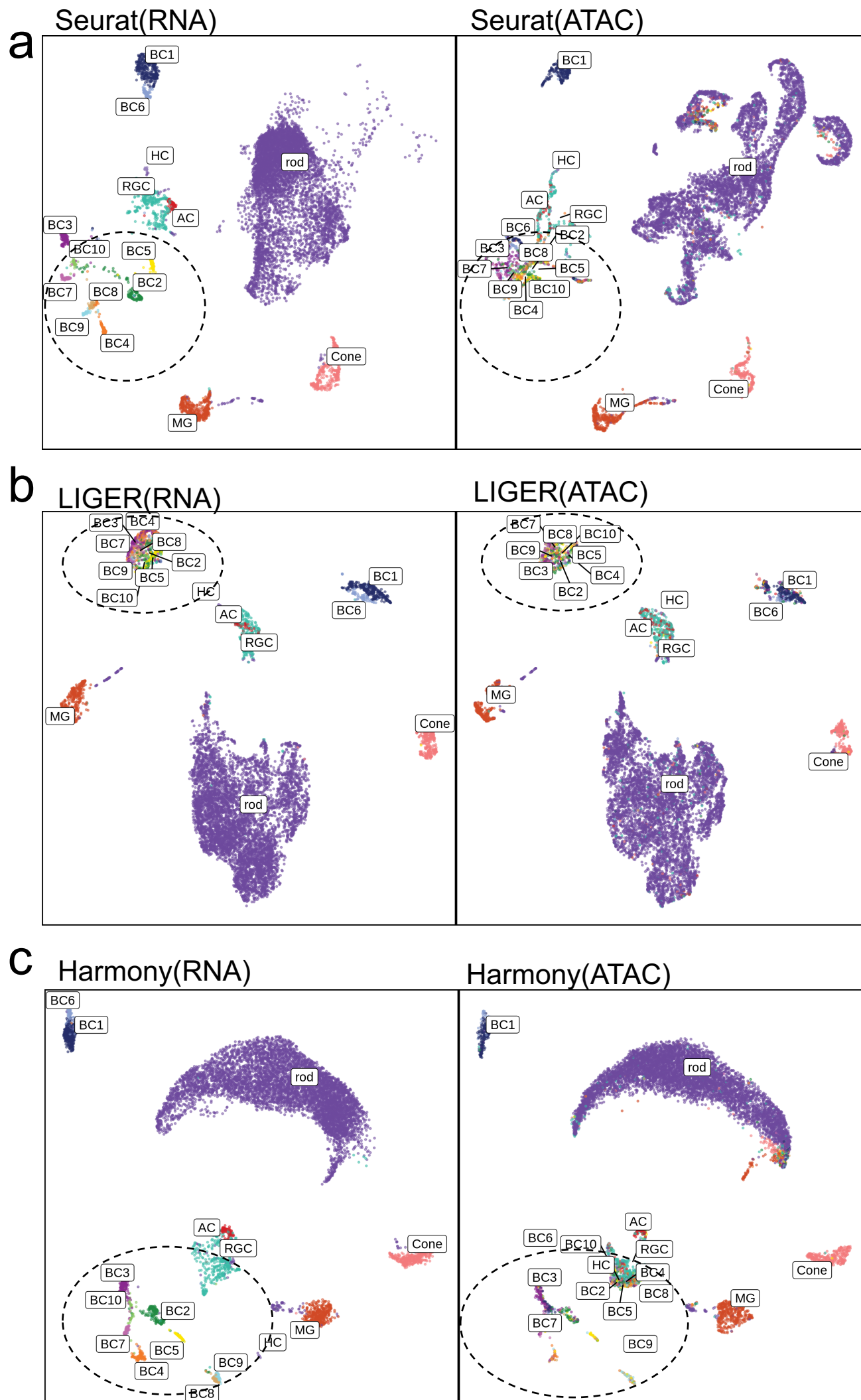


Fig. S9. UMAP visualization of mouse retina cells in the *in silico* co-embeddings generated by Seurat, LIGER, and Harmony. From top to bottom are the results for Seurat (a), LIGER (b), and Harmony (c) respectively. The left panel shows cells from the RNA-seq data. The right panel shows cells from the ATAC-seq data. Cells were colored based on cell types identified in **Supplementary Fig. 8**. The oval regions were zoomed in **Fig. 4 g-j**.

Supplementary Fig. 10

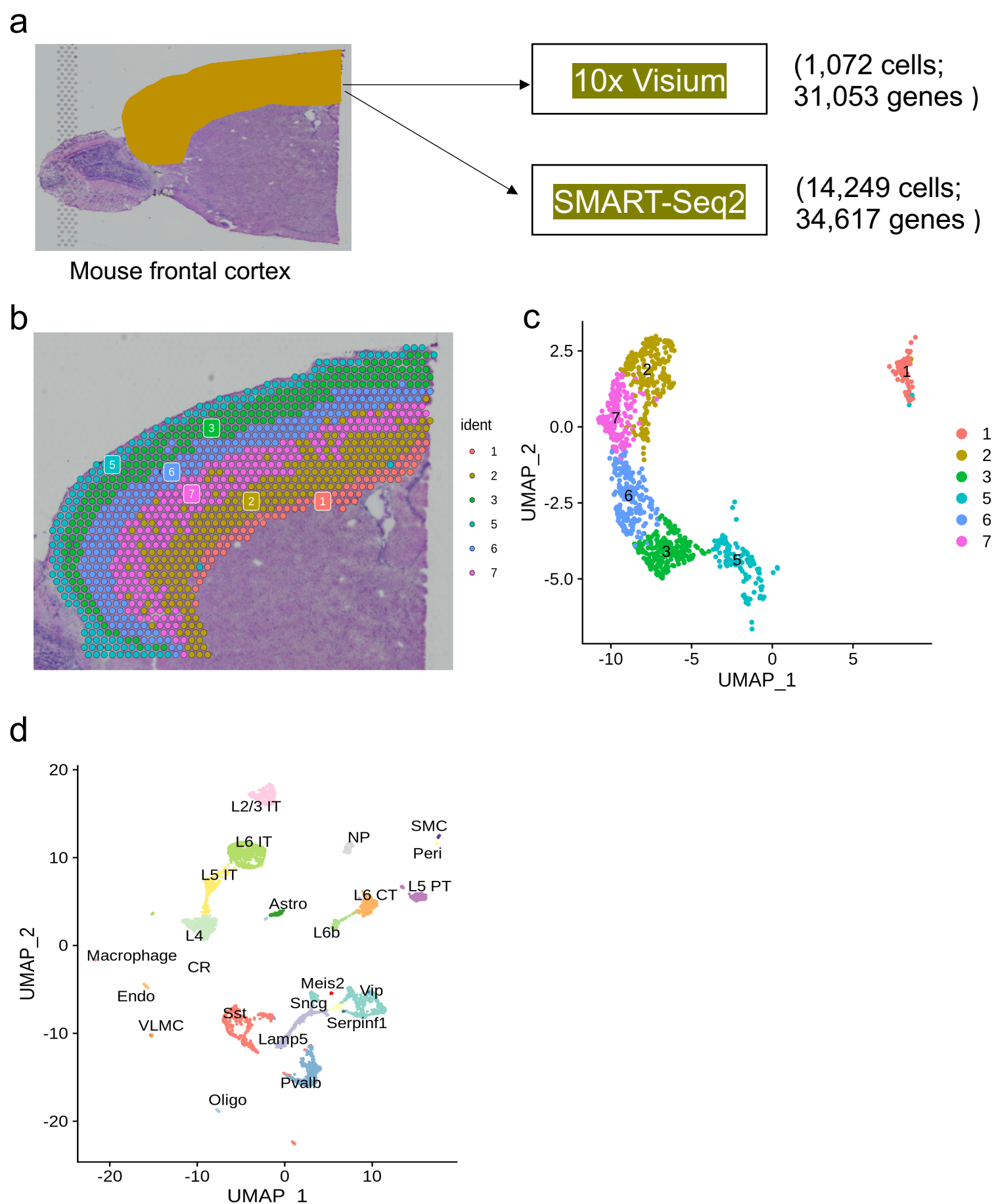


Fig. S10. Integrating 10x Visium spatial transcriptomics data with SMART-Seq2 scRNA-seq data from mouse frontal cortex cells. Schematic representation of data used for integration **(a)**. Histology image of mouse frontal cortex overlaying with cells from the 10x Visium technology **(b)**. UMAP view of the RNA expression of the 1,072 spots in the 10x Visium data **(c)**. UMAP view of the transcriptomes of 14,249 frontal cortex cells produced by SMART-Seq2 technology **(d)**. Cell-type labels in **(d)** are derived from the published SMART-Seq2 dataset.

Supplementary Fig. 11

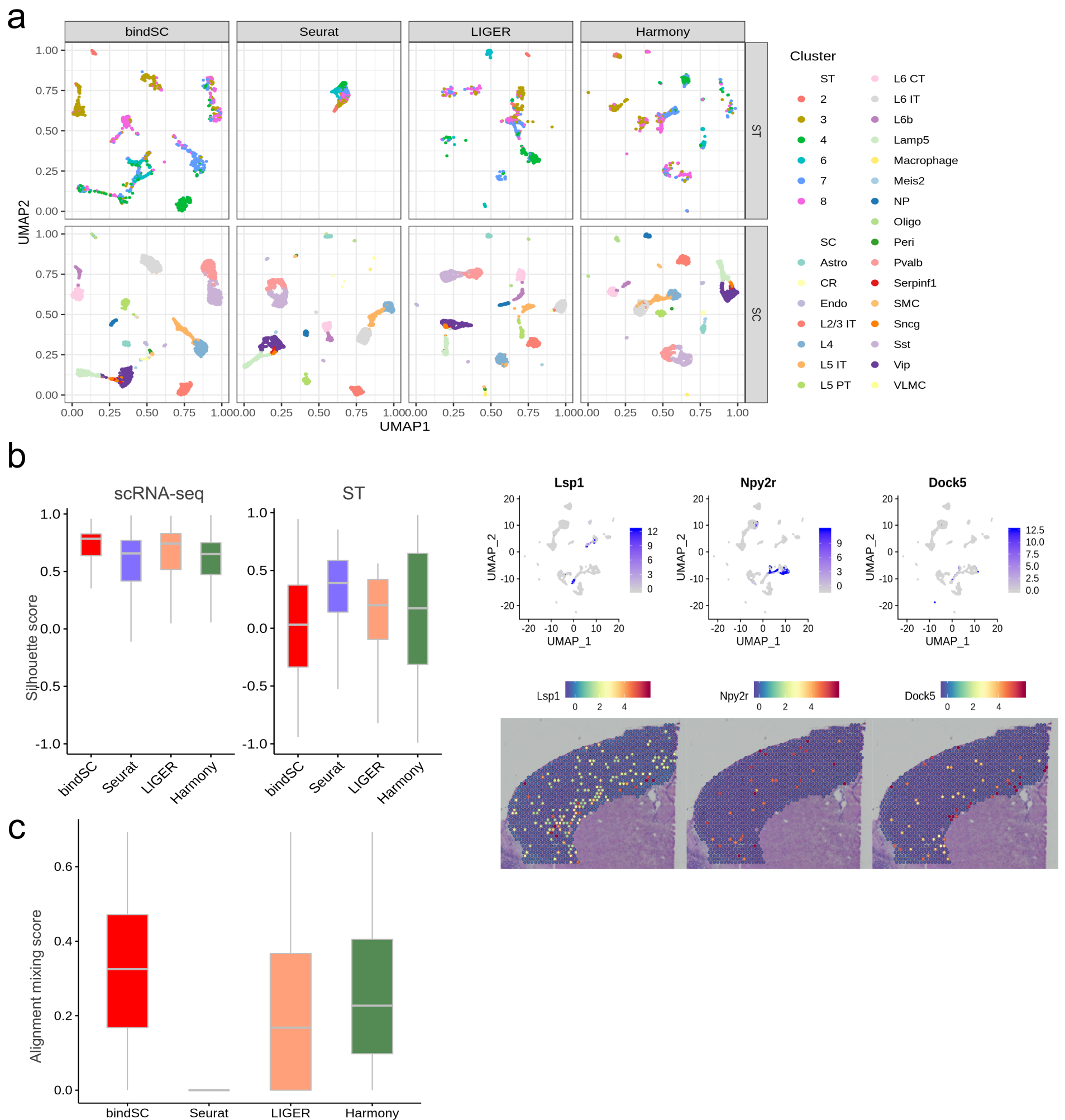


Fig. S11 Performance of four methods on integrating spatially resolved transcriptomic (ST) data with dissociated scRNA-seq data from mouse frontal cortex cells. Related to **Fig. 5a**. UMAP of cells from mouse frontal cortex datasets, separated by sequencing technology with ST on the top panel and scRNA-seq data on the bottom panel (**a**). Cell-type labels are consistent with those from **Supplementary Fig. 10c-d**. Comparison of cell-type classification based on Silhouette scores (**b**). Comparison of dataset alignment based on alignment mixing scores (**c**). Gene expression profiles of three Lamp5-related marker genes *Lsp1*, *Npy2r*, and *Dock5* from the scRNA-seq data (**d**) and the ST data (**e**).

Supplementary Fig. 12

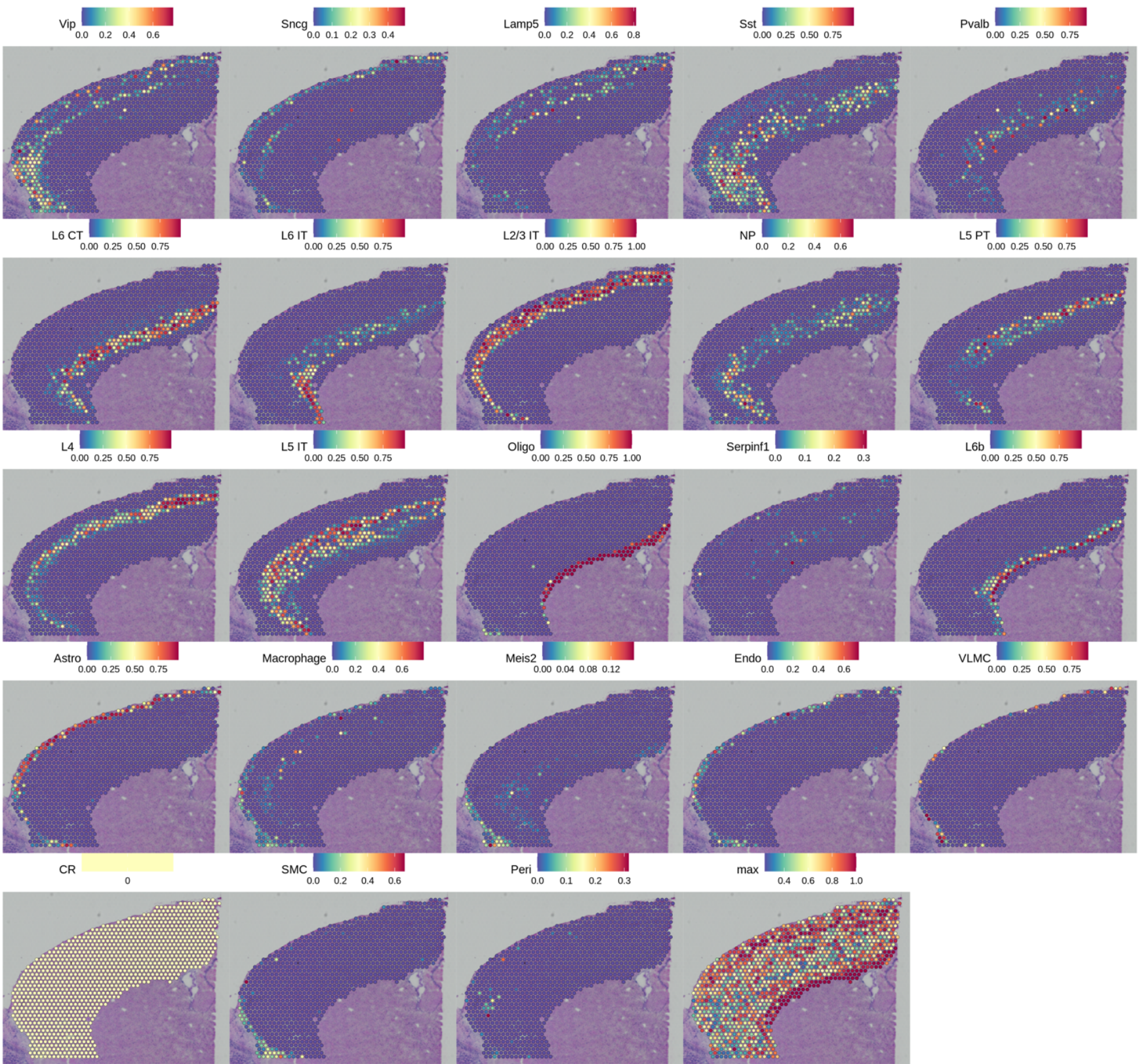


Fig. S12. Cell types mapped by Seurat onto mouse brain histology images.

Supplementary Fig. 13

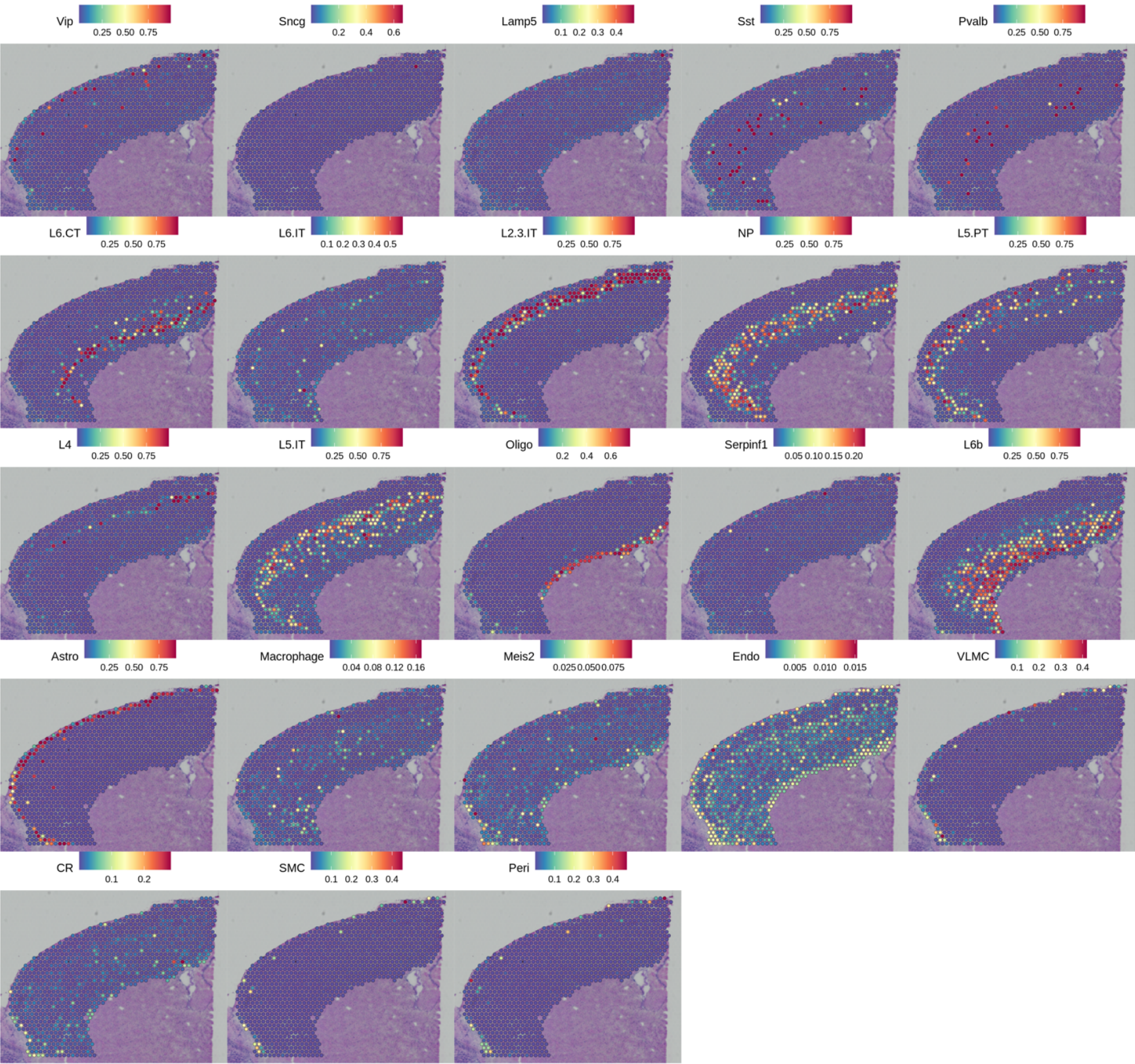


Fig. S13. Cell types mapped by LIGER onto mouse brain histology images.

Supplementary Fig. 14

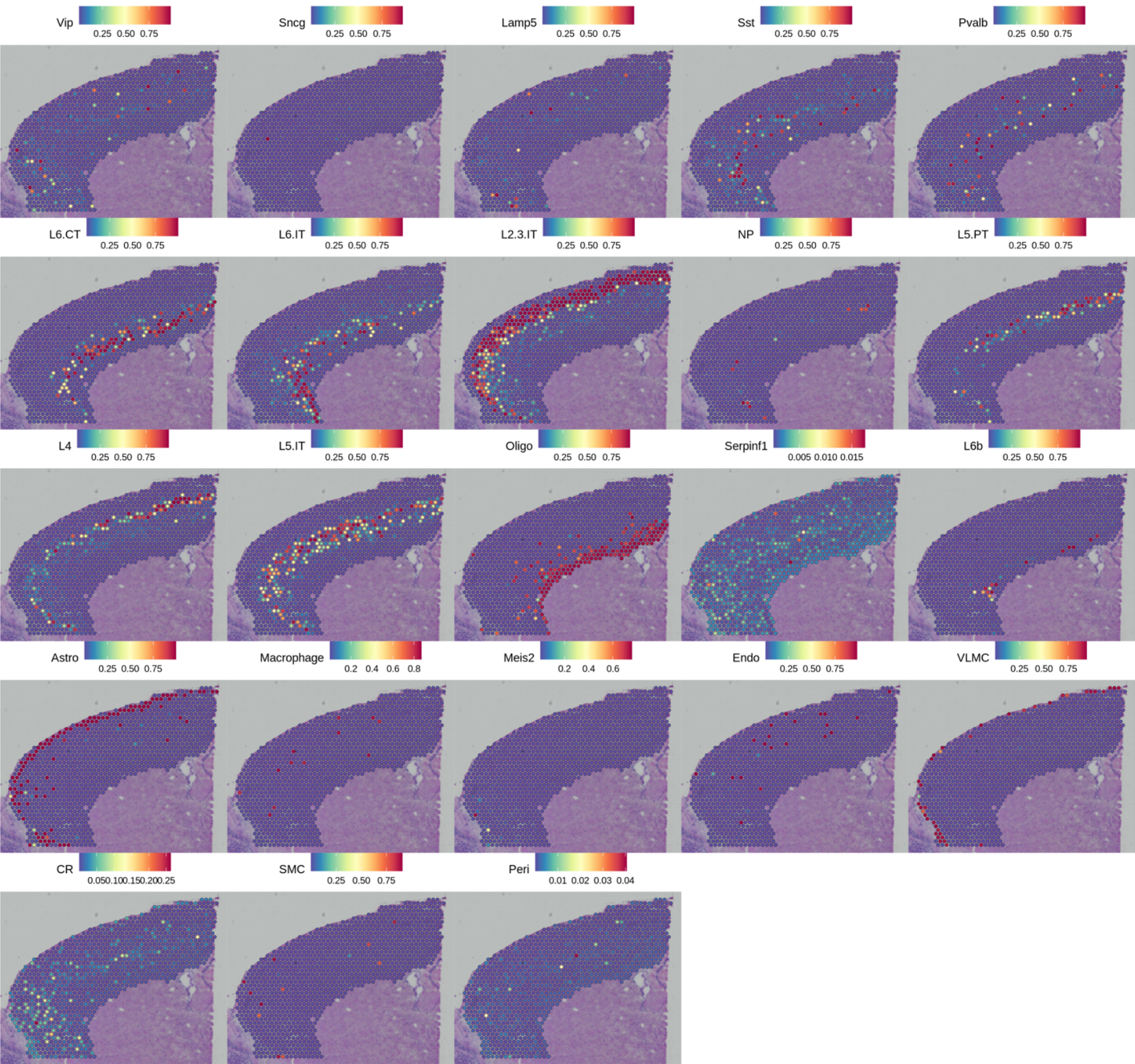


Fig. S14. Cell types mapped by Harmony onto mouse brain histology images.

Supplementary Fig. 15

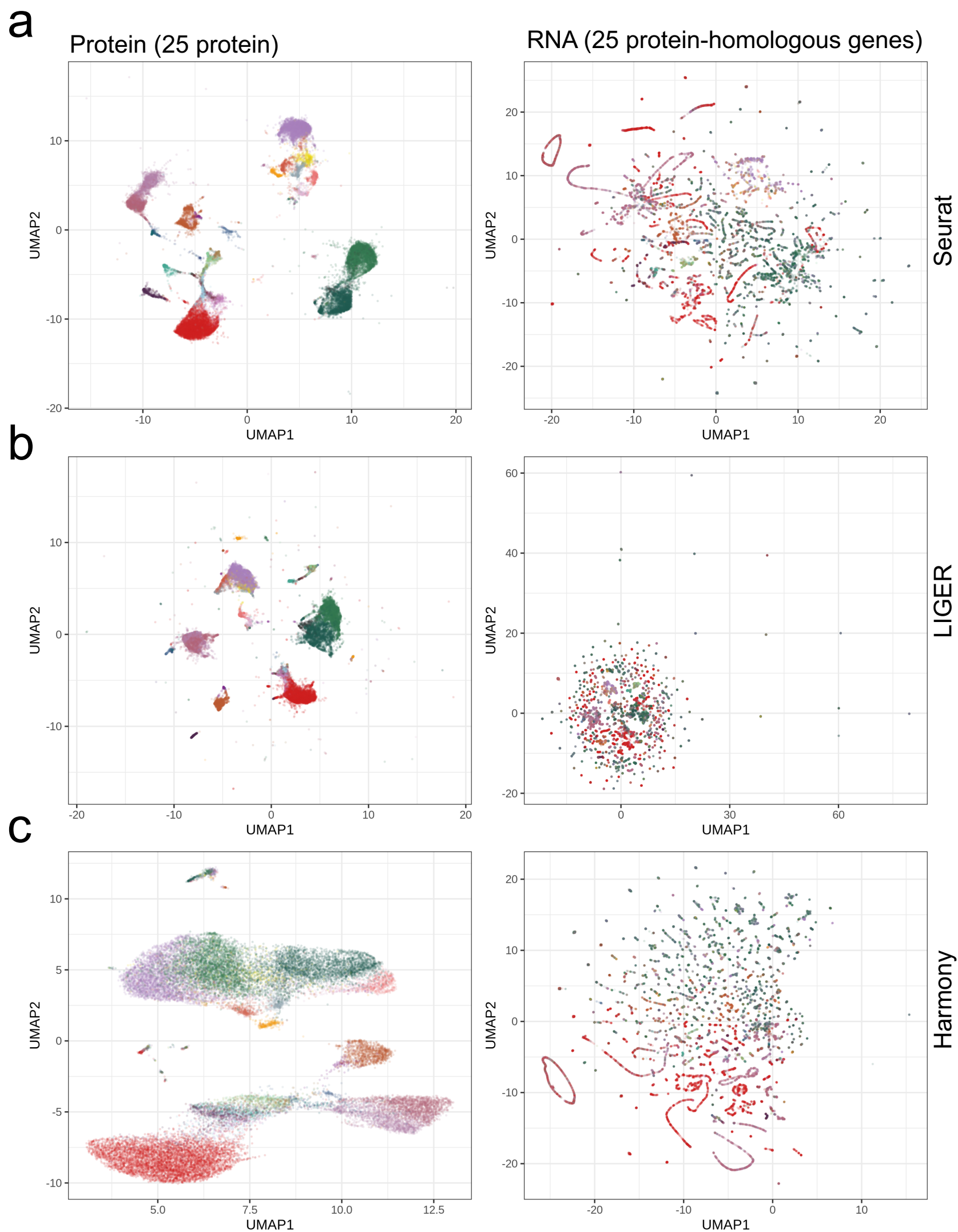


Fig. S15. Performance of three methods on integration of transcriptomic and proteomic data. The cluster colors for each modality are consistent with those in **Fig. 6**.

Table S1. Summary of datasets evaluated in this study. Also listed are the key parameters for running bindSC, Seurat, LIGER, and Harmony on each dataset.

Dataset	Simulation	A549 cell-line data	Mouse skin data	Mouse retina cell atlas	Mouse frontal cortex data	Human bone marrow data
Main results	Fig. 2	Fig. 3	Fig. S7	Fig. 4	Fig.5	Fig. 6
Task	Theoretical validation	scRNA&scATAC	scRNA&scATAC	scRNA&scATAC	scRNA&ST	scRNA&protein
Multi-omics	N.A.	sci-CAR	SHARE-seq	10x Genomics Multiome ATAC+RNA kit	10x Visium	CITE-seq
X (dimension)	1,000x(1,000~10,000)	6,005x4,759	25,594x4,894	9,383x6,275	14,249x2,316	30,672x25
Y(dimension)	1,000x(1,000~10,000)	3,628x24,953	25,594x74,161	9,383x59,353	1,072x2,316	30,672x3,000
Z0 (dimension)	1,000x(1,000~10,000)	3,628x4,759	25,594x4,894	9,383x6,275	1,072x2,316	30,672x25
bindSC						
E	5	5	15	15	15	10
a	0.9	0.3	0.3	0.3	0.1	0.5
delta	0.01 (default)	0.01 (default)	0.01 (default)	0.01 (default)	0.01 (default)	0.01 (default)
Seurat reductions	CCA	CCA	CCA	CCA	CCA	CCA
k (CCA)	5	5	15	15	15	10
LIGER						
k (optimizeALS)	5	5	15	15	15	10
lamda	5 (default)	5 (default)	5 (default)	5 (default)	5 (default)	5 (default)
Harmony dim.use	5	5	15	15	15	10

Table S2. Simulation results with 5,000 cells.

		Alignment mixing score					Silhouette score				
SNR	MR	bindSC	Seurat	LIGER	Harmony	CCA	bindSC	Seurat	LIGER	Harmony	CCA
0	0	0.69	0.69	0.69	0.67	0.69	0.81	0.9	0.07	0.87	0.83
0	0.1	0.69	0.69	0.67	0.67	0.69	0.81	0.88	0.04	0.85	0.83
0	0.2	0.69	0.69	0.69	0	0.69	0.82	0.87	0.08	0.6	0.82
0	0.3	0.69	0.69	0.69	0	0.69	0.82	0.86	0.06	0.38	0.82
0	0.4	0.69	0.69	0.67	0	0.69	0.84	0.87	0.04	0.29	0.82
0	0.5	0.69	0.69	0.67	0	0.69	0.81	0.89	0.03	0.36	0.66
0	0.6	0.69	0.69	0.69	0	0.67	0.83	0.88	0.06	0.22	0.53
0	0.7	0.69	0.67	0.69	0	0.67	0.82	0.77	0.01	0.23	0.5
0	0.8	0.67	0.61	0.67	0	0.67	0.77	0.7	0	0.2	0.47
0	0.9	0.67	0.5	0.67	0	0.65	0.75	0.5	0	0.24	0.38
0	1	0.65	0.5	0.67	0	0.67	0.12	0.24	-0.02	0.18	0.29
0.25	0	0.69	0.69	0.69	0.2	0.69	0.81	0.74	0	0.51	0.74
0.25	0.1	0.69	0.67	0.67	0	0.69	0.81	0.63	0.01	0.51	0.71
0.25	0.2	0.69	0.67	0.67	0	0.67	0.81	0.63	0.02	0.32	0.63
0.25	0.3	0.69	0.67	0.69	0	0.67	0.82	0.67	0.02	0.27	0.63
0.25	0.4	0.69	0.67	0.67	0	0.69	0.83	0.62	0.03	0.21	0.59
0.25	0.5	0.69	0.65	0.67	0	0.67	0.81	0.64	0.01	0.07	0.45
0.25	0.6	0.69	0.61	0.69	0	0.67	0.81	0.51	0.02	0.03	0.36
0.25	0.7	0.69	0.56	0.69	0	0.67	0.8	0.44	0.01	0.01	0.41
0.25	0.8	0.67	0.61	0.69	0	0.67	0.79	0.47	0	0.07	0.35
0.25	0.9	0.67	0.56	0.67	0	0.67	0.8	0.3	-0.01	0	0.3
0.25	1	0.61	0.56	0.69	0	0.67	0.07	0.2	-0.02	-0.01	0.27
0.5	0	0.69	0.56	0.69	0	0.67	0.82	0.29	0	0.26	0.55
0.5	0.1	0.69	0.56	0.67	0	0.67	0.83	0.35	0	0.22	0.52
0.5	0.2	0.69	0.56	0.69	0	0.67	0.81	0.34	0.01	-0.01	0.47
0.5	0.3	0.69	0.5	0.69	0	0.67	0.83	0.27	0.02	0	0.41
0.5	0.4	0.69	0.56	0.69	0	0.67	0.83	0.27	0.01	0	0.39
0.5	0.5	0.69	0.56	0.67	0	0.67	0.82	0.23	0	-0.01	0.37
0.5	0.6	0.69	0.56	0.69	0	0.67	0.81	0.24	0.01	-0.01	0.3
0.5	0.7	0.69	0.56	0.69	0	0.67	0.81	0.27	-0.02	-0.01	0.33
0.5	0.8	0.67	0.5	0.67	0	0.67	0.78	0.24	-0.03	0.02	0.29
0.5	0.9	0.67	0.5	0.69	0	0.65	0.8	0.23	-0.01	-0.02	0.35
0.5	1	0.65	0.56	0.67	0	0.67	0.13	0.23	0	0	0.27

Table S3. Simulation results with 10,000 cells.

		Alignment mixing score					Silhouette score				
SNR	MR	bindSC	Seurat	LIGER	Harmony	CCA	bindSC	Seurat	LIGER	Harmony	CCA
0	0	0.69	0.69	0.67	0.67	0.69	0.81	0.88	0.44	0.86	0.81
0	0.1	0.69	0.69	0.67	0	0.69	0.81	0.9	0.31	0.83	0.81
0	0.2	0.69	0.69	0.67	0	0.69	0.82	0.89	0.3	0.45	0.81
0	0.3	0.69	0.69	0.69	0	0.69	0.81	0.9	0.21	0.39	0.8
0	0.4	0.69	0.69	0.67	0	0.69	0.82	0.89	0.16	0.4	0.76
0	0.5	0.69	0.67	0.69	0	0.67	0.82	0.9	0.3	0.44	0.71
0	0.6	0.67	0.67	0.69	0	0.67	0.82	0.89	0.15	0.39	0.65
0	0.7	0.67	0.67	0.67	0	0.65	0.81	0.83	0.11	0.31	0.57
0	0.8	0.67	0.42	0.69	0	0.56	0.81	0.68	0.1	0.36	0.47
0	0.9	0.67	0.33	0.67	0	0.61	0.77	0.6	0.05	0.34	0.23
0	1	0.65	0.2	0.67	0	0.56	-0.01	0.25	0.01	0.27	0.14
0.25	0	0.69	0.67	0.67	0	0.69	0.82	0.83	0.39	0.56	0.75
0.25	0.1	0.69	0.67	0.67	0	0.67	0.82	0.82	0.22	0.55	0.72
0.25	0.2	0.69	0.67	0.67	0	0.67	0.82	0.84	0.26	0.22	0.7
0.25	0.3	0.69	0.67	0.67	0	0.67	0.81	0.81	0.23	0.27	0.63
0.25	0.4	0.67	0.65	0.67	0	0.67	0.8	0.75	0.13	0.2	0.56
0.25	0.5	0.69	0.65	0.67	0	0.65	0.82	0.73	0.19	0.21	0.52
0.25	0.6	0.67	0.61	0.67	0	0.67	0.82	0.71	0.01	0.22	0.49
0.25	0.7	0.67	0.42	0.67	0	0.65	0.82	0.63	0.03	0.11	0.44
0.25	0.8	0.67	0.33	0.67	0	0.67	0.81	0.48	0.08	0.05	0.43
0.25	0.9	0.65	0.33	0.67	0	0.65	0.79	0.14	0.01	0.08	0.29
0.25	1	0.65	0.33	0.67	0	0.65	0.03	0.33	-0.01	0.1	0.28
0.5	0	0.69	0.42	0.67	0	0.67	0.82	0.42	0.29	0.34	0.55
0.5	0.1	0.67	0.5	0.67	0	0.67	0.83	0.45	0.1	0	0.53
0.5	0.2	0.69	0.5	0.67	0	0.67	0.81	0.41	0.18	-0.01	0.5
0.5	0.3	0.69	0.33	0.67	0	0.65	0.83	0.43	0.1	-0.01	0.5
0.5	0.4	0.69	0.42	0.67	0	0.65	0.82	0.41	0.07	0	0.46
0.5	0.5	0.67	0.42	0.67	0	0.65	0.82	0.38	0.11	0	0.42
0.5	0.6	0.67	0.33	0.67	0	0.65	0.82	0.32	0.01	0	0.37
0.5	0.7	0.67	0.33	0.69	0	0.65	0.81	0.32	0	-0.02	0.35
0.5	0.8	0.67	0.33	0.69	0	0.67	0.82	0.37	0.06	-0.01	0.33
0.5	0.9	0.5	0.33	0.67	0	0.65	0.74	0.29	0.02	-0.01	0.3
0.5	1	0.2	0.33	0.67	0	0.65	0.32	0.3	0.02	0	0.29