

Supplemental information

hECA: The cell-centric assembly of a cell atlas

Sijie Chen, Yanting Luo, Haoxiang Gao, Fanhong Li, Yixin Chen, Jiaqi Li, Renke You, Minsheng Hao, Haiyang Bian, Xi Xi, Wenrui Li, Weiyu Li, Mingli Ye, Qiuchen Meng, Ziheng Zou, Chen Li, Haochen Li, Yangyuan Zhang, Yanfei Cui, Lei Wei, Fufeng Chen, Xiaowo Wang, Hairong Lv, Kui Hua, Rui Jiang, and Xuegong Zhang

hECA: the cell-centric assembly of a cell atlas

Supplemental figures

- **Figure S1.** Filtering candidate T cell subpopulations, Related to Figure 2.
- **Figure S2.** CD4/CD8 T cell population definition, Related to Figure 2.
- **Figure S3.** Cell type and organ distribution of CD19 expressed cells, Related to Figure 3.
- **Figure S4.** Cell type and organ distribution of CD22 expressed cells, Related to Figure 3.
- **Figure S5.** Example gene portrait page of PTPRC, Related to Figure 1.
- **Figure S6.** Example cell type portrait page of Fibroblast, Related to Figure 1.
- **Figure S7.** Example organ portrait page of Brain, Related to Figure 1.
- **Figure S8.** Performance of hECA and other batch integration methods on heart label transfer task, Related to STAR Methods.
- **Figure S9.** Using hECA neuron data for label transfer, Related to STAR Methods.
- **Figure S10.** Overview of all cell types from different sample origins, Related to STAR Methods.
- **Figure S11.** Identification of 3 cell states in epithelial cells can also be retrieved by comparing with hECA data, Related to STAR Methods.
- **Figure S12.** Using hECA lung immune cells for label transfer and DEG finding, Related to STAR Methods.

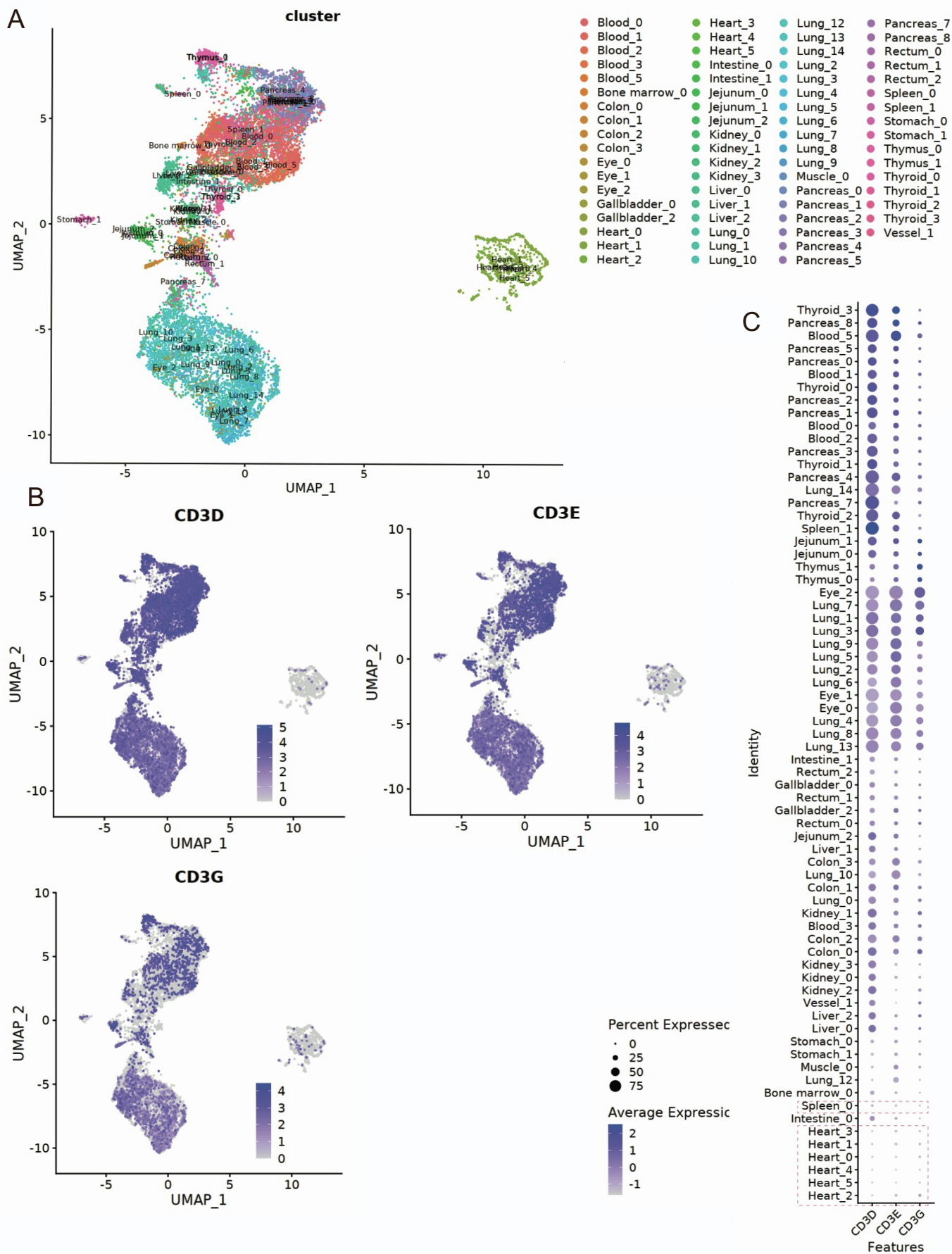
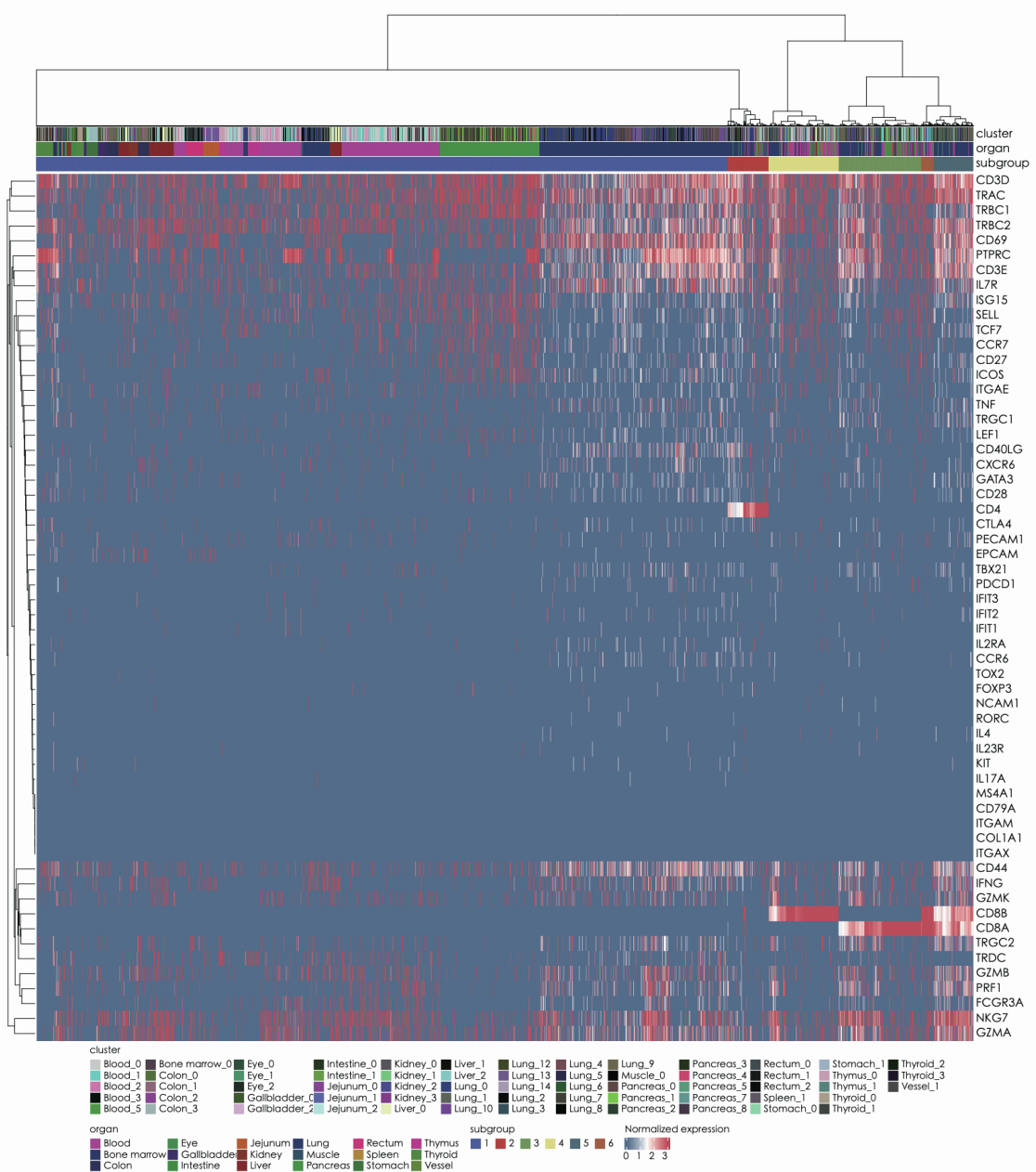


Figure S1. Filtering candidate T cell subpopulations, Related to Figure 2. (A) A UMAP showing the within-organ clustering results. (B) The general T cell markers' expressions (CD3D, CD3E, CD3G). (C) Per-cluster marker gene expressions.

A



B

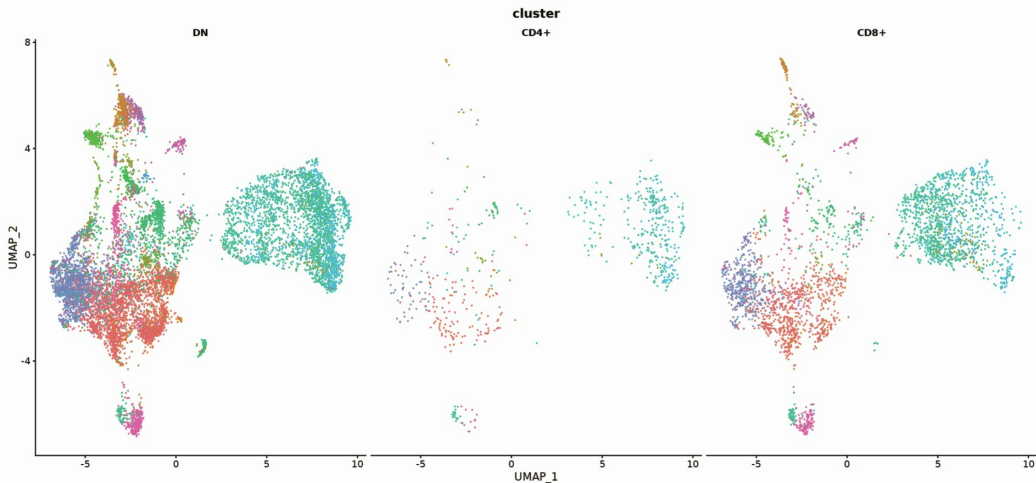


Figure S2. CD4/CD8 T cell population definition, Related to Figure 2. (A) A heatmap showing the hierarchical clustering results based on the CD4/CD8A/CD8B genes. Other T cell signature genes are also listed on the heatmap. (B) Split view of CD4 positive T cells, CD8 positive T cells, and double-negative T cells (Note that sequencing dropouts caused some falsely recognized double-negative T cells).

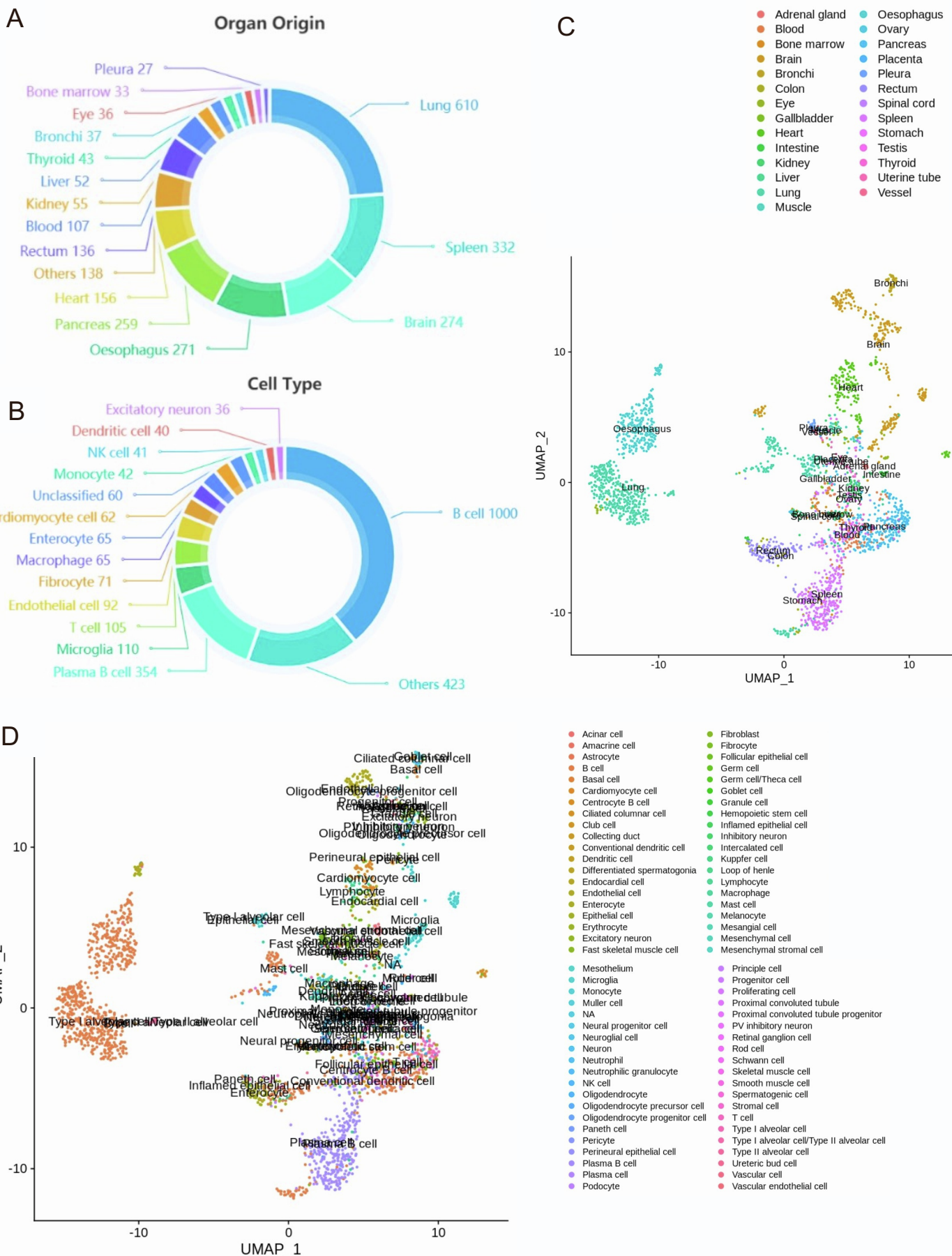


Figure S3. Cell type and organ distribution of CD19 expressed cells, Related to Figure 3. The number of CD19 expressed cells from different organs (A) and cell types (B). Visualization of CD19 expressed cells in UMAP labeling organs (C) and cell types (D).

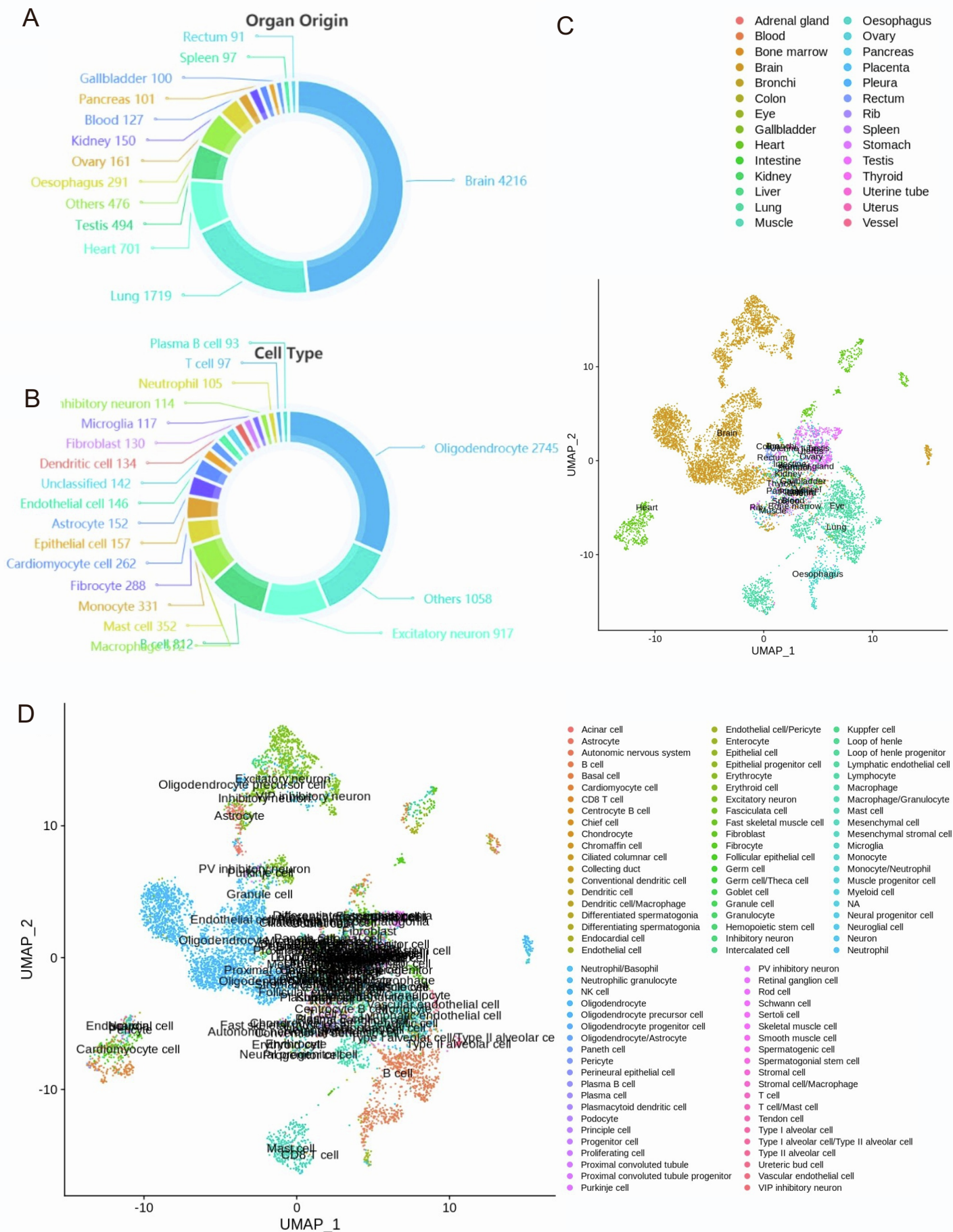


Figure S4. Cell type and organ distribution of CD22 expressed cells, Related to Figure 3. The number of CD22 expressed cells from different organs (A) and cell types (B). Visualization of CD22 expressed cells in UMAP labeling organs (C) and cell types (D).

Home > Gene Portrait

Gene Portrait

Select a Gene

Gene Symbol :

Group By :

Select Organs :
Organs : 294 organs selected

[Submit](#)

Gene Expression Profiling

Please enter search criteria

Sort by: [download](#)

PTPRC

Organ	% Expr
Rectum	3.95%
Ileum	3.35%
Jejunum	4.57%
Spinal cord	3.99%
Thyroid	11.22%
Stomach	2.39%
Intestine	1.49%
Blood	20.48%
Thymus	7.29%
Liver	8.74%
Pancreas	5.25%
Rib	0.42%
Uterine tube	4.37%
Prostate	0.98%
Adrenal gland	1.87%
Gallbladder	11.84%
Adipose	12.11%
Muscle	1.56%
Skin	0.94%
Bone marrow	6.82%
Spleen	9.99%
Uterus	1.96%
Ovary	0.64%
Pleura	7.45%
Colon	3.15%
Placenta	6.16%
Kidney	2.81%
Vessel	7.41%
Testis	0.11%
Bronchi	2.53%
Oesophagus	1.95%
Lung	33.53%
Eye	1.65%
Heart	2.64%
Brain	5.76%
Bladder	0.00%
Duodenum	0.00%
Ureter	0.00%

About the Gene

Full Name
Protein Tyrosine Phosphatase Receptor Type C

Aliases
T200; Receptor-Type Tyrosine-Protein Phosphatase C; CD45 Antigen; GPI80; CD45; L-CA; LCA; Protein Tyrosine Phosphatase, Receptor Type, C Polypeptide; T200 Leukocyte Common Antigen; Leukocyte Common Antigen; T200 Glycoprotein; EC 3.1.3.48; CD45R; PTPRC; B220; LY5

Location
chr1:198,638,457-198,757,476; chr1:198,608,170-198,726,605; chr1:198,607,801-198,726,545

Known as markers of
[CD8 T cell](#) [Common myeloid progenitor\(CMP\)](#) [NK cell](#) [Innate lymphoid cell\(ILC\)](#) [T cell](#) [Eosinophilic granulocyte](#) [Adaptive lymphoid cell](#)

Description
The protein encoded by this gene is a member of the protein tyrosine phosphatase (PTP) family. PTPs are known to be signaling molecules that regulate a variety of cellular processes including cell growth, differentiation, mitosis, and oncogenic transformation. This PTP contains an extracellular domain, a single transmembrane segment and two tandem intracytoplasmic catalytic domains, and thus is classified as a receptor type PTP. This PTP has been shown to be an essential regulator of T- and B-cell antigen receptor signaling. It functions through either direct interaction with components of the antigen receptor complexes, or by activating various Src family kinases required for the antigen receptor signaling. This PTP also suppresses JAK kinases, and thus functions as a regulator of cytokine receptor signaling. Alternatively spliced transcripts variants of this gene, which encode distinct isoforms, have been reported. [provided by RefSeq, Jun 2012]

Link
[GeneCard](#), [NCBI](#), [Ensembl](#), [Wikigenes](#).

Figure S5. Example gene portrait page of PTPRC, Related to Figure 1. Basic information about PTPRC and distribution of expression level grouped by organs are shown in the “About the Gene” and “Gene Expression Profiling” panel, respectively. Users can select which gene to visualize and group by either organ or cell types in the “Select a Gene” panel.

Fibroblast

"A connective tissue cell which secretes an extracellular matrix rich in collagen and other macromolecules. Flattened and irregular in outline with branching processes; appear fusiform or spindle-shaped."

[<http://en.wikipedia.org/wiki/Fibroblast>, ISBN:0517223651, MESH:A11.329.228, MESH:D005347]

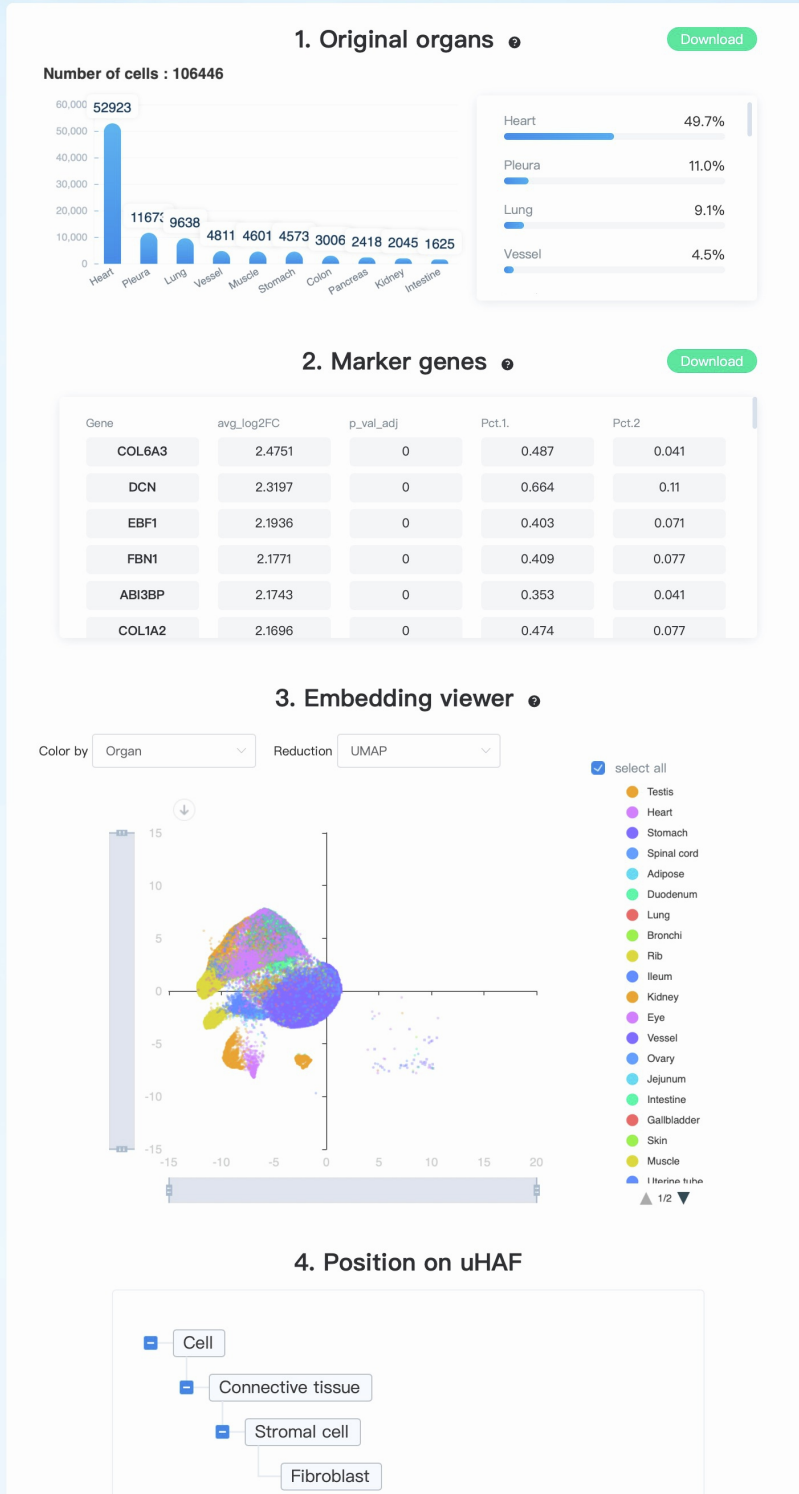
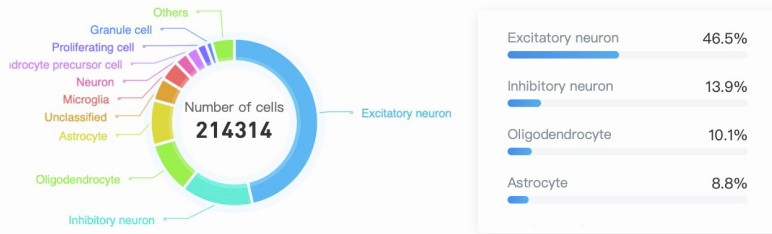


Figure S6. Example cell type portrait page of Fibroblast, Related to Figure 1. A brief description of Fibroblast is provided under the title. “Original organs” section shows the number and proportion of fibroblast cells in each organ. “Marker genes” section lists the differentially expressed genes in fibroblast and their characteristic expression ranges. “Embedding viewer” section visualizes the distribution of cells on 2D DensMAP space, colored by organs. It also provides 2D PCA and UMAP visualization of cells colored by sequencing technology, data source or expression level of certain gene. “Position on uHAF” marks the position of Fibroblast on the unified hierarchical annotation framework (uHAF).

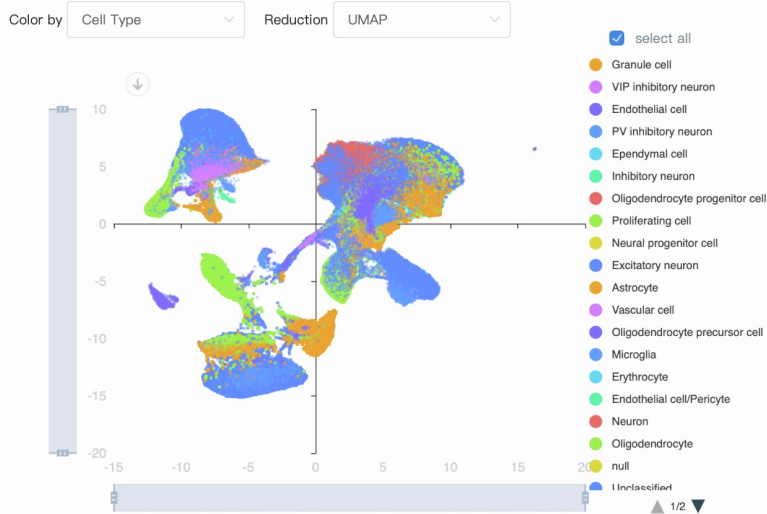
Brain

A brain is an organ that serves as the center of the nervous system. In a human, the cerebral cortex contains approximately 14–16 billion neurons, and the estimated number of neurons in the cerebellum is 55–70 billion. The function of brain includes perception, motor control, homeostasis, learning and memory, etc.

1. Cell type composition

[Download](#)


2. Embedding viewer



3. Anatomy relationship

Belongs to the system: **Nervous**

Other related organs:



4. Position on uHAF

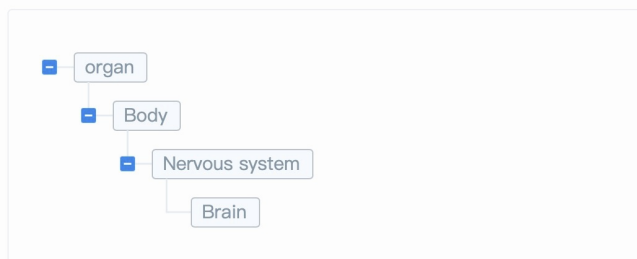
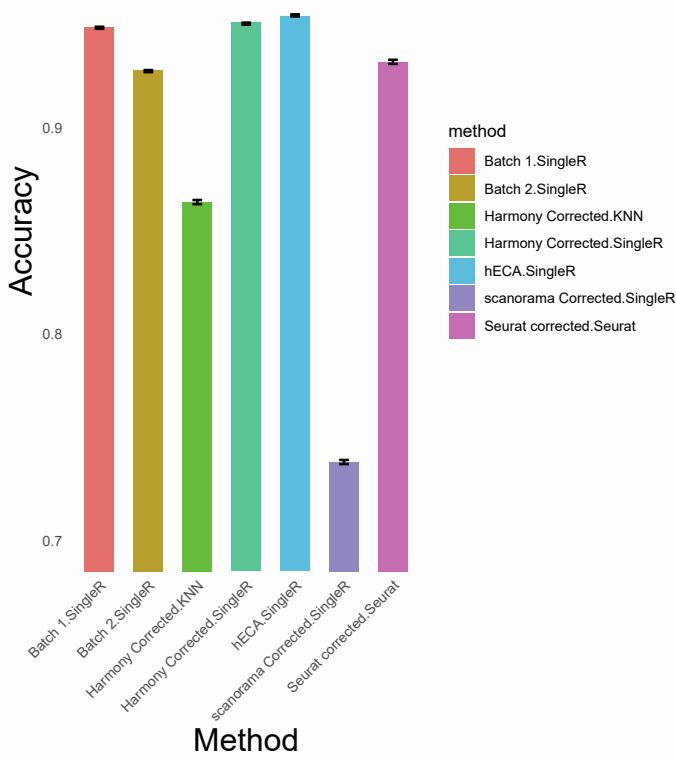
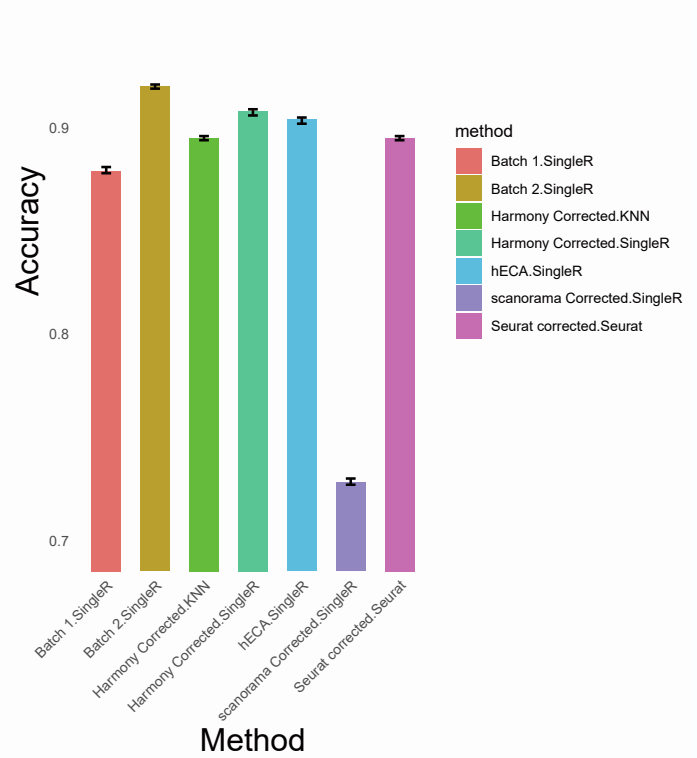


Figure S7. Example organ portrait page of Brain, Related to Figure 1. A brief description of Brain is provided under the title. “Cell type composition” section shows the number and proportion of brain cells in each cell type. “Embedding viewer” section visualizes the distribution of cells on 2D UMAP space, colored by cell types. It also provides 2D PCA and DensMAP visualization of cells colored by sequencing technology, data source or expression level of certain gene. “Anatomy relationship” shows the brain belongs to the nervous system and is related to eye and skin. “Position on uHAF” marks the position of Brain on the unified hierarchical annotation framework (uHAF).

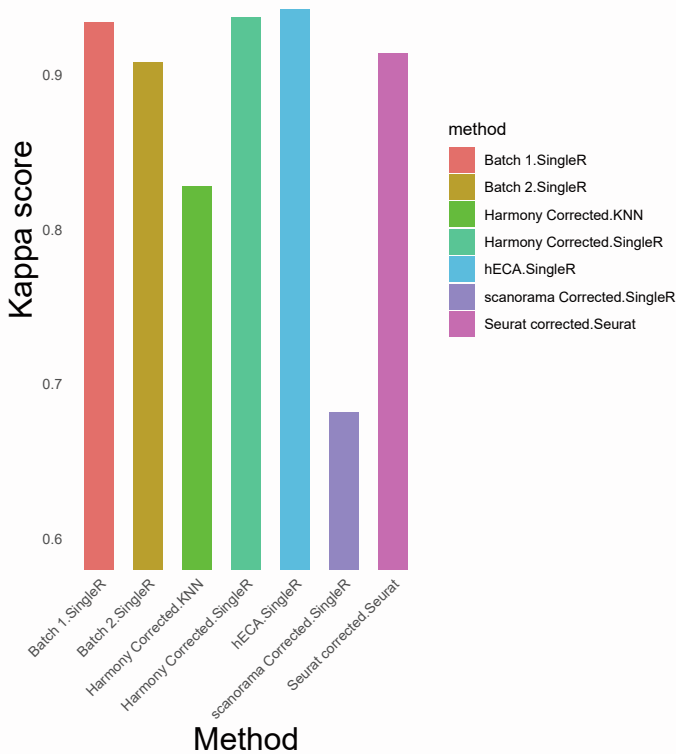
A) Accuracy on Litvi..uková et al. Nature, 2020



B) Accuracy on Tucker et al. Circulation, 2020



C) Kappa score on Litvi..uková et al. Nature, 2020



D) Kappa score on Tucker et al. Circulation, 2020

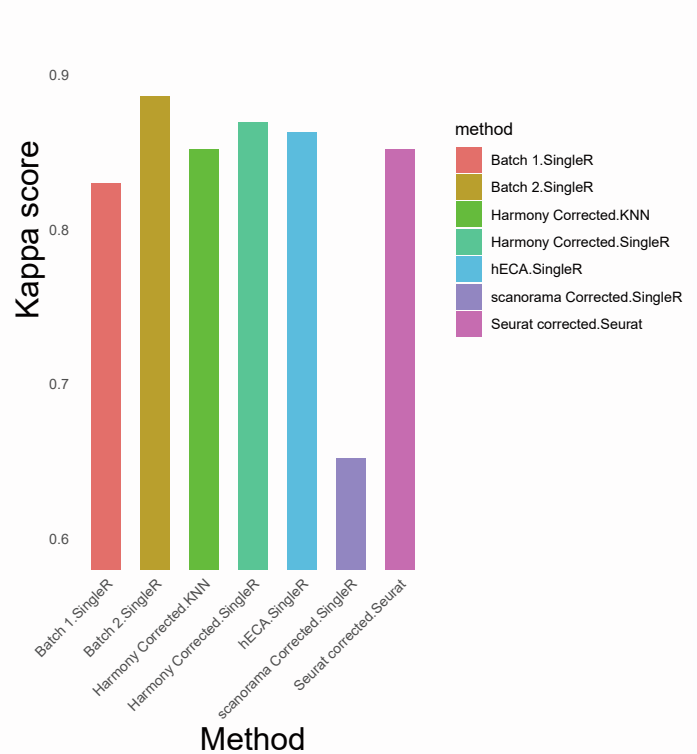


Figure S8. Performance of hECA and other batch integration methods on heart label transfer task, Related to STAR Methods. (A) The accuracy of 6 experiment on (Litviňuková et al., 2020) (B) The accuracy of 6 experiment on (Tucker et al., 2020). (C) The kappa score of 6 experiment on (Litviňuková et al., 2020) (D) The kappa score of 6 experiment on (Tucker et al., 2020).

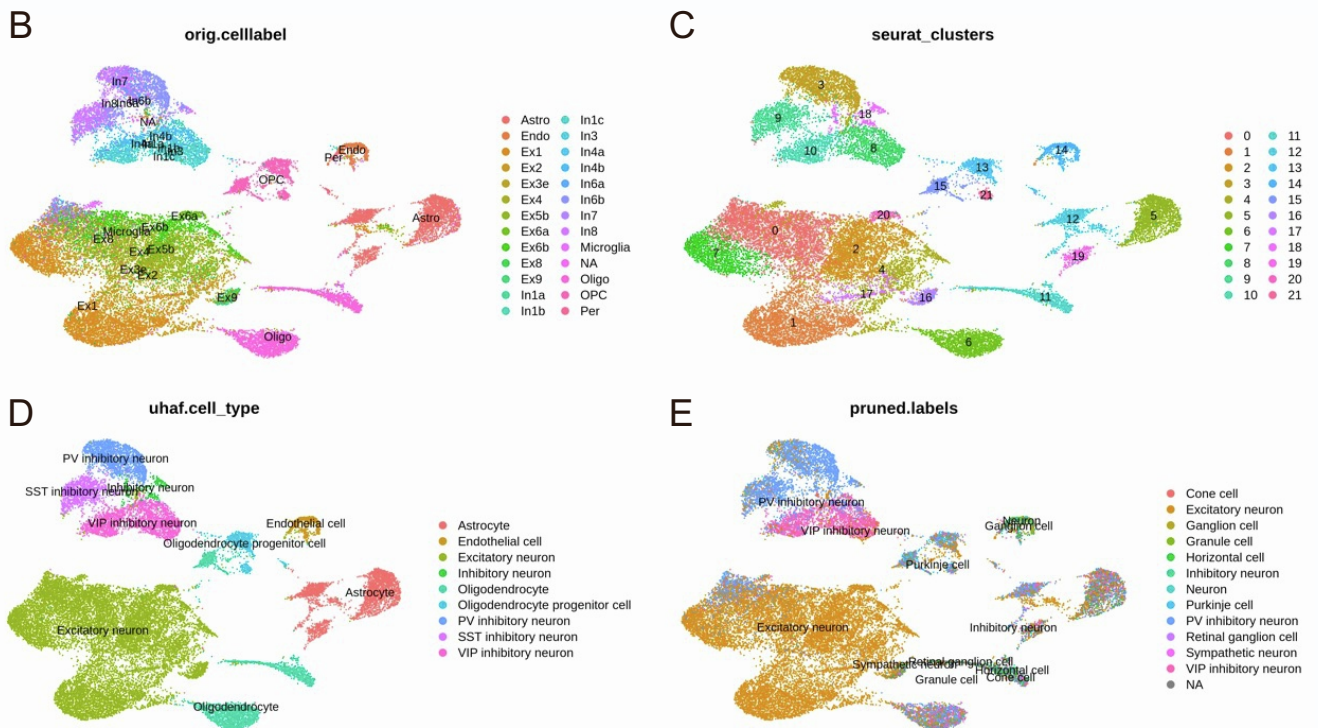
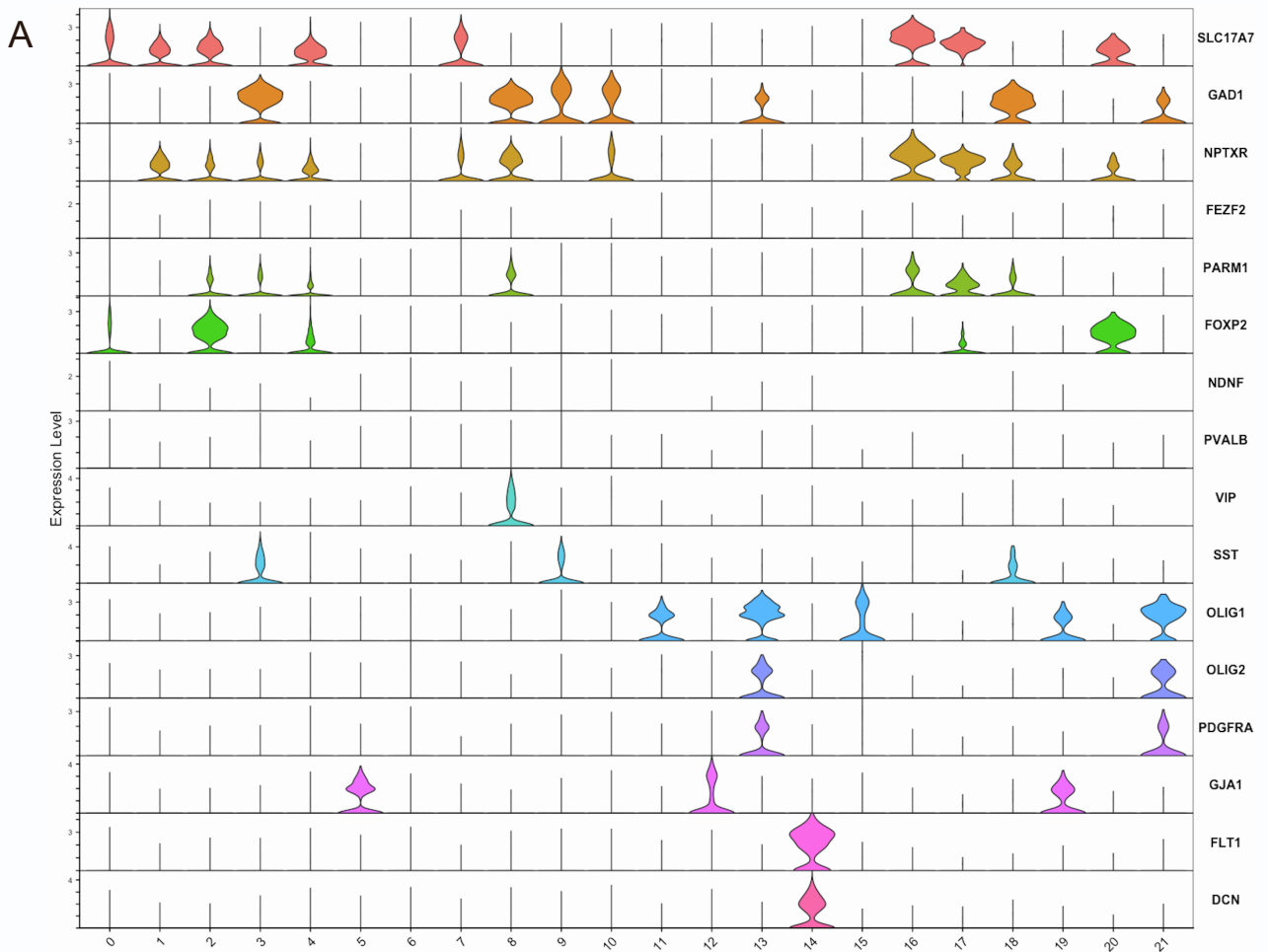


Figure S9. Using hECA neuron data for label transfer. (A) Well-known cell type markers used to determine the cell type, Related to STAR Methods. UMAP of query cells with the original cell type in uHAF name (B), Seurat cluster ID (C), uHAF cell type by manually annotation (D), and transferred labels predicted by SingleR (E).

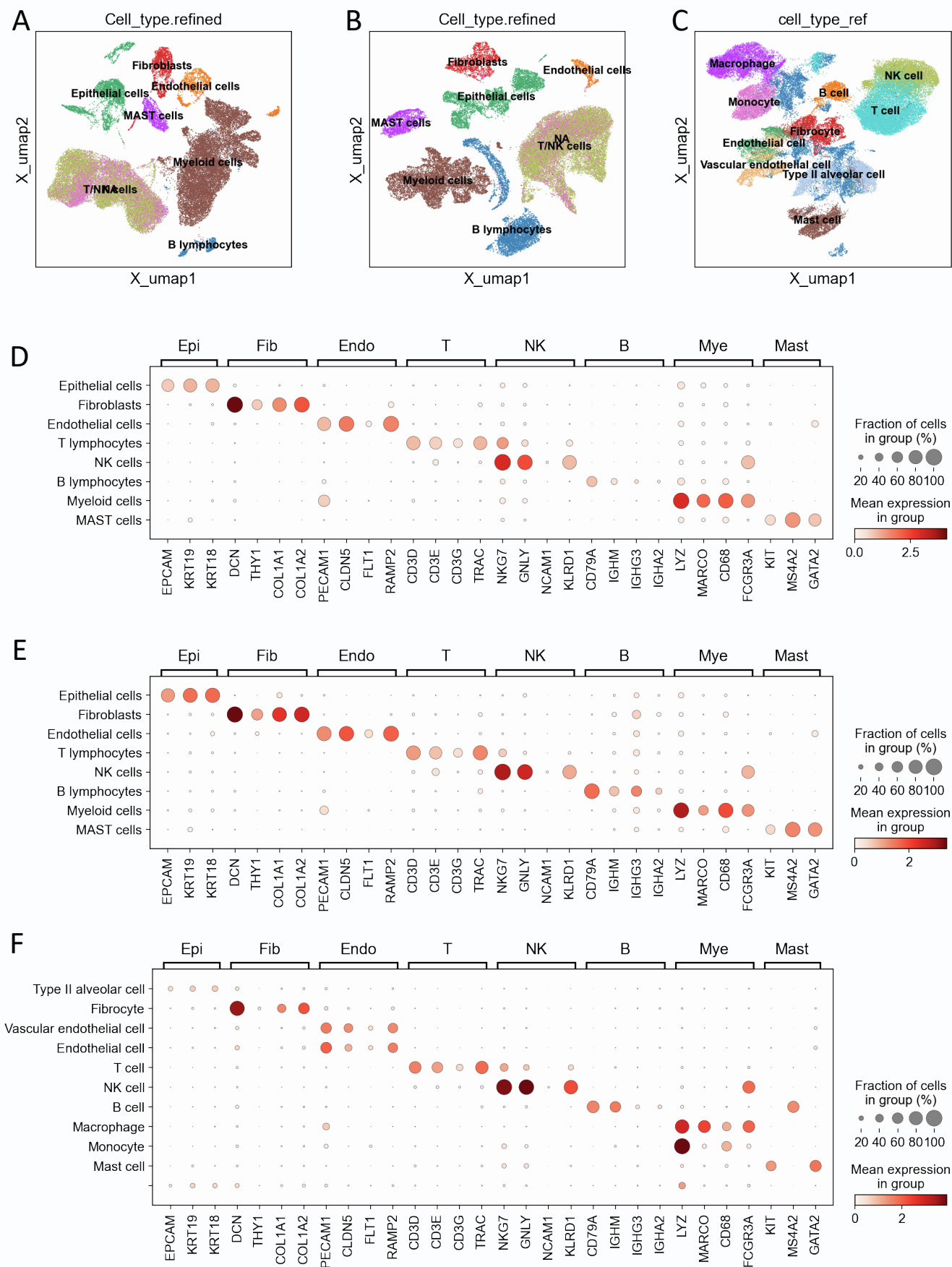


Figure S10. Overview of all cell types from different sample origins, Related to STAR Methods. (A-C) UMAP projection for normal lung (nLung), tumor lung (tLung) from paper and normal lung from hECA database (hECA). (D-F) Dotplot shows marker genes for major cell types in nLung, tLung and hECA.

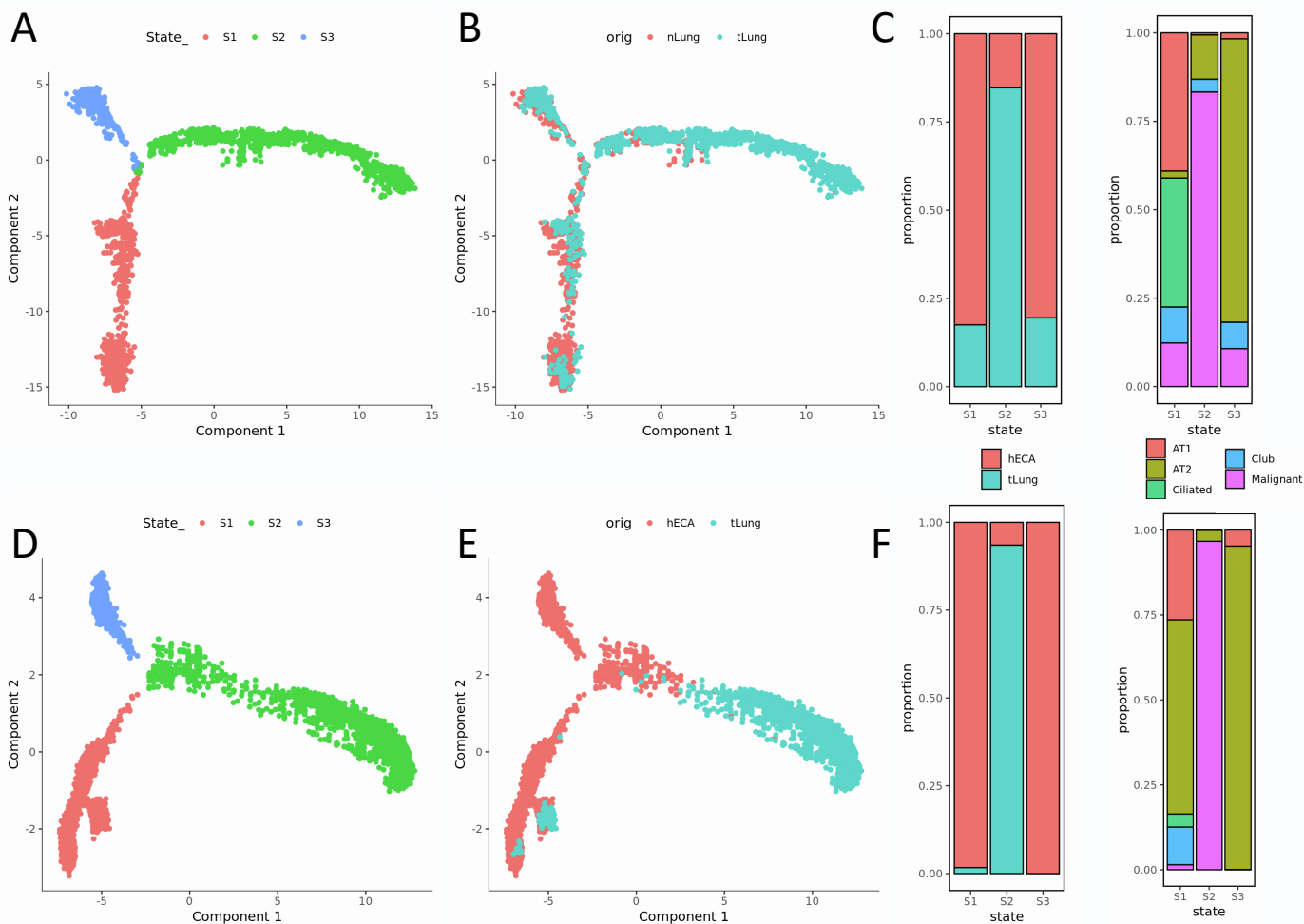
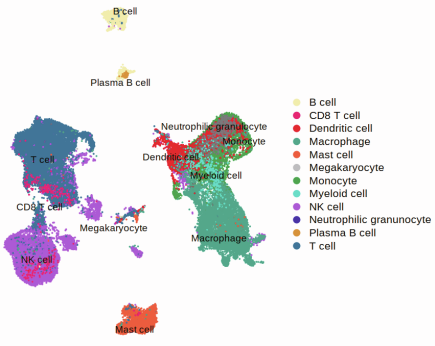


Figure S11. Identification of 3 cell states in epithelial cells can also be retrieved by comparing with hECA data, Related to STAR Methods. (A-B) monocle trajectory plot of epithelial cells from nLung and tLung, colored by cell states (A) and cell origins (B). (C) composition of cell origins for three cell states in A and B. (D-E) monocle trajectory plot of epithelial cells from hECA and tLung, colored by cell states (D) and sample origin (E). (F) composition of cell origins for three cell states in D and E.

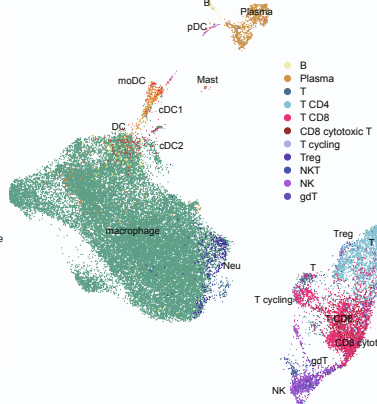
(A)

#	Reference Data	Methods	ACC	ARI
1	hECA batch1: Microwell-seq lung immune	Seurat v3 Label Transfer	82.84%	0.31
2	hECA batch2: 10xGenomics lung immune	Seurat v3 Label Transfer	92.53%	0.41
3	hECA unintegrated: lung immune	Seurat v3 Label Transfer	92.05%	0.37
4	hECA integrated: lung immune	Seurat v3 CCA integration + Label Transfer	92.50%	0.38

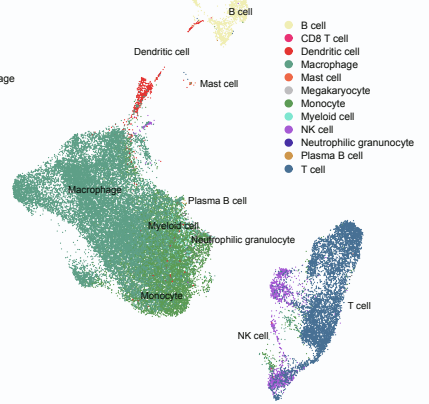
(B) Lung immune cells (hECA)



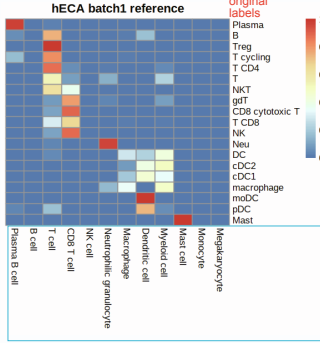
(C) Lung immune cells (COVID)



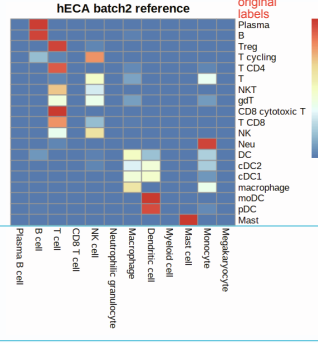
(D) hECA transferred Label



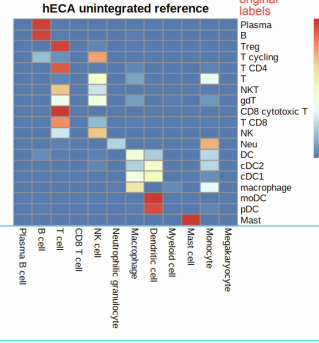
(E)



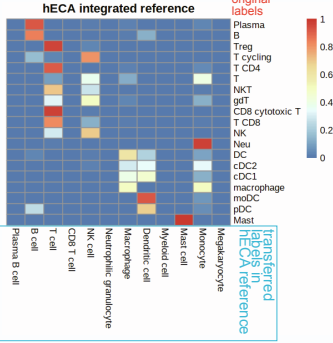
(F)



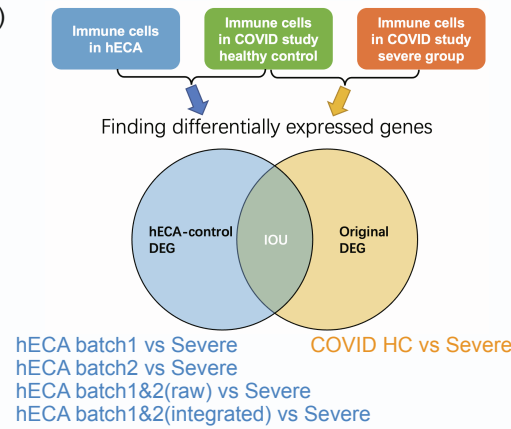
(G)



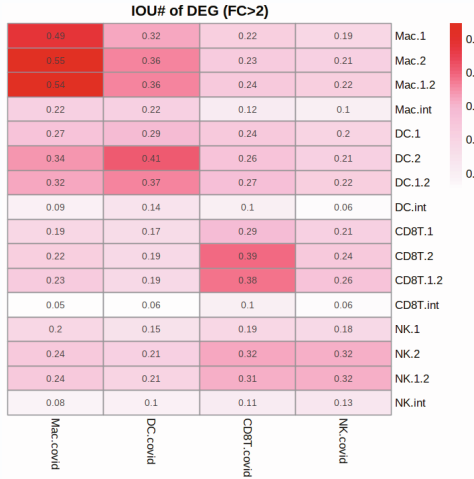
(H)



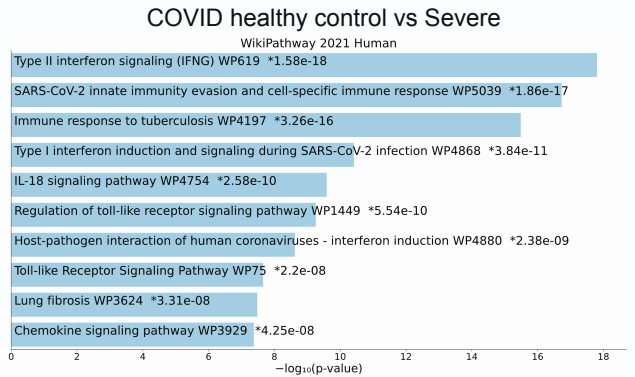
(I)



(J)



(K)



(L)

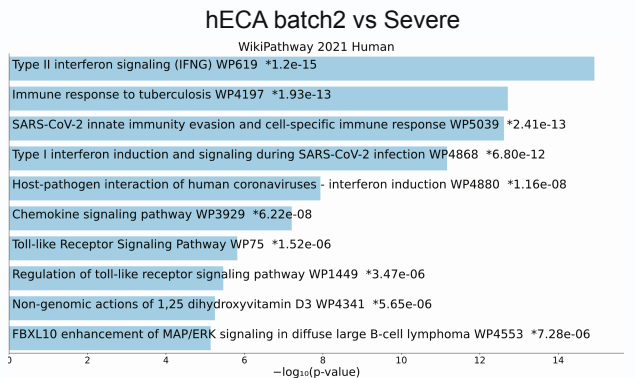


Figure S12. Using hECA lung immune cells for label transfer and DEG finding, Related to STAR Methods. (A) Label transfer experiment designs and performances. ACC: label transfer accuracy. ARI: adjusted Rand index (B) UMAP visualization of hECA lung immune cell-type labels. (C) UMAP visualization of a COVID-19 lung immune cell-type labels (GSE145926). COVID-19: coronavirus disease-2019 (D) UMAP visualization of cell-type labels transferred from hECA to the COVID-19 dataset. (E-H) Fractions of cells mapped to hECA labels in different cell types. The right row indices represent the labels derived from the original COVID-19 study. The bottom column indices represent the labels transferred from hECA. (I) Designs of the DEG (differentially expressed genes) finding experiments. We replaced the internal HC (healthy control) group in the COVID study with external hECA healthy control data and evaluated the consequences. (J) Consistency between the hECA-control DEGs and original DEGs. IOU: Intersection over union, the number of shared DEGs in two compared conditions divided by the number of total DEGs. (K-L) Pathway enrichment analysis of the original DEGs and the hECA-control DEGs.