

A reference single-cell regulomic and transcriptomic map of cynomolgus monkeys

Jiao Qu^{1,#}, Fa Yang^{1,#}, Tao Zhu^{1,#}, Yingshuo Wang^{2,#}, Wen Fang¹, Yan Ding¹, Xue Zhao¹, Xianjia Qi³, Qiangmin Xie², Qiang Xu¹, Yicheng Xie^{2,*}, Yang Sun^{1,4,*} and Dijun Chen^{1,*}

¹State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210023, China;

²The Children's Hospital, Zhejiang University School of Medicine, National Clinical Research Center for Child Health, Hangzhou 310052, China;

³Shanghai XuRan Biotechnology Co., Ltd., 1088 Zhongchun Road, Shanghai 201109, China;

⁴Chemistry and Biomedicine Innovation Center (ChemBIC), Nanjing University, Nanjing 210023, China.

[#]These authors contributed equally to this work.

*Correspondence: dijunchen@nju.edu.cn (Dijun Chen), yangsun@nju.edu.cn (Yang Sun) or ycxie@zju.edu.cn (Yicheng Xie)

Abstract

Non-human primates (NHP) are attractive laboratory animal models that accurately reflect both developmental and pathological features of humans. Here we present a compendium of cell types from the cynomolgus monkey *Macaca fascicularis* (denoted as ‘Monkey Atlas’) using both single-cell chromatin accessibility (scATAC-seq) and RNA sequencing (scRNA-seq) data at the organism-wide level. The integrated cell map enables in-depth dissection and comparison of molecular dynamics, cell-type composition and cellular heterogeneity across multiple tissues and organs. Using single-cell transcriptomic data, we inferred pseudotime cell trajectories and cell-cell communications to uncover key molecular signatures underlying their cellular processes. Furthermore, we identified various cell-specific *cis*-regulatory elements and constructed organ-specific gene regulatory networks at the single-cell level. Finally, we performed a comparative analysis of single-cell landscapes among mouse, cynomolgus monkey and human, and we showed that cynomolgus monkey has significantly higher degree of cell-type similarity to human than mouse. Taken together, our study provides a valuable resource for NHP cell biology.

Key words: Non-human primate; Cell atlas; scRNA-seq; scATAC-seq

Introduction

Non-human primates (NHP) are phylogenetically close to humans and show various human-like characteristics, including genetics, organ development, physiological function, pathological response and biochemical metabolism. Hence NHP are extremely valuable as experimental animal models in medical research and drug development¹. Since cells are the fundamental unit of all life, direct comparison of cell identities and cell-type compositions between organisms across organs would help to transfer knowledge in primates to medical research. In this regard, it is of vital importance to understand the cellular composition and heterogeneity of primate organs.

Rapid advances in single-cell multi-omics technologies have enabled molecular quantification of thousands of cells at once, leading to meticulous insight into organ compositions and mechanisms driving cellular heterogeneity². Previous studies³⁻⁷ have mapped the single-cell landscapes across multiple organs in humans and mice, expanding our knowledge about the cellular heterogeneity underlying normal development and aging. Three-dimensional multicellular culture systems combined with single-cell transcriptome sequencing technology enables to chart the cellular and molecular dynamic changes during organ growth and development⁸⁻¹⁰. In addition, extensive efforts⁹⁻¹² have been achieved to investigate how cells are perturbed in various disease conditions, including cancer and neurological disorders.

Mice have long been used as a representative model organism for mammalian development and physiology. Recently, extensive comparative analyses based on single-cell transcriptomics data have shown that both cell types and associated gene regulatory networks are conserved between human and mouse¹³⁻²⁰, which provides a new perspective for explaining disease mechanisms and finding targets for disease intervention. However, it has been widely recognized that there are significant differences between mice and humans in terms of development and physiology²¹. From genetic perspective, primate experiments are more useful as they can better simulate human diseases and promote scientific research owing to high genetic similarity between primates and humans²². Although the potential

importance and values of NHP models in basic research are indispensable, an organism-wide single-cell atlas is still pending for primates. Here we present a compendium of single-cell regulomic and transcriptomic data from *Macaca fascicularis* (cynomolgus monkeys) that comprises 40 distinct cell types from 16 organs and tissues, greatly extending our current view²³⁻²⁵ of single cell landscapes in this model species. This cell atlas -- which we denote 'Monkey Atlas' -- represents a new resource for NHP cell biology.

Results

Mapping a cynomolgus monkey multi-organ cell atlas by multi-omic analysis

To generate a reference cell map of monkey, we performed both single-cell RNA sequencing (scRNA-seq; 10x Genomics; n=174,233) and scATAC-seq (10x Genomics; n=66,566) on more than 240,000 high-quality cells from 16 tissues and organs in one male or/and one female cynomolgus monkeys (**Fig. 1a** and **Supplementary Fig. 1**). We integrated all of the scRNA-seq data using canonical correlation analysis (CCA)²⁶ to correct for batch effects. Unsupervised clustering based on t-distribution stochastic neighbor embedding (t-SNE) resolved major cell types, including epithelial, ciliated epithelial, mesenchymal, immune, endothelial, spermatid, and secretory cell populations. These cells could be subdivided into 40 transcriptionally distinct clusters with cluster-specific markers (**Fig. 1b,c**). Due to technical and financial constraints, not every organ was analyzed in each monkey or by data modality. Nevertheless, the overall gene expression patterns or cell composition for the same organs or functional related organs (e.g., stomach, liver, spleen and colon from the digestive system) are quite similar (**Supplementary Figs. 2 and 3**); the analysis of multiple organs from the same monkey enable us to obtain data that is controlled for uncertain effects (such as age, sex, diet, environment and so on).

To analyze scATAC-seq cells from the organs of liver, colon, uterus, spleen, lung, heart and kidney, we created a count matrix of fragments across the genome. We demonstrated the overall high quality of scATAC-seq data based on the enrichment analysis of accessible DNA sequences relative to the transcriptional start site (TSS) and the size distribution of unique

fragments (**Supplementary Fig. 4**). T-SNE clustering analysis of scATAC-seq data revealed ten major cell types annotated based on chromatin accessibility at the promoter regions of well-characterized marker genes (**Fig. 1d**). For the organs with matched scRNA-seq and scATAC-seq data, we performed cross-modality integration analysis using mutual nearest neighbors (MNNs) approach (**Supplementary Fig. 5**; see **Methods**). We assigned cell type cluster labels from matched scRNA-seq data to scATAC-seq cells. This revealed that cell types identified by scRNA-seq and scATAC-seq are highly consistent (**Fig. 1e**), highlighting the quality of the dataset.

Epithelial cell heterogeneity and developmental dynamics across organs

Epithelial cells account for the largest part in the integrated cell map. To dissect epithelial heterogeneity, we extracted epithelial cells and performed unsupervised sub-clustering analysis (**Fig. 2a**). The analysis identified 14 cell clusters (E01-E14), including basal cells, secretory cells, ciliated cells and non-ciliated cells, according to distinct pattern of marker gene expression (**Fig. 2a,b**). It is worth noting that we observed a large proportion of ciliated epithelial cells in various tissues (**Fig. 2c** and **Supplementary Fig. 6a,b**). Ciliated cells play an important role in cleaning pathogenic microorganisms and signal transduction²⁷. Gene ontology (GO) enrichment analysis based on differentially expressed genes revealed ciliated cell subpopulations with different biological functions. For example, subpopulations of E09 (*SCGB1D2*^{high} ciliated cells) and E12 (*GLIPR1L1*^{low}) enriched for pathways related to cellular response to stimulus, cell communication and intracellular signal transduction, whereas E05 subpopulation related to RNA biosynthetic process (**Fig. 2d**).

To explore the developmental dynamics of epithelial cells, we performed pseudotime trajectory analysis using both Monocle2²⁸ and RNA velocity²⁹. We determined the cluster E03 (*CYB5A*^{high} secretory cells) as the start point of trajectory based on estimated latent time by RNA velocity (**Fig. 2f**). Accordingly, highly expressed genes in E03 are functional related to ATP metabolic process and purine ribonucleoside monophosphate metabolic process (**Fig. 2g**). Epithelial cells were arranged into a trajectory with two bifurcations and three cell states with E03 as the root (**Fig. 2e**). It is worth noticing that ciliated epithelial cells are in different

states of differentiation. For example, cluster E06 was predominant in bifurcation 2 (**Fig. 2h** and **Supplementary Fig. 6e**). Some marker genes are not in the E06 subgroup, such as *CST6*, *PDZK1IP1*, *KRT19*, and *PSCA*, they had low relative expression in state 3 (**Supplementary Fig. 6d**). From the perspective of sample type, almost all of the epithelial cells of testis tissue are present in bifurcation 2 (major in state3) (**Fig. 2h** and **Supplementary Fig. 6c**). Differential genes such as *GABRR1*, *GRIK2*, *CEP41*, *CFAP20*, *TPGS1* in state3 showed enrichment in presynaptic membrane potential and protein polyglutamylation (**Fig. 2i**). This may be due to the ability of testis to produce sperm and male hormones. We speculate that *GLIPR1L1*^{high} ciliated cells are the main effector epithelial cells of the testis.

Stromal cellular heterogeneity

Stromal cells are an important component of body tissues³⁰. In the stromal compartment, we identified 11 clusters (S01-S11) belonging to four major cell types including endothelial cells, fibroblasts, FibSmo cells and smooth muscle cells (**Fig. 3a-c** and **Supplementary Fig. 7a**). Although these cell clusters were identified in all organ tissues, the heterogeneity of stromal cells was observed in different organs (**Fig. 3d,e** and **Supplementary Fig. 7b**). Most mesenchymal cells were generated from kidney; almost all fibroblasts in testis are from S05 (*DCN*^{high}*APOD*^{high} fibroblasts); there are a large number of *TAGLN*^{high}*MUSTN1*^{low} smooth muscle cells in aorta tissues but few other smooth muscle cells (**Fig. 3d**). Considering that stromal cells have a certain differentiation potential³¹, we applied RNA velocity analysis to explore developing states of stromal cells (**Fig. 3f**). The results show that fibroblasts have the capacity of differentiation to smooth muscle cells and endothelial cells (**Fig. 3g,i** and **Supplementary Fig. 7c-e**). It is worth noting that fibroblasts of cynomolgus monkeys have a strong metabolic ability rather than the ability to synthesize collagen (**Fig. 3h**).

Heterogeneity of immune cells

Immune cells are essential for maintaining body homeostasis³². We identified 72,284 immune-related cells from the investigated organs, including B cells, T cells and myeloid cells, and these cells were further grouped into 13 major clusters (I01-I13) based on known or novel gene signatures (**Fig. 4a,b** and **Supplementary Fig. 8a,b**). Although the annotated

immune cell clusters can be found in all organs (**Fig. 4c,d**), the relative proportion of immune cells varied greatly in different organs. For example, We noticed that NKT_cell_CD3D^{high}_GZMK^{high}_GzmB^{high} cells vary widely in muscle tissues compared to other tissues (**Fig. 4e**). Subsequently, we analyzed differentially expressed genes of NKT_cell_CD3D^{high}_GZMK^{high}_GzmB^{high} cells in muscle tissues. We observed that mitochondria-related genes (ATP6, COX3 and ND1) were the main differentially expressed genes in NKT_cell_CD3D^{high}_GZMK^{high}_GzmB^{high} cells of muscle tissue compared to other tissues (**Fig. 4f** and **Supplementary Fig. 8c**).

Dynamics of cell-cell interactome

To decipher the dynamics of intercellular communications in different tissues, we employed CellPhoneDB to identify potential ligand-receptor pairs among the major cell types. We observed that there are strong intercellular interactions among stromal cells, epithelial cells and myeloid cells (**Fig. 5a,b**). Generally, the intensity and pattern of cellular interactions between cells are tissue-specific (**Fig. 5c**). For example, tongue and uterus tissues show stronger cellular interactions, while intercellular interactions in testicular tissue is weaker than other tissues (**Supplementary Figs. 9** and **10**). To chart the rewiring of molecular interactions regulating cell-cell interactions, we mapped ligand-receptor pairs in specified cell subpopulations in different organs (**Fig. 5d**). In brief, the “CD99-PILRA” ligand-receptor pair is specific in the interaction between stromal cells and myeloid cells, particularly in adipose, aorta, and colon. As an inhibitory receptor of immunoglobulin-like type 2 receptor (PILR), PILRA has been shown to bind to the CD99 ligand for immune regulation³³. The “CCL4L2-VSIR” pair occurred exclusively in the interaction of myeloid cells and T cells. In contrast, the “LGALS9-CD44” pair contributed to most immune cell related interactions. Accordingly, CD44 plays a role in innate immunity and subsequent adaptive responses, and has extensive inflammatory and proliferative effects on a variety of cell types^{34,35}. Taken together, these results predicted the possible molecular mechanisms underlying cell-cell communication in various tissues (**Fig. 5d**).

Single-cell chromatin landscape of major organs in cynomolgus monkey

To deconstruct the gene regulation principles of complex tissues in cynomolgus monkey, we examined the single-cell chromatin accessibility landscape of major organs including colon, kidney, lung, uterus, heart, liver and spleen by scATAC-seq. In total, we generated scATAC-seq profiles from 66,566 cells after quality control. We identified 22 distinct cell clusters in the integrated cell map according to cluster-specific *cis*-elements and visualized single-cell profiles with uniform manifold approximation and projection (UMAP) (**Fig. 6a** and **Supplementary Fig. 11a**). For example, clusters 1-4 demonstrated accessibility at *cis*-elements neighboring B cell genes, including *CD22*, *MS4A1* and *TNFRSF13C*, while the cluster 22 demonstrated accessibility at *cis*-elements neighboring T cell genes, including *CD3D* and *IL7R* (**Supplementary Fig. 12**). We detected 397,773 *cis*-elements across all clusters, ranging from 3,046 to 75,001 peaks in each clusters (**Fig. 6b**). As expected, most of *cis*-elements were derived from promoters, intronic or distal intergenic regulatory regions. We observed that most cell clusters are organ-specific (**Fig. 6c**) and cluster-specific *cis*-elements exhibited organ-specific accessibility accordingly (**Fig. 6d**).

Comparison analysis of scATAC-seq and scRNA-seq data highlighted concordant patterns of chromatin accessibility and gene expression across clusters, exemplified by marker genes (*POU2F2* and *TCF21*) in specific cell types (**Fig. 6e,f**). We also computed TF deviation scores using chromVAR³⁶, which measured the accessibility of TF binding “footprint” genome-wide in each single cell. Indeed, TF deviation scores for *POU2F2*, a B-cell-specific transcription factor involves in cell immune response by regulating B cell proliferation and differentiation^{37,38}, were increased in B cell clusters (**Fig. 6f** and **Supplementary Fig. 11b**). Similarly, the TF deviation scores for *TCF21*, an essential regulator of fibroblasts in development³⁹, were increased in the fibroblasts cell cluster (**Fig. 6f** and **Supplementary Fig. 11b**). Furthermore, we applied Cicero⁴⁰ to identify co-accessible *cis*-elements at genome-wide, as exemplified at the gene locus of *EGFL7* (**Fig. 6g**), an endothelium-specific secreted factor mostly produced by blood vessel endothelial cells during development⁴¹⁻⁴³. We observed increased enhancer-promoter connections in endothelial cell clusters at the promoter of *EGFL7*. Overall, our scATAC-seq data provide a rich resource for unbiased

discovery of cell types and regulatory DNA elements in cynomolgus monkey.

Cell-type specific and organ-specific transcriptional gene regulatory networks

TFs are important regulators controlling cell identity and tissue-specific gene expression. To infer cell-type and organ-specific transcriptional regulatory programs based on the monkey cell transcriptional landscape, we applied SCENIC (single-cell regulatory network inference and clustering) to identify TF regulons. We identified several TF regulon modules that were active in either cell-type (n=8; **Supplementary Fig. 13a,b**) or organ-specific manners (n=6; **Fig. 7a,b**). Subsequently, we analyzed representative TF regulons across different cell types (n=7; **Supplementary Figs. 13c and 15**) or different organs (n=16; **Fig. 7c and Supplementary Fig. 14**). The identified TF regulons are highly cell-type or organ-specific based on regulon activity scores (**Fig. 7d and Supplementary Fig. 13d**). Finally, The representative TF regulons and their associated target genes were organized into cell-type specific or organ-specific gene regulatory networks (**Fig. 7e and Supplementary Fig. 13e**).

In the cell-type specific gene regulatory networks, we observed that CREM specifically controls proliferation-related target genes such as DAZL and HEY2 in spermatid cells. Immune-related TFs such as IRF2, FLI1 and IKZF3 are shown to regulate immune cell identity genes such as S100A4 and CD48. FEV, a known TF that regulates the development of hematopoietic stem cells⁴⁴, extensively link to target genes actively expressed in immune cells and epithelial cells (**Supplementary Fig. 13e**).

In the organ-specific gene regulatory networks, we found that target genes of ZNF770 and CTCF are specifically expressed in tongue. The spermatid cell-specific TF regulon CREM regulated genes actively expressed in testis. ETS-factors (ELK3, ERG, and FLI1) together with pre-/immature-B TFs (POU2F2) positively regulated genes showed elevated activity in heart and muscle (**Fig. 7e**).

To further confirm the unbiased inference of organ-specific TF regulons based on scRNA-seq data (see **Fig. 7d**), we validated the organ-specific TF regulons using organ-matched

scATAC-seq data. We therefore measured chromatin accessibility at *cis*-elements containing a specific TF binding motif using chromVAR³⁶ and accessibility changes were analyzed in binding sites for the above identified organ-specific TFs (denoted as TF deviation scores). In general, TF deviation scores showed similar organ-specific patterns to regulon activity scores (Fig. 7f). For example, the HOXD8 regulon is kidney-specific and showed high TF deviation scores in kidney, while the regulon activity and TF deviation scores of ONECUT1 are both liver-specific (Fig. 7g). These analyses emphasize the unbiased prediction of organ-specific gene regulatory networks at the single-cell level.

Comparison of cell landscapes among human, mouse and cynomolgus monkey

The cynomolgus monkey cell landscape offers the opportunity to compare the cellular components and transcriptomic dynamics across species with similar organ compositions. Here we integrated scRNA-seq data from the non-human primate cynomolgus monkey (by this study), human¹⁶ and mouse⁴⁵ with matched organs/tissues using orthologous genes for cross-species analysis (Fig. 8a and Supplementary Fig. 16a,b). The integrated cell map consists of 338,932 cells (Fig. 8b), which were grouped into 15 major cell types (Fig. 8c and Supplementary Fig. 16c,d). Although the cell-type compositions largely varied in the three species (Fig. 8d,i), the expression patterns of representative marker genes and transcriptomic similarity of cell types were overall consistent across species (Fig. 8e,f and Supplementary Fig. 14e,f). Consistent with previous single-cell comparative analyses^{3,18-20}, the gene expression patterns of the major cell types are conserved in all three species, including immune, stromal and epithelial cells (Fig. 8g). As expected, cynomolgus monkey and human showed significantly higher cell-type similarity in orthologous gene expression than comparisons in other species (Fig. 8h).

We next investigated the transcriptomic dynamics of the same cell types among different species, with a specific focus on varying cell types across species such as hepatocytes and ciliated cells (Fig. 8i). To this end, we performed pairwise comparison of gene expression in liver for hepatocytes and in stomach for ciliated cells (Fig. 8j). In the differential analysis of gastric ciliary cells between human and cynomolgus monkey, we found that *LYZ* was highly

expressed in cynomolgus monkeys⁴⁶. *LYZ* has a dual role of immune defense and digestive function⁴⁷. As a gastric lipase, *LIPF* is expressed in human chief cells and promotes lipid metabolism⁴⁸. *ORM1*, as an acute-phase protein, was highly specifically expressed in human hepatocytes and had a certain promoting effect on liver regeneration.

Discussion

Non-human primates (NHP) are similar to humans in terms of anatomy, physiology and biochemical metabolism. Cynomolgus monkeys, a well-established laboratory animal model, have outstanding contributions to the scientific field⁴⁹. Although several single cell transcriptomic atlases have recently been established in cynomolgus monkeys based on a few organs (including ovary, lung, heart and artery)²³⁻²⁵, an organism-wide single-cell map is still lacking in this model species. Here, we chart a reference cell map of cynomolgus monkeys (named 'Monkey Atlas') using both scATAC-seq and scRNA-seq data across multiple organs, allowing us to gain deeper insights into the molecular dynamics and cellular heterogeneity of the cynomolgus monkeys organism.

As a proof of concept, we have performed various analyses based on the Monkey Atlas to show its wide uses, including the discovery of new putative cell types, the identification of key regulators in organ specification, and the ability to compare cell types across organs and species. For instance, our data shows that ciliated cells present in various organs of cynomolgus monkeys and the different ciliated cell subpopulations show various functions related to metabolic process, signal transduction, and cellular response to stimulus (Fig. 2). This observation somehow expands our notion that ciliated cells are generally found in respiratory system⁵⁰ with vital role in cleaning pathogenic microorganisms and signal transduction²⁷.

Recently, comprehensive reference cell maps across organs have been established in human^{3,51} and mouse^{45,52}. Although the Monkey Atlas does not provide exhaustive characterization of all organs in cynomolgus monkeys, it does offer a rich dataset of the most

populously studied organs in biology. In this regard, we performed cross-species integration analysis of cell maps to explore the molecular and cellular differences among the three species with comprehensive single cell data. We noticed that cynomolgus monkeys and human both have abundant immune cells and epithelial cells and a comparative composition of cell types in matched organs. This indicates that cynomolgus monkeys are ideal models to study complex diseases.

In conclusion, the Monkey Atlas provides valuable information about the most populous and important cell populations in NHP, and presents a foundation for preclinical studies.

Methods

Organ tissue collection

The cynomolgus monkey sample collection and research conducted in this study were approved by the Research Ethics Committee of the Changchun Biotechnology Development Co., Ltd. (Approval Number: 21001). Tissues were collected from 16 organs including trachea, spleen, stomach, kidney, uterus, tongue, testis, muscle, lung, liver, heart, colon, breast, bladder, adipose and aorta. To be more specific, tissues were cut into 1-2 mm³ pieces in RPMI-1640 medium (Gibico) with 12% fetal bovine serum (FBS, Gibico). Then the tissues were enzymatically digested with gentleMACS (Miltenyi) according to manufacturer's instruction. Cells were passed through a 70 µm cell trainer (Miltenyi) and centrifuged at 300 g for 5 min at 4 °C. The pelleted cells were re-suspended in red blood cell lysisbuffer (Beyotime) and incubated 1 min to lyse red blood cells. After wash twice with 1XPBS (Gibico), the cell pellets were re-suspended in sorting buffer (PBS with supplemented with 1% FBS). The single cells were captured in the 10× Genomics Chromium Single Cell 3' Solution, and RNA-seq libraries were prepared following the manufacturer's protocol (10× Genomics). The libraries were subjected to high-throughput sequencing on the Novaseq6000 platform, and 150-bp paired-end reads were generated.

scRNA-seq and data processing

Single-cell gene expression data were aligned to the *Macaca fascicularis* reference genome (macFas6) and processed for barcode assignment and unique molecular identifier (UMI) counting using the CellRanger v3.1.0 pipeline (10x Genomics). Filtered count matrices from the CellRanger pipeline were converted to sparse matrices using Seurat package (v4.0.0) in R⁵³, and cells expressing more than 4000 genes or less than 200 genes and more than 20% of mitochondrial genes expressing in UMI counts were filtered out before downstream analysis. Filtered data were then log normalized and scaled, with cell-cell variation due to UMI counts and percent mitochondrial reads regressed out. Then, we log normalised and scaled the filtered data to avoid cell-to-cell variation caused by UMI counts and percent mitochondrial reads removal.

As the samples involved the integration of large multi-organ samples such as trachea and spleen, Seurat's Robust Principal Component Analysis (RPCA) method was adopted for data integration. Cell clustering was performed at 0.8 resolution using the "FindClusters" function, and cell identity were defined using the top 20 principal components (PCs), and 17 clusters were identified. Dimensionality was reduced by the "RunUMAP" function and by visual Uniform Manifold Approximation and Projection (UMAP). Different types of cells were extracted for subgroup cell clustering, and their first 20 PCS were used for clustering. In the end, we identified 40 different subgroups. To ensure the accuracy of subsequent analysis, all 40 different subgroups were processed to remove double cells. Wilcoxon rank-sum test (FindAllMarkers function with default parameters) was used to identify markers for each cluster. Marker genes for each cluster are shown in **Supplementary Table S1**.

Creating a reference package for *Macaca fascicularis*

First, the FASTA and GTF files of the *Macaca fascicularis* reference genome were downloaded from the Ensembl database (version 6.0). Then, GTF files were filtered since it contains entries for non-polyA transcripts that overlap with protein-coding gene models. Because of the overlapping annotations, these entries can cause reads to be flagged as mapped to multiple genes (multi-mapped). Therefore, these entries were removed from the GTF file.

Finally, we create a Reference Package using genome FASTA and filtered GTF files.

Gene-set enrichment analysis

Gene-set enrichment analysis (GSEA) is a gene-based enrichment analysis method. The clusterProfiler package⁵⁴ in R was performed in all the gene-set enrichment analyses in this study.

Gene-set variation analysis

Gene-set variation analysis (GSVA)⁵⁵ starts from gene expression amount and multiple pathway information. Unsupervised samples were classified according to pathway activity changes. Three states defined by pseudo time analysis were choose as groups and genes corresponding to the three states were utilized to conduct Gene-set variation analysis in R.

Trajectory analysis using Monocle v.2

R package Monocle2 (version 2.99.3) was used to illustrate the epithelial cells state transition. In general, UMI count matrix of epithelial cells, cell phenotype information and gene annotation information, and the `negbinomial.size()` parameter were used to create a `CellDataSet` object. The variable genes obtained from epithelial cell types were detected by Seurat to sort cells in pseudotime. We used the `DDRTree` method and `orderCells` function for dimensional reduction and cell ordering. The `secretory_cell_CYB5Ahigh` cluster (E03) was defined as the root state argument and aligned via the “`orderCells`” function.

Trajectory analysis using RNA velocity

We sorted the `possorted_genome_bam.bam` file in the `outs` folder generated after each organ sample ran CellRanger, and then use Velocity's `run10x` function to generate loom files. To run `run10x`, we needed to prepare `rmsk.gtf` and the `genes.gtf` of the genome as input files. In this paper, RNA Velocity was used for trajectory analysis of epithelial cells and stromal cells.

Cell-cell interaction analysis

Cell-cell interactions among the cell types were estimated by CellPhoneDB (v2.1.1)⁵⁶ with default parameters (10% of cells expressing the ligand/receptor) and using version 2.0.0 of the database. We converted the normalized *Macaca fascicularis* genes into human homologous and used them as inputs. Interactions with p-value < 0.05 were considered significant. We considered only ligand-receptor interactions based on the annotation from the database, for which only and at least one partner of the interacting pair was a receptor, thus discarding receptor-receptor and other interactions without a clear receptor.

Create cisTarget databases for *Macaca fascicularis*

Firstly, chromosomes were cut at start and stop sites. EnsemblID, SymbolID, and annotation information from the GTF file were extracted to make the genes.bed file. The fasta file was generated from genes.bed file (10kb up- and 10kb downstream of the TSS) with 'bedtools getfasta'. Human motifs were downloaded from the CIS-BP database. Then, we extracted Motif_ID and TF_Name from the motifs file and outputted them as motifs_list. After, we converted all motif files in the pwms_all_motifs folder into motif.cb files. Finally, the fasta file, motifs_list file, and motif.cb files were used as input files to create cisTarget databases for *Macaca fascicularis*(macFas6).

Gene regulatory network

To identify cell type-specific gene regulatory networks, we performed Single-cell Regulatory Network Inference and Clustering (v0.10.0; a Python implementation of SCENIC)³ in our *Macaca fascicularis* dataset. First, the original expression data were normalized by dividing the gene count for each cell by the total number of cells in that cell and multiplying by 10,000, followed by a log_{1p} transformation. Next, we used the normalized counts to generate the co-expression module, using the GRNboost2 algorithm implemented in the arboreto package(v.0.1.3). Finally, we used pySCENIC with its default parameters to infer co-expression modules, cis-regulatory were used to filter motif analysis (RcisTarget, using macFas6 motif set), in order to only retain the corresponding transcriptional speculative direct binding target module enrichment factor (TF). The rest of the modules were trimmed

due to the lack of motif target support. An AUCell value matrix was generated to represent the regulators in each cell with different activity. The final gene regulatory networks consisted of 86 regulons for our *Macaca fascicularis* dataset as shown in figure 7a. GRNs can be visualized using igraph package in R.

scATAC-seq data pre-processing

The scATAC-seq sequencing data are pre-processed by cellranger-atac (v1.2.0) with the count command line. The running parameters are used by default except for “--force-cells=”. The “--force-cell” is 10000 for liver, lung and colon, 8000 for spleen, and the rest of the organs have no restriction on this parameter. For the subsequent scATAC-seq data processing and analysis, we used the ArchR (v1.0.1) package⁵⁷. *Macaca fascicularis* genome were constructed and annotated by createGenomeAnnotation and createGeneAnnotation function respectively. Then arrow file were created by createArrowFiles function with the default parameters. We used the addDoubletScores function to infer the doublet, filterDoublets function was used to remove the potential doublets with the “filterRatio = 1.0” parameter. ArchR project was created by ArchRProject function with the default parameters. For dimensionality reduction, we use the addIterativeLSI function in ArchR with the following parameters: “iterations = 4, clusterParams = list (resolution = c(0.2, 0.4, 0.6), sampleCells = 10000, n.start = 10, maxClusters = 6), varFeatures = 20000, dimsToUse = 1:50, scaleDims=FALSE”. Next, the Harmony package were utilized to remove the batch effect by addHarmony function⁵⁸. AddClusters function was used to cluster cells by its default parameters. For single cell embedding, we selected the reducedDims object with harmony and used addTSNE function with the parameter “perplexity = 30” for visualization.

Marker genes identification and cluster annotation

To identify the marker gene, gene scores were calculated when the ArchR project was created and stored in the arrow file. Then getMarkerFeatures function was used to identify the cluster-specific “expressed” genes with the default parameters. To visualize the marker genes in embedding, we used addImputeWeights function to run the MAGIC⁵⁹ to smooth gene scores across the nearby cells. For track plot, we used the plotBrowserTrack function

with the default parameters except for “tileSize = 100”.

Peak calling and TF binding motif analysis.

Before peak calling, we used `addGroupCoverages` function with default parameters to make pseudo-bulk replicates. Then `addReproduciblePeakSet` function was used with its default parameters except for “genomeSize = 2.7e09” to call accessible chromatin peaks using MACS2(v2.2.7.1)⁶⁰. For cell type specific peak analysis, firstly, `getMarkerFeatures` function was applied with peak matrix to identify marker peaks. Then `getMarkers` function with parameter “cutOff = FDR <= 0.01 & Log2FC >= 1” was conducted to get the differential peaks. For TF motif enrichment analysis, *Macaca fascicularis* motif was downloaded from the CIS-BP database (<http://cisbp.ccb.utoronto.ca/>). Then motif annotation were added to ArchR project by `addMotifAnnotations` function and TF motif enrichment in differential peaks were compute by the `peakAnnoEnrichment` function with “cutOff = FDR <= 0.1 & Log2FC >= 0.5” parameter. For motif footprint analysis, we first used `getPosition` function to locate relevant motifs, then `getFootprints` function was used to compute our interest motif footprints with its default parameters. At last footprint patterns were illustrated in plot by `plotFootprints` function with the following parameters: “normMethod = Subtract, smoothWindow = 10”.

Integrative analysis of scRNA-seq and scATAC-seq data

In order to better align and integrate the scATAC-seq data, we extracted and annotated the scRNA-seq data of the 7 organs corresponding to scATAC-seq. We first used the `FindTransferAnchors` function from the Seurat package and aligned the data with `addGeneIntegrationMatrix` function in ArchR with “unconstrained integration” mode. We found in our result that most of the predicted scores > 0.5. To improve the accuracy of the predictions and better integrate the two datasets, we applied the “constrained integration” mode again to integrate the scATAC-seq and scRNA-seq data. Briefly, we annotated the scATAC-seq data with cell types based on the gene scores of scATAC-seq. Then, a restricted list were created such that gene expression similarity was calculated only in the same cell type for both scATAC-seq and scRNA-seq. For peak to gene linkage analysis, we used the

addPeak2GeneLinks function to compute peak accessibility and gene expression with the parameters “corCutOff = 0.45, resolution = 1”.

Data availability

The raw files are available from China National GeneBank (CNGB) (<https://db.cngb.org/>): RNA: CNP0002427; ATAC: CNP0002441. Gene counts and metadata are available at Zenodo (<https://zenodo.org/>): DOI: 10.5281/zenodo.5881495.

Acknowledgements

The authors acknowledge the High Performance Computing Center of Nanjing University for providing high performance computing (HPC) resources. This work is supported by the National Natural Science Foundation of China (No. 32070656, 81872877, 82173819 and 81872876), the Nanjing University Deng Feng Scholars Program and the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions.

Author contributions

D.C., Y.S. and Y.X. conceived and designed the study. Q.J., Y.W., Y.D. and X.Q. performed animal experiments. F.Y., T.Z. and W.F. performed data analyses. D.C. and J.Q. wrote the manuscript with input from X.Z. and Q.X.. All authors reviewed and approved the submitted version.

Additional information

Competing interests: The authors declare no competing interests.

References

1. Bhatt, D.L. & Mehta, C. Adaptive Designs for Clinical Trials. *N Engl J Med* **375**, 65-74 (2016).
2. Chappell, L., Russell, A.J.C. & Voet, T. Single-Cell (Multi)omics Technologies. *Annu Rev Genomics Hum Genet* **19**, 15-41 (2018).
3. Han, X. *et al.* Construction of a human cell landscape at single-cell level. *Nature* **581**, 303-309 (2020).
4. He, S. *et al.* Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biol* **21**, 294 (2020).
5. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091-1107 e17 (2018).
6. Schaum, N. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-+ (2018).
7. Tabula Muris, C. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590-595 (2020).
8. Yu, Q. *et al.* Charting human development using a multi-endodermal organ atlas and organoid models. *Cell* **184**, 3281-3298 e22 (2021).
9. Saviano, A., Henderson, N.C. & Baumert, T.F. Single-cell genomics and spatial transcriptomics: Discovery of novel cell states and cellular interactions in liver physiology and disease biology. *J Hepatol* **73**, 1219-1230 (2020).
10. Williams, J.W. *et al.* Single Cell RNA Sequencing in Atherosclerosis Research. *Circ Res* **126**, 1112-1126 (2020).

11. Chen, H., Ye, F. & Guo, G. Revolutionizing immunology with single-cell RNA sequencing. *Cell Mol Immunol* **16**, 242-249 (2019).
12. Colonna, M. & Brioschi, S. Neuroinflammation and neurodegeneration in human brain at single-cell resolution. *Nat Rev Immunol* **20**, 81-82 (2020).
13. Ranzoni, A.M. *et al.* Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis. *Cell Stem Cell* **28**, 472-487 e7 (2021).
14. Andueza, A. *et al.* Endothelial Reprogramming by Disturbed Flow Revealed by Single-Cell RNA and Chromatin Accessibility Study. *Cell Rep* **33**, 108491 (2020).
15. Xing, Q.R. *et al.* Diversification of reprogramming trajectories revealed by parallel single-cell transcriptome and chromatin accessibility sequencing. *Sci Adv* **6**(2020).
16. Han, X.P. *et al.* Construction of a human cell landscape at single-cell level. *Nature* **581**, 303-+ (2020).
17. Wang, J. *et al.* Tracing cell-type evolution by cross-species comparison of cell atlases. *Cell Rep* **34**, 108803 (2021).
18. La Manno, G. *et al.* Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell* **167**, 566-+ (2016).
19. Hodge, R.D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61-+ (2019).
20. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Systems* **3**, 346-+ (2016).
21. Nakamura, T. *et al.* Data Descriptor: Single-cell transcriptome of early embryos and cultured embryonic stem cells of cynomolgus monkeys. *Scientific Data* **4**(2017).

22. VandeBerg, J.L. & Williams-Blangero, S. Advantages and limitations of nonhuman primates as animal models in genetic research on complex diseases. *J Med Primatol* **26**, 113-9 (1997).
23. Ma, S. *et al.* Single-cell transcriptomic atlas of primate cardiopulmonary aging. *Cell Research* **31**, 415-432 (2021).
24. Wang, S. *et al.* Single-Cell Transcriptomic Atlas of Primate Ovarian Aging. *Cell* **180**, 585-+ (2020).
25. Zhang, W.Q. *et al.* A single-cell transcriptomic landscape of primate arterial aging. *Nature Communications* **11**(2020).
26. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411-+ (2018).
27. Spassky, N. & Meunier, A. The development and functions of multiciliated epithelia. *Nat Rev Mol Cell Biol* **18**, 423-436 (2017).
28. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982 (2017).
29. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
30. Buckley, C.D., Barone, F., Nayar, S., Benezech, C. & Caamano, J. Stromal cells in chronic inflammation and tertiary lymphoid organ formation. *Annu Rev Immunol* **33**, 715-45 (2015).
31. Soliman, H. *et al.* Multipotent stromal cells: One name, multiple identities. *Cell Stem Cell* **28**, 1690-1707 (2021).

32. Natoli, G. & Ostuni, R. Adaptation and memory in immune responses. *Nat Immunol* **20**, 783-792 (2019).
33. Ma, P. *et al.* Immune Cell Landscape of Patients With Diabetic Macular Edema by Single-Cell RNA Analysis. *Front Pharmacol* **12**, 754933 (2021).
34. Orian-Rousseau, V. & Sleeman, J. CD44 is a multidomain signaling platform that integrates extracellular matrix cues with growth factor and cytokine signals. *Adv Cancer Res* **123**, 231-54 (2014).
35. Wu, C. *et al.* Galectin-9-CD44 Interaction Enhances Stability and Function of Adaptive Regulatory T Cells. *Immunity* **41**, 270-282 (2014).
36. Schep, A.N., Wu, B.J., Buenrostro, J.D. & Greenleaf, W.J. chromVAR : inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nature Methods* **14**, 975-+ (2017).
37. Clerc, R.G., Corcoran, L.M., Lebowitz, J.H., Baltimore, D. & Sharp, P.A. The B-Cell-Specific Oct-2 Protein Contains Pou Box-Type and Homeo Box-Type Domains. *Genes & Development* **2**, 1570-1581 (1988).
38. Hodson, D.J. *et al.* Regulation of normal B-cell differentiation and malignant B-cell survival by OCT2. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E2039-E2046 (2016).
39. Acharya, A. *et al.* The bHLH transcription factor Tcf21 is required for lineage-specific EMT of cardiac fibroblast progenitors. *Development* **139**, 2139-2149 (2012).
40. Pliner, H.A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell* **71**, 858-871 e8 (2018).

41. Parker, L.H. *et al.* The endothelial-cell-derived secreted factor Eglf7 regulates vascular tube formation. *Nature* **428**, 754-758 (2004).
42. Soncin, F. *et al.* VE-statin, an endothelial repressor of smooth muscle cell migration. *EMBO J* **22**, 5700-11 (2003).
43. Fitch, M.J., Campagnolo, L., Kuhnert, F. & Stuhmann, H. Eglf7, a novel epidermal growth factor-domain gene expressed in endothelial cells. *Developmental Dynamics* **230**, 316-324 (2004).
44. Wang, L. *et al.* Fev regulates hematopoietic stem cell development via ERK signaling. *Blood* **122**, 367-75 (2013).
45. Han, X.P. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091-+ (2018).
46. Janiak, M.C., Burrell, A.S., Orkin, J.D. & Disotell, T.R. Duplication and parallel evolution of the pancreatic ribonuclease gene (RNASE1) in folivorous non-colobine primates, the howler monkeys (*Alouatta* spp.). *Sci Rep* **9**, 20366 (2019).
47. Badiyan, S.N. *et al.* Clinical Outcomes of Patients With Recurrent Lung Cancer Reirradiated With Proton Therapy on the Proton Collaborative Group and University of Florida Proton Therapy Institute Prospective Registry Studies. *Pract Radiat Oncol* **9**, 280-288 (2019).
48. Busslinger, G.A. *et al.* Human gastrointestinal epithelia of the esophagus, stomach, and duodenum resolved at single-cell resolution. *Cell Rep* **34**, 108819 (2021).
49. Capitanio, J.P. & Emborg, M.E. Contributions of non-human primates to neuroscience research. *Lancet* **371**, 1126-35 (2008).

50. Sachs, N. *et al.* Long-term expanding human airway organoids for disease modeling. *EMBO J* **38**(2019).
51. He, S. *et al.* Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biology* **21**(2020).
52. Kalucka, J. *et al.* Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**, 764-+ (2020).
53. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502 (2015).
54. Yu, G.C., Wang, L.G., Han, Y.Y. & He, Q.Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics-a Journal of Integrative Biology* **16**, 284-287 (2012).
55. Hanzelmann, S., Castelo, R. & Guinney, J. GSEA: gene set variation analysis for microarray and RNA-Seq data. *Bmc Bioinformatics* **14**(2013).
56. Efremova, M., Vento-Tormo, M., Teichmann, S.A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* **15**, 1484-1506 (2020).
57. Granja, J.M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403-411 (2021).
58. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289-1296 (2019).
59. van Dijk, D. *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-729.e27 (2018).

60. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).

Figure Legends

Figure 1. Single-cell landscapes of 16 organs from cynomolgus monkeys

(a) Workflow of sample collection and single-cell transcriptome analysis for 16 organs and chromatin accessibility analysis for 7 organs from normal cynomolgus monkey. **(b)** Cell type identification (scRNA-seq), including 174,233 cells, 17 major cell types, and 40 cell subtypes. TSNE of cells were either colored by major cell types (left) or colored by cell subtypes (right). **(c)** Heatmap with the scaled expression levels of cell type-specific marker genes (left). 18 marker genes expression were randomly selected to exemplify the specificity of marker gene in the right. **(d)** Overview of cell type identification in scATAC-seq analysis. 66,566 cells and 10 major cell types were identified and cells were colored by major cell types. **(e)** Sankey diagram of scRNA-seq and scATAC-seq data cell-type mapping.

Figure 2. The heterogeneity and developmental state of epithelial cells

(a) Distribution of 14 epithelial cell subtypes on the UMAP. **(b)** Dot plot shows representative differentially expressed genes (DEGs) across epithelial subtypes. The size of dot is proportional to the fraction of cells which express specific genes. Color intensity corresponds to the relative expression of specific genes. **(c)** Bar plot shows the percentage of cell subtypes in each organ. **(d)** Heatmap shows functional pathways of ciliated epithelial cell enrichment, darker color represents a higher enrichment score. **(e)** Semi-supervised pseudotime trajectory of subtypes (E01-E14) of epithelial cells by monocle2. Trajectory is colored by cell states (left) and pseudotime (right). **(f)** Unsupervised pseudotime trajectory of subtypes (E01-E14) of epithelial cells by RNA velocity. Trajectory is colored by cell subtypes. Arrowhead direction represents the trend of cell pseudo-temporal differentiation. **(g)** Bar plot shows functional pathways of E03 subtype. **(h)** Heatmap showing the scaled expression of differentially expressed genes across pseudotime from **(e)**. Genes (on the right of the heatmap) are assigned to specific cell states based on their expression levels. Bar plots at the top of the heatmap are scale diagrams of different cell types, samples and cell states during pseudotime differentiation trajectory. **(i)** Heatmap showing functional pathways enriched in three cell states (S1, S2, S3) by GSVA analysis.

Figure 3. The heterogeneity of stromal cells

(a) Distribution of 11 stromal cell subtypes on the UMAP. (b) Dot plots showing representative top differentially expressed genes (DEGs) across stromal subtypes. Dot size is proportional to the fraction of cells expressing specific genes. Color intensity corresponds to the relative expression of specific genes. (c) Feature plot shows the expression of marker genes. (d) Bar plot showing the percentage of cell subtypes in each organ. (e) Distribution of stromal cells in different organs on the UMAP. (f) Unsupervised pseudotime trajectory of subtypes (S01-S11) of stromal cells by RNA velocity. Trajectory is colored by cell subtypes. The arrow direction is the trend of cell pseudo-temporal differentiation. (g) The UMAPs showing the pseudotime differentiation trajectory of fibroblasts, smooth muscle cells and endothelial cells. (h) Bar plot shows functional pathways of fibroblasts. (i) Heatmap showing the scaled expression levels of cell type-specific marker genes along pseudotime differentiation trajectory. Examples of marker expression are shown in the right UMAPs.

Figure 4. Immune cell heterogeneity

(a) Distribution of 13 major immune cell types on the UMAP. (b) Dot plots shows representative top differentially expressed genes (DEGs) across stromal subtypes. Dot size varies synchronously with the fraction of cells expressing specific genes. Color intensity corresponds to the relative expression of specific genes. (c) Distribution of immune cells in different organs on the UMAP. (d) Chord diagram maps the different cell types in all organs as a whole. The width of the arrow represents the proportion of cell types. (e) The bar chart shows the proportion of different cell types in the same organ. (f) Scatter plots shows pairings of gene expression between muscle and other organs. Each point represents a DEG, and its size is proportionate to the fold change.

Figure 5. Dynamics of cell-cell communication networks

(a) Chordal diagram of the integrated cell-cell interaction network among the major cell types of 16 organs. (b) Heatmap shows interaction intensity of cellular interactome from (a). Block sizes and colors are proportional to the interaction frequency. (c) Sankey diagram

shows the cell-cell interactions of different cell types in 16 organs. The thickness of lines represents the strength of cell-cell interactions. **(d)** Dot plot of interactions between selected cell subtypes in different organs. Each row represents a ligand-receptor pair, and each column defines a cell-cell interaction pair.

Figure 6. Single-cell chromatin landscape of major organs of cynomolgus monkeys

(a) Cell type identification (scATAC-seq), including 66,566 cells, and 22 cell subtypes. Shown is the tSNE of cells colored by cell subtypes. **(b)** Bar plot shows the number of reproducible peaks identified from each cluster. The peaks are classified to four category: distal, exonic, intronic and promoter. **(c)** Percentage of cell subtypes in each organ shown by bar plot. **(d)** Heatmap of 80,270 marker peaks across 22 subclusters identified by bias-matched differential testing ($FDR \leq 0.01$ and $\text{Log}_2FC \geq 3$). **(e)** Chromatin accessibility and gene expression of 52,229 significantly ($R > 0.45$ and $FDR < 0.1$) linked peak-gene pairs illustrated by heatmap. **(f)** Profile of POU2F2 and TCF21 gene accessibility, gene expression (inferred from scRNA-seq) and TF motif activity. **(g)** Visualization of the EGFL7 locus with the maximum number of peak-gene pairs shown by genome browser track (chr15: 1,772,849–1,832,850).

Figure 7. Organ-specific transcriptional regulatory networks

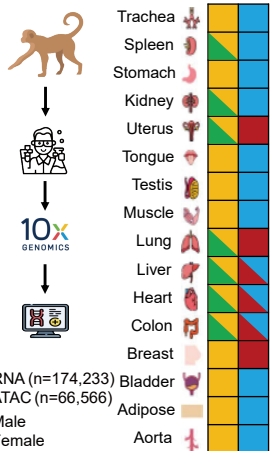
(a) Identification of regulon modules using SCENIC. Heatmap (left) shows the similarity of different regulons ($n=87$) based on the AUCell score. Eight regulon modules were identified based on regulon similarity. UMAPs (right) illustrate the average AUCell score distribution for different regulon modules (in different colors). **(b)** Wordcloud plots shows enrichment of organs in different regulon modules. **(c)** Representative transcription factors and corresponding binding motifs in different regulon modules. **(d)** Heatmap showing transcription factors enriched in different organs. Color depth represents the level of regulon-specific score. **(e)** Integrated gene-regulatory networks of the regulons. Regulon-associated TFs are highlighted in blue rectangles and target genes in circles. Target genes (in circles) are colored according to their highly expressed organs. **(f)** Heatmap showing gene-activity scores of marker genes in the indicated organs. **(g)** Violin plot showing motif activity (measured by TF chromVAR deviations) of TF regulons highlighted in (f).

Figure 8. Comparison of cell landscapes of human, mouse and cynomolgus monkey

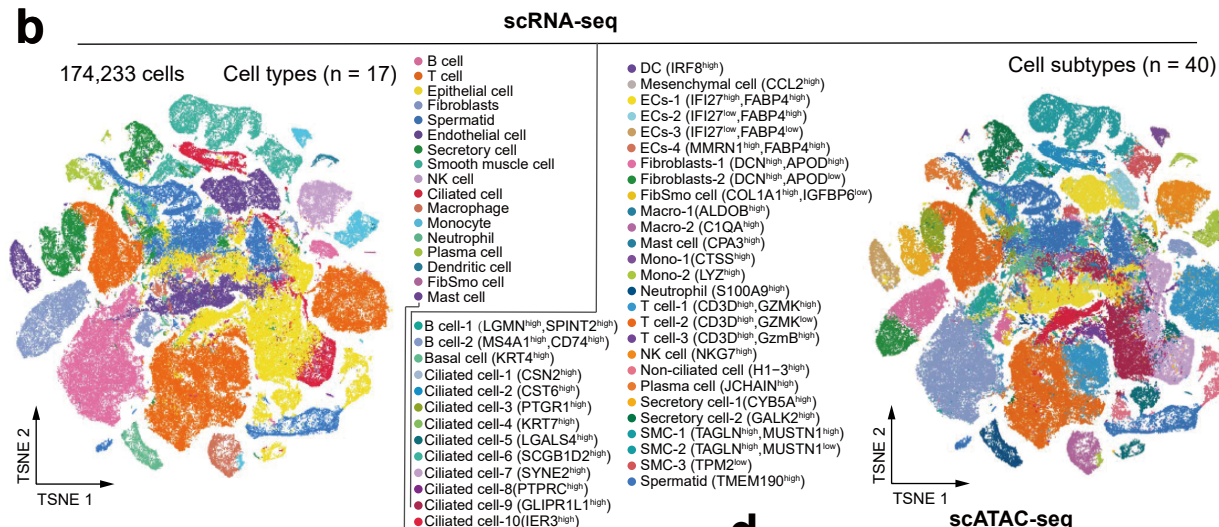
(a) Integration of data from 16 organs from cynomolgus monkey, 11 organs from human, and 9 organs from mouse. **(b)** Distribution of cells from cynomolgus monkey, human and mouse on the UMAP. **(c)** Distribution of 15 major cell types on the UMAP. **(d)** Bar plot showing the percentage of cell types in cynomolgus monkey, human and mouse. **(e)** Dot plots showing representative top differentially expressed genes (DEGs) across cell types. Dot size is proportional to the fraction of cells expressing specific genes. Color intensity corresponds to the relative expression of specific genes. **(f)** Feature Plot showing the expression of marker genes. **(g)** Correlation of orthologous gene expression between human, mouse and monkey pseudo-cell types (n=212) based on AUROC scores. The AUROC scores are calculated by MetaNeighbor to measure the similarity of cell type. The hierarchical clustering is calculated with pseudo-cell types. **(h)** Box plot showing correlation of top genes expression in cynomolgus monkey, human and mouse integrated data (Spearman method). Dots represent different cell types. **(i)** The proportion of hepatocytes in liver and ciliated cells in stomach of cynomolgus monkey, human and mouse samples. **(j)** Scatter plots showing a pairwise comparison of gene expression across species organs. Each point represents a DEG, and its size is proportionate to the fold change. Two scatter dots on the left show DEGs of Hepatocyte in the cynomolgus monkey, human and mouse samples, and Ciliated cells on the right.

Macaca fascicularis

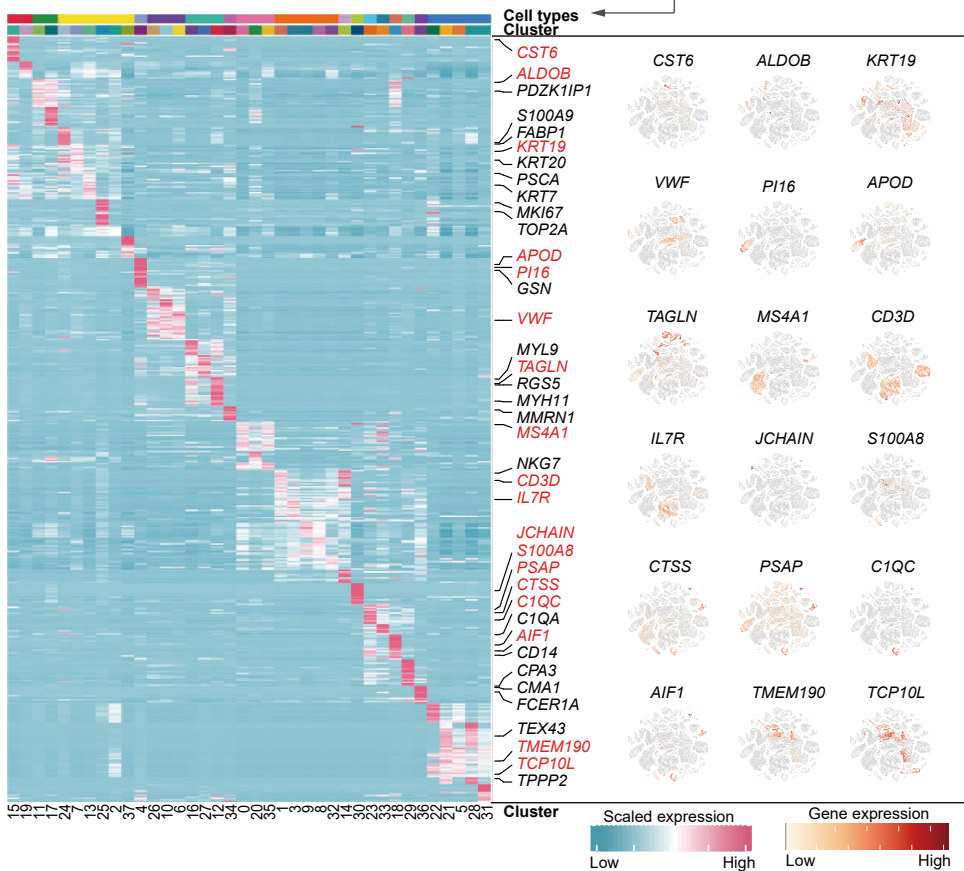
a



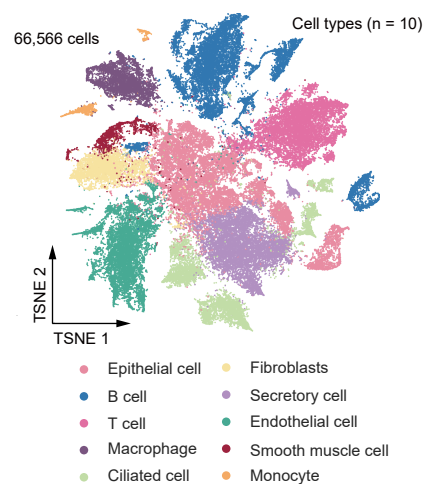
b



c



d



e

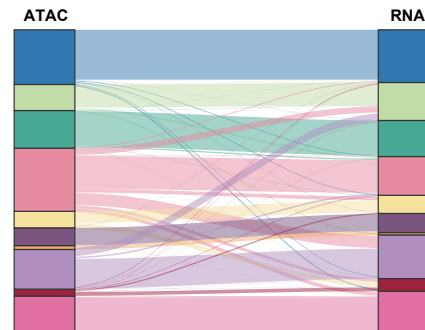
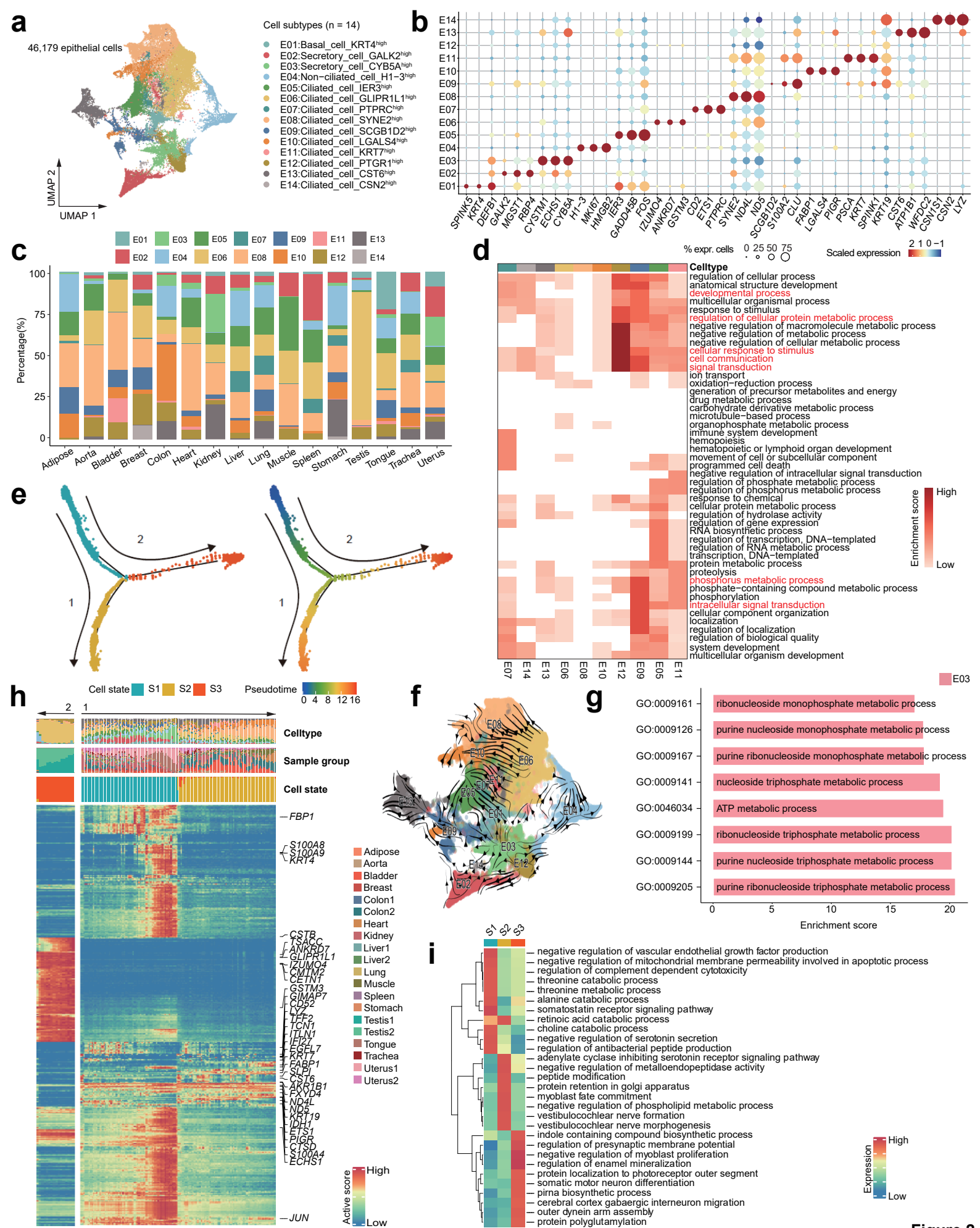


Figure 1



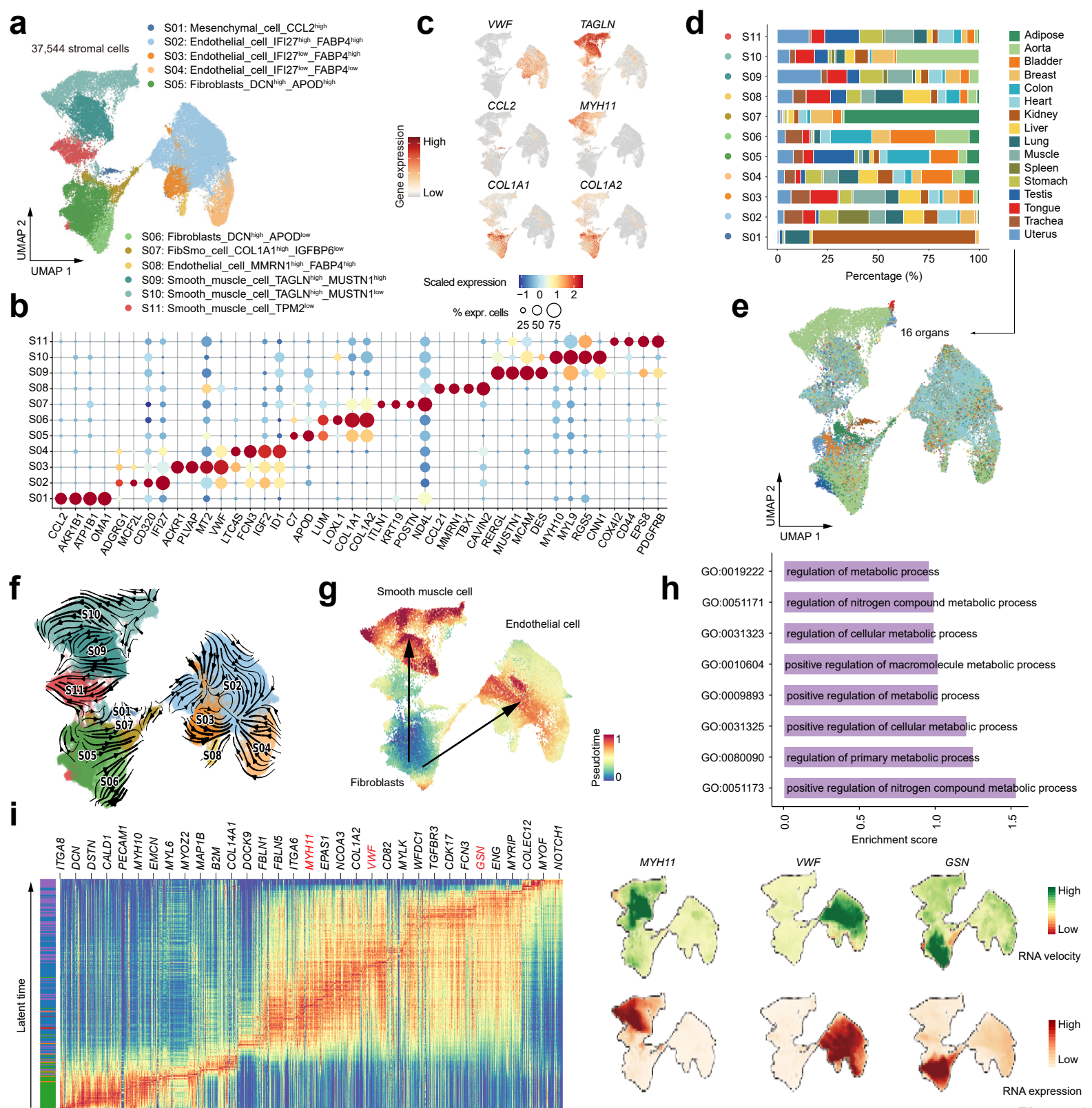


Figure 3

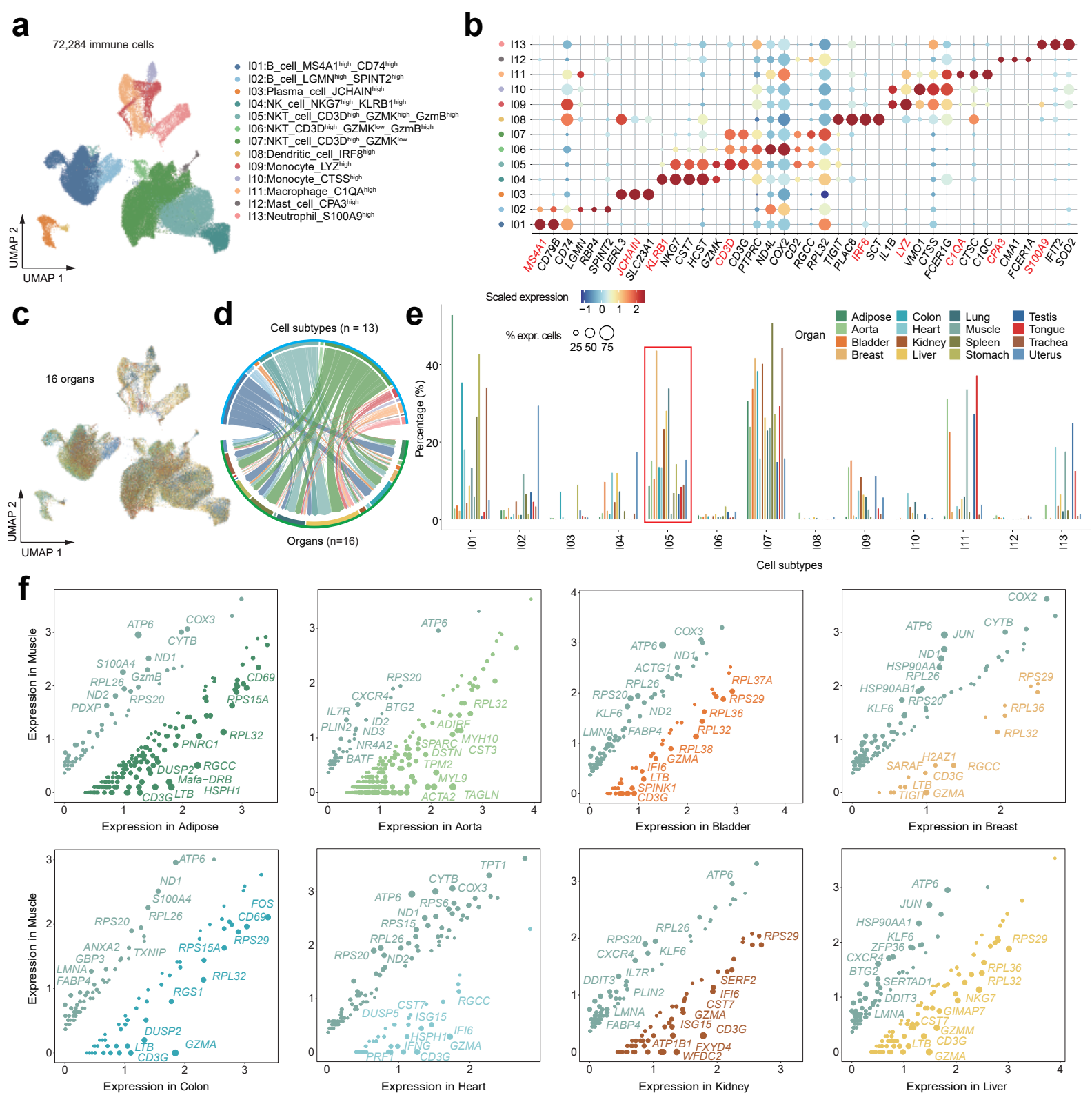
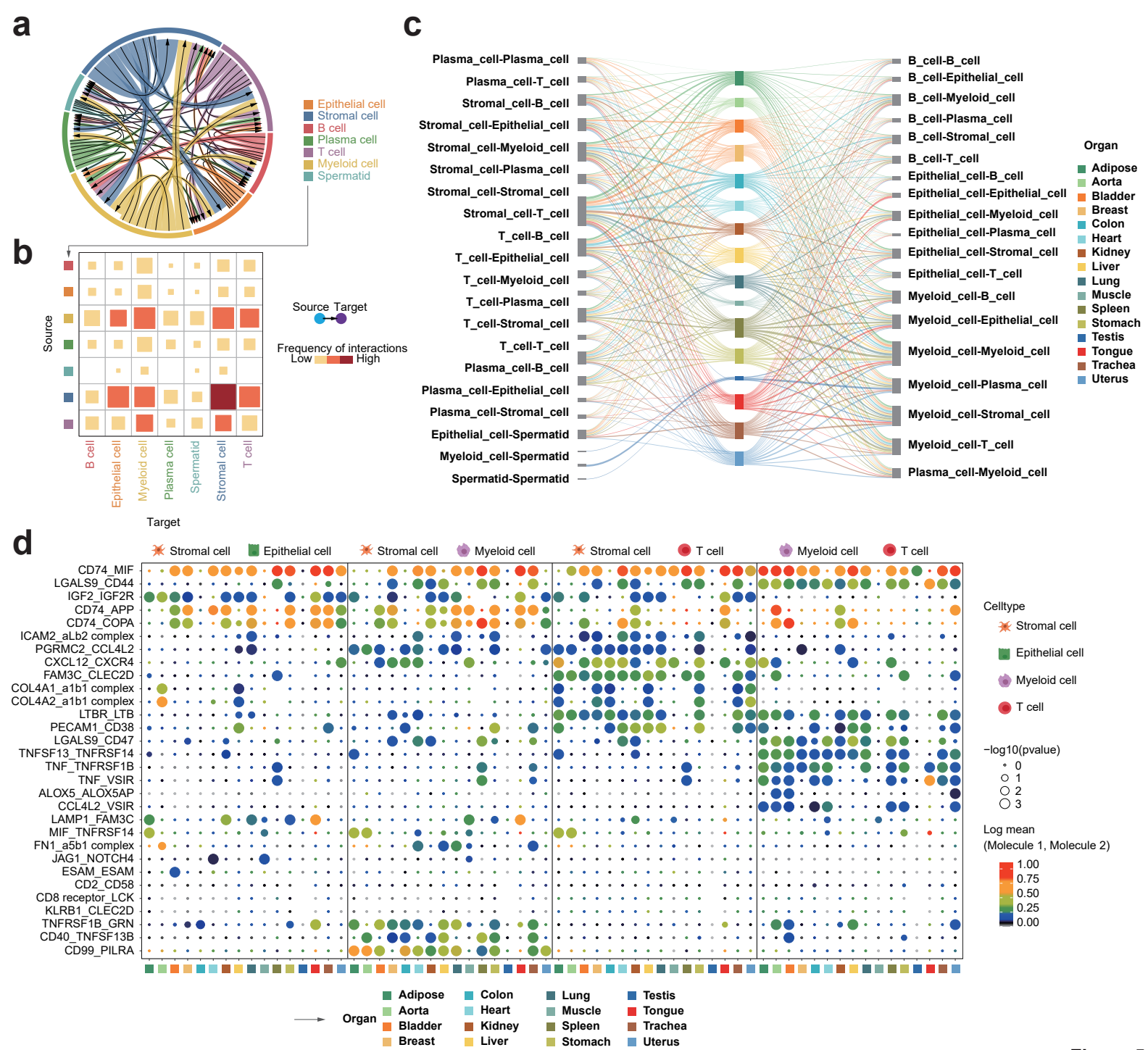


Figure 4



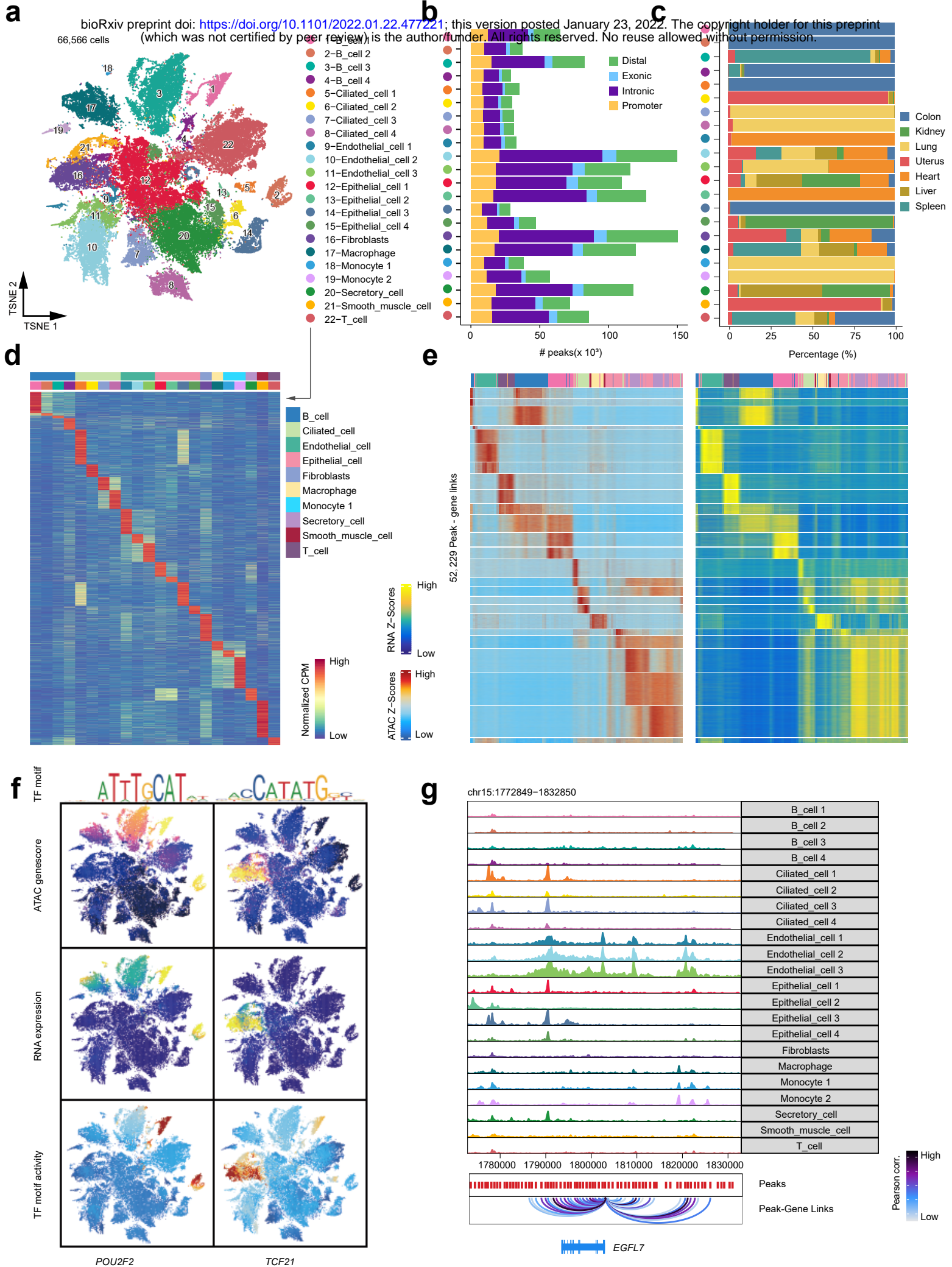


Figure 6

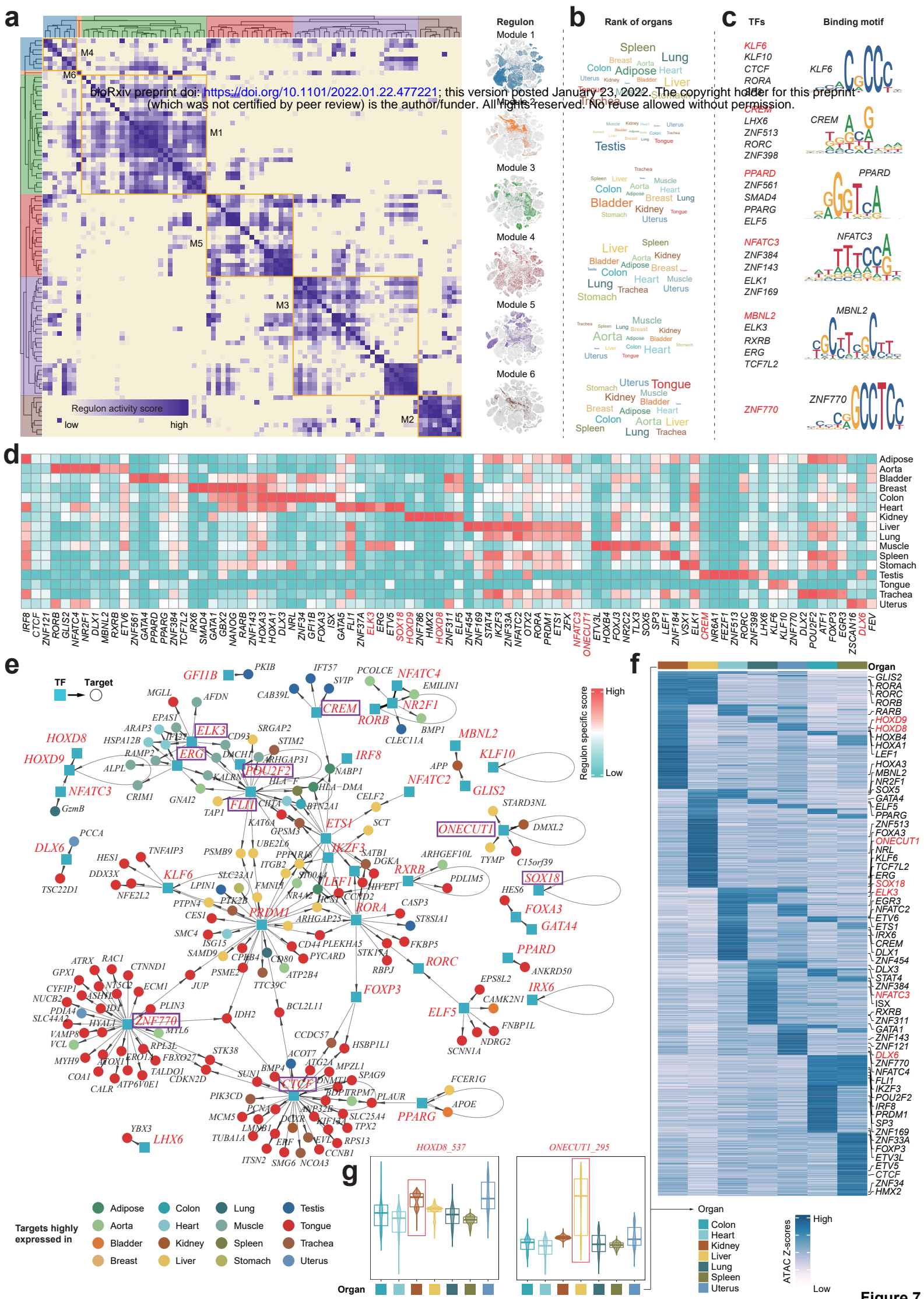


Figure 7

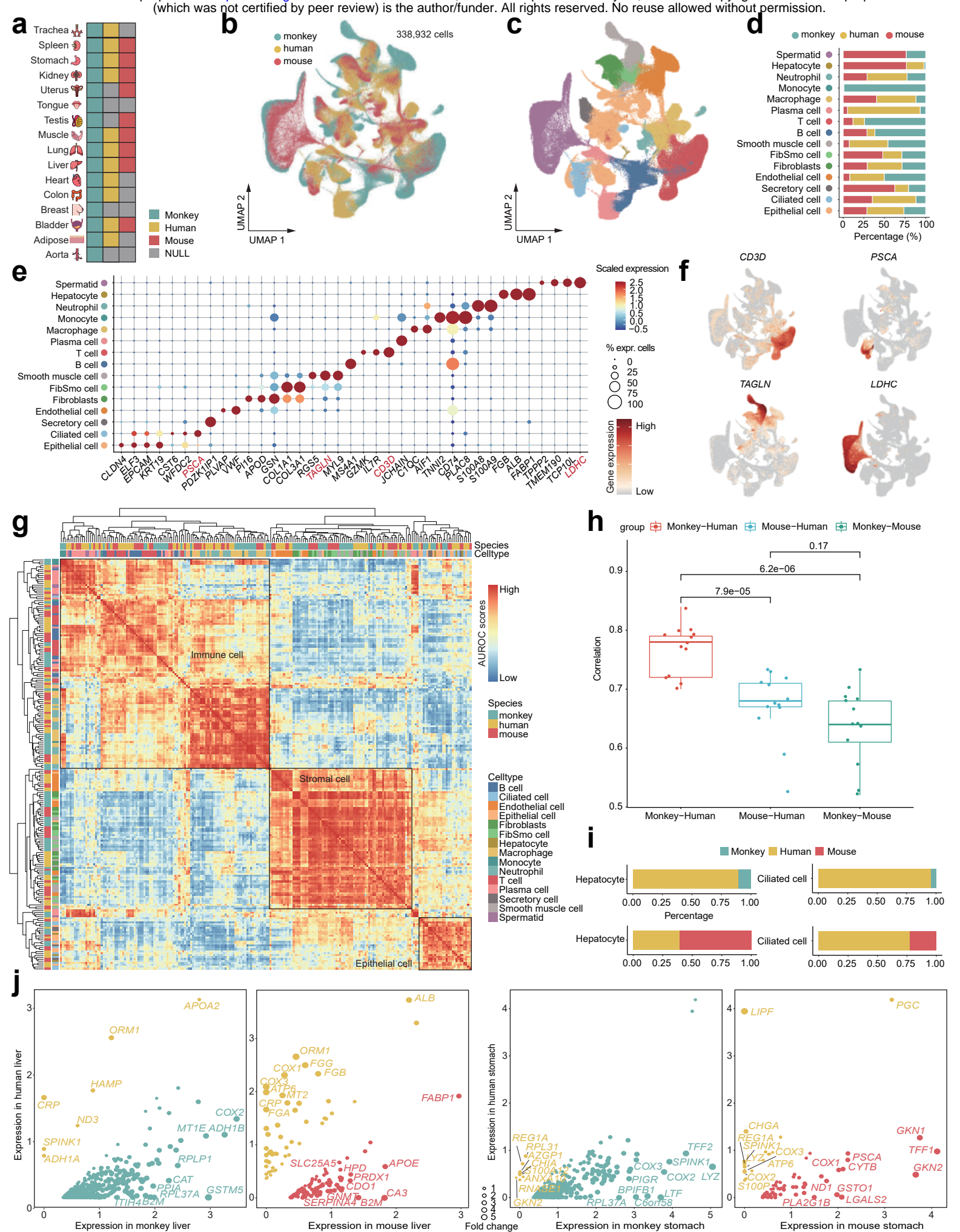
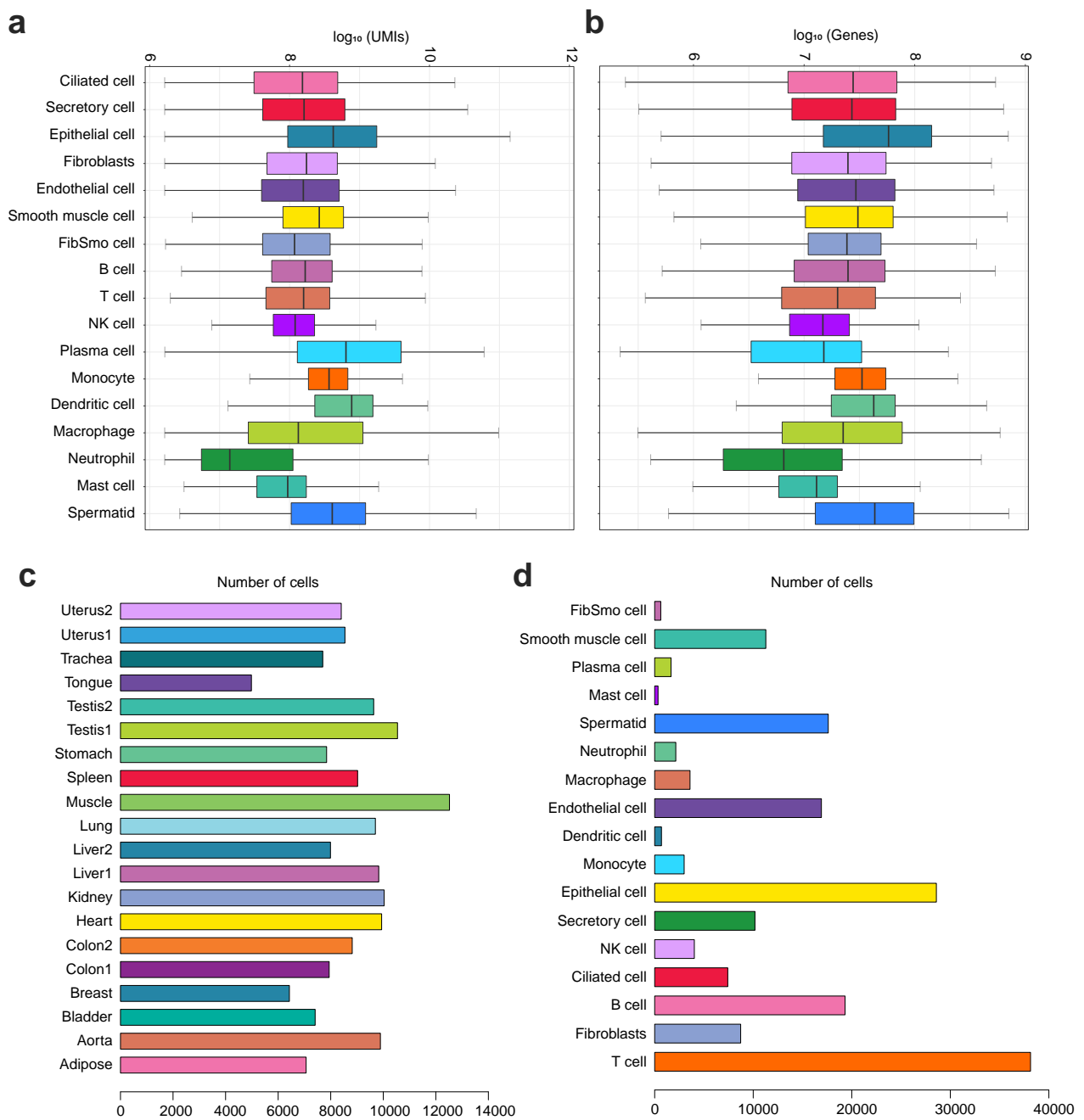
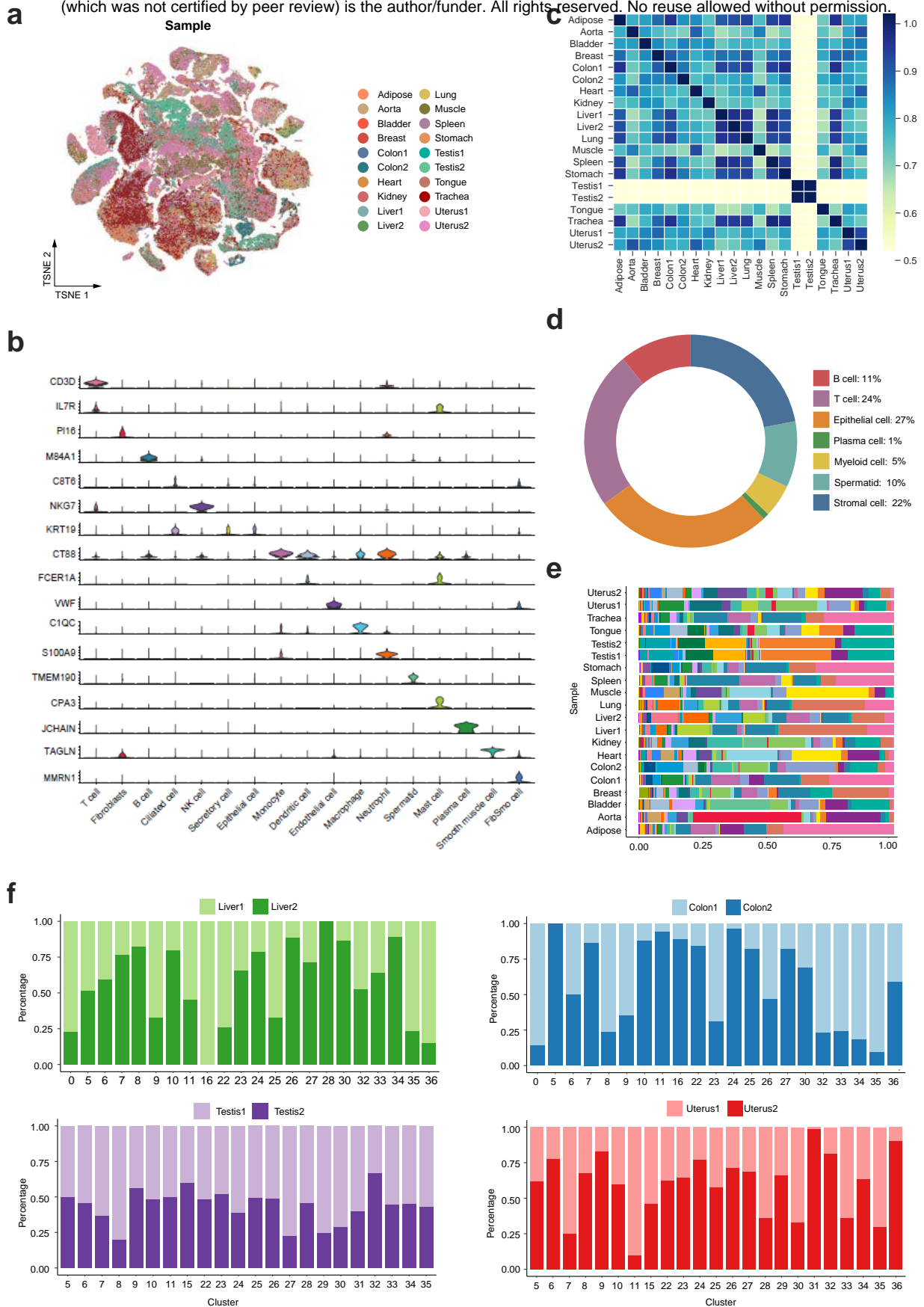


Figure 8

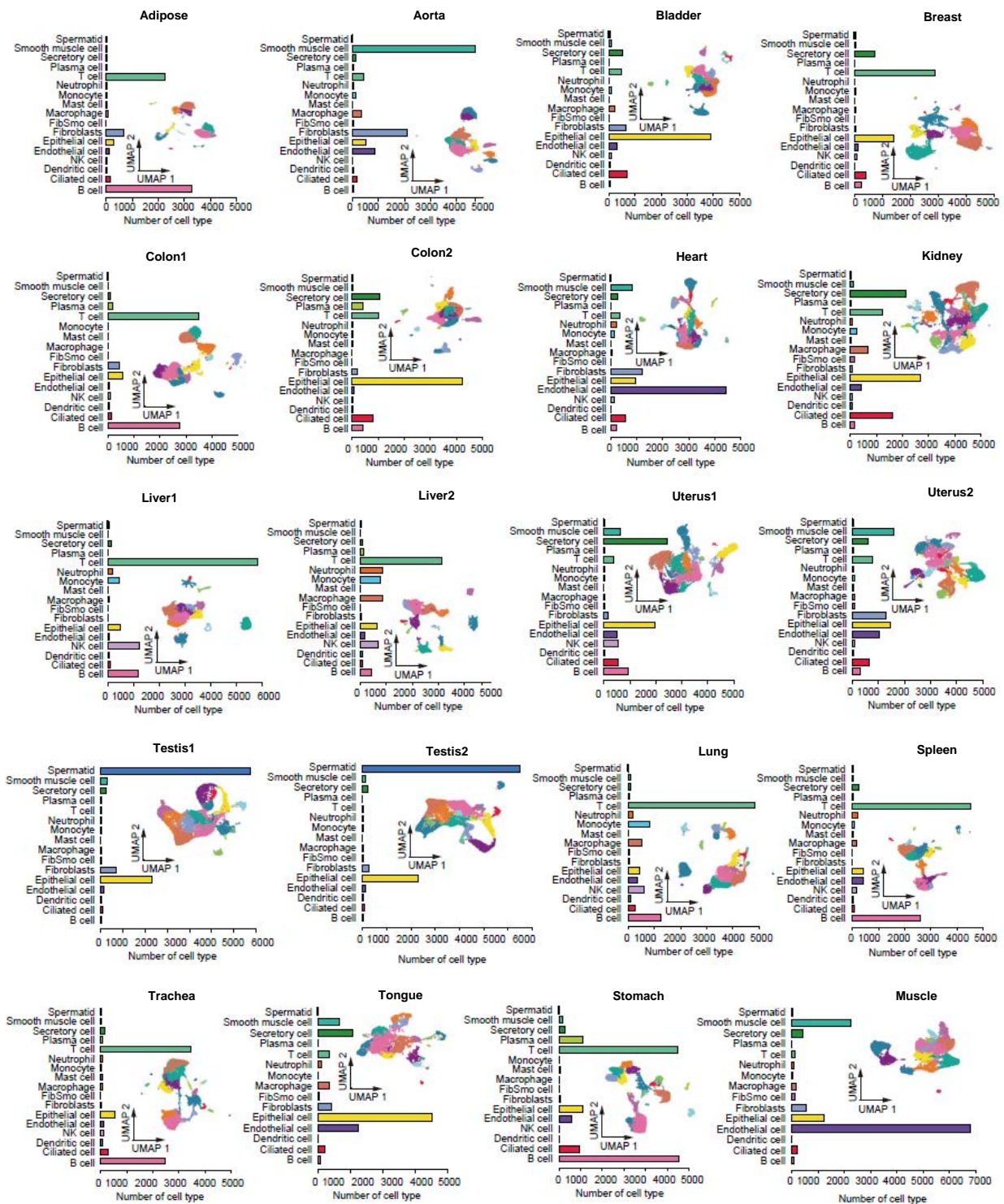
Supplementary Figures



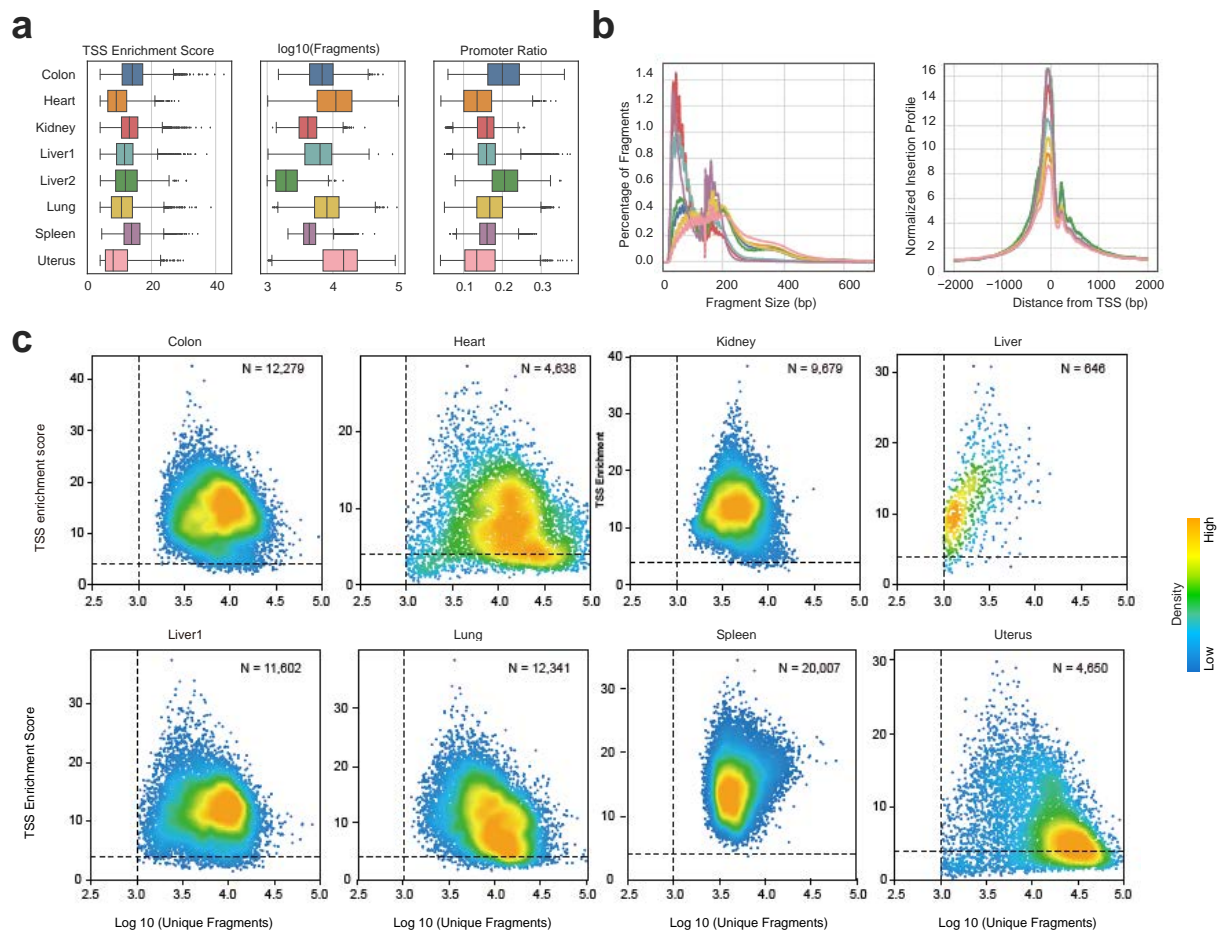
Supplementary Fig. 1: Data quality of scRNA. Related to Figure 1. (a-b) Box plot showing the number of UMIs and genes in major cell types respectively. **(c-d)** Bar plot showing the number of cells in major cell types and in samples.



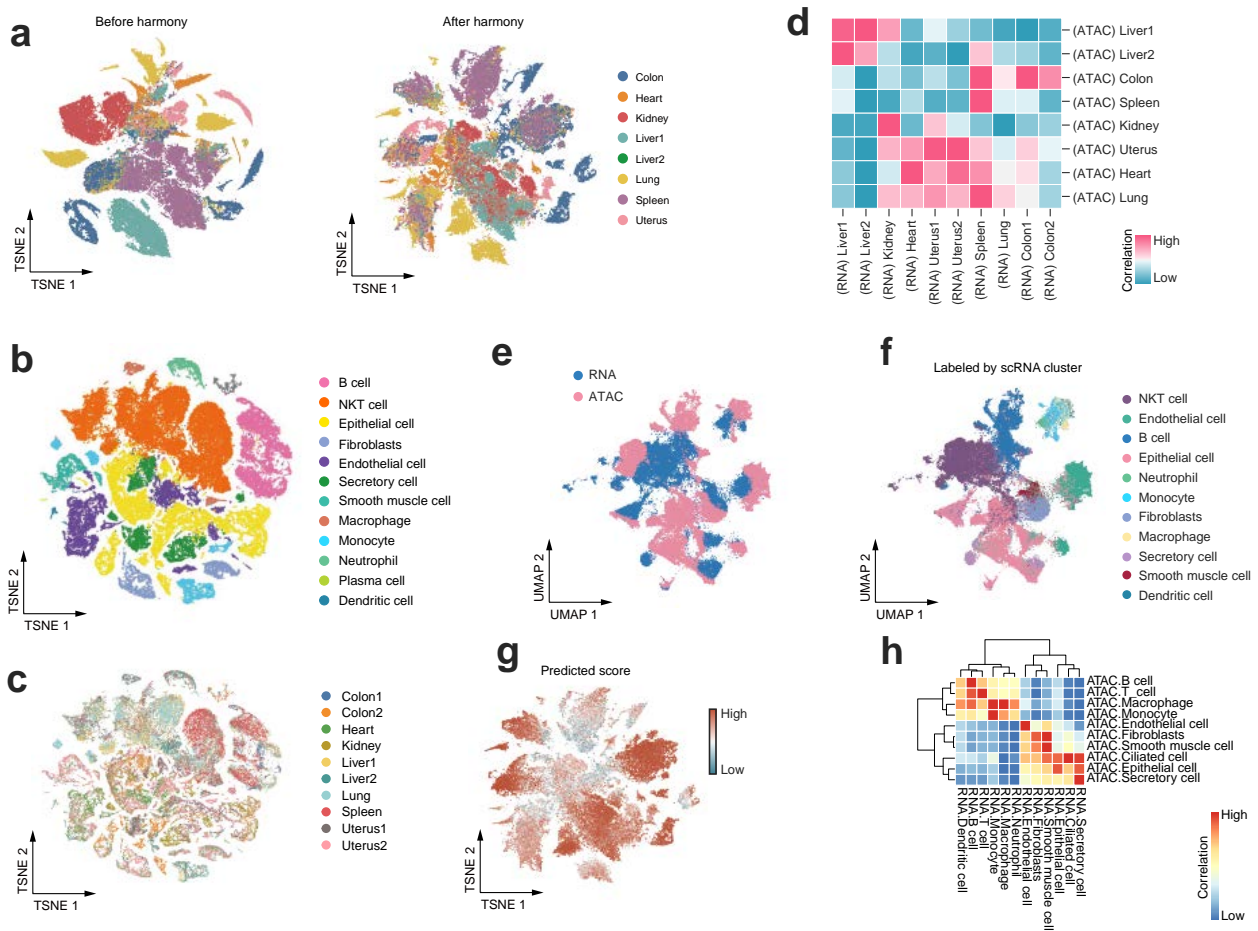
Supplementary Fig. 2: Data quality of scRNA. Related to Figure 1. (a) Distribution of 20 samples of cells on the tSNE. (b) Violin diagram showing the expression of marker genes in major cell types. (c) Heatmap showing correlation of highly expressed genes in 20 samples. (d) Circle diagram showing the cell proportions of seven cell types. (e) Bar plot showing the percentage of cells in each sample. (f) Bar Plot showing the proportion of cells in different clusters from repeated samples of liver, colon, testis and uterus.



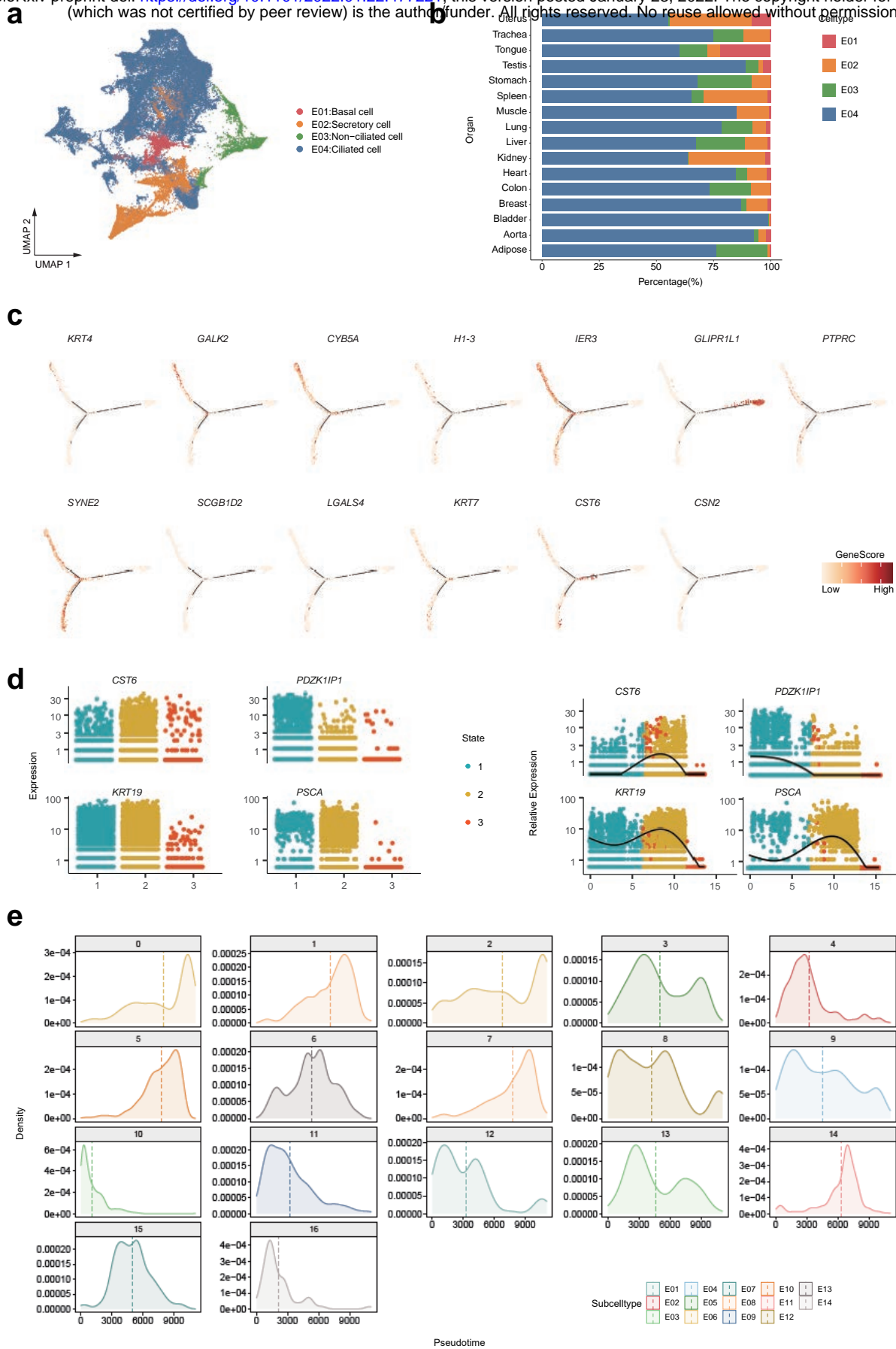
Supplementary Fig. 3: The UMAPs of all organs and bar plots of the proportion of different cell types (scRNA). Related to Figure 1. Distribution of different clusters on the UMAP in 20 samples, and bar plot shows the cell proportion of 20 samples of major cell types.



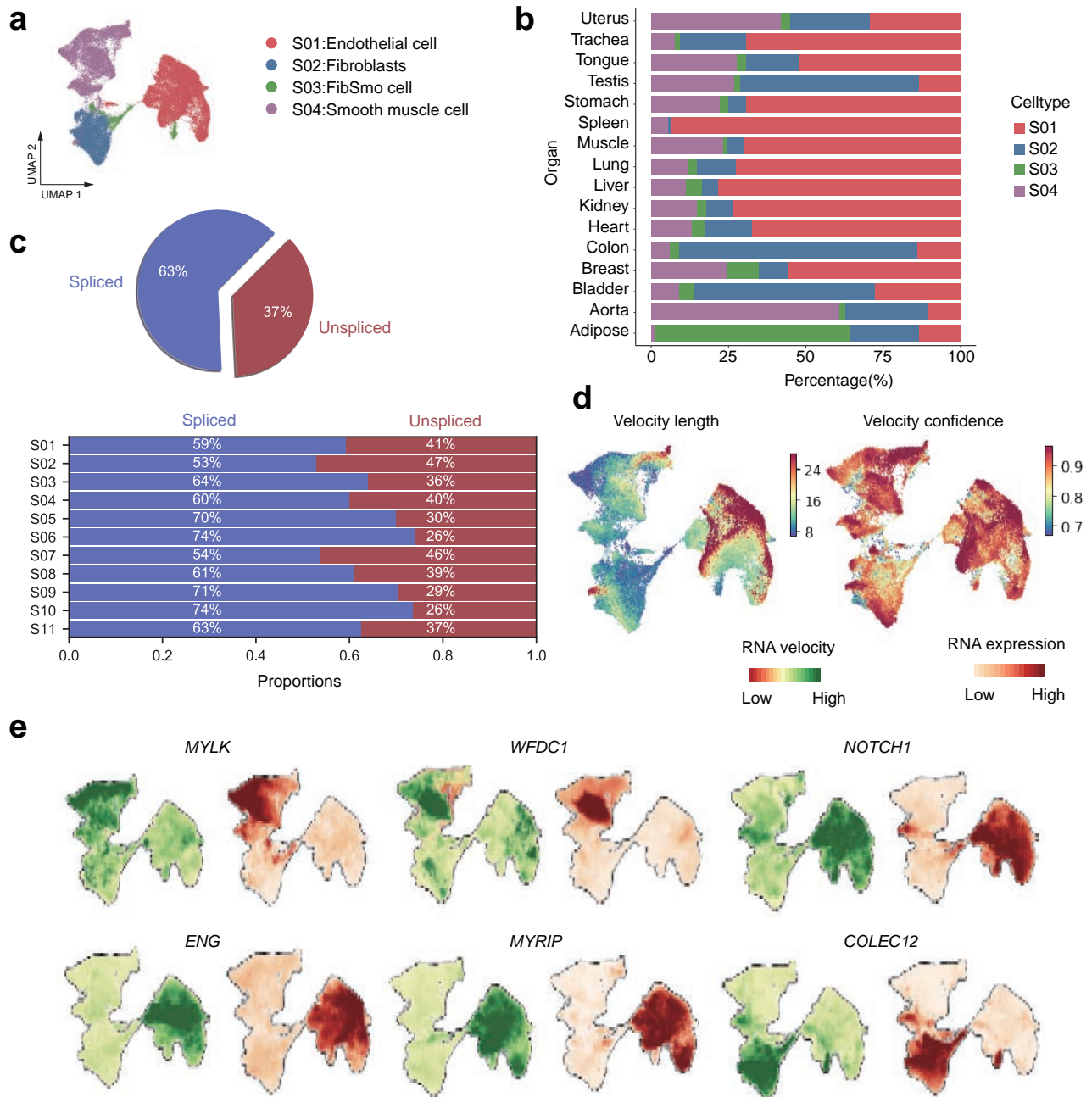
Supplementary Fig. 4: Data quality of scATAC. Related to Figure 1. (a) Box plot showing the distribution of TSS enrichment score, number of fragments and promoter ratio across eight samples. Each dot represents a cell. (b) Fragment size distributions of eight samples (left). Aggregate TSS insertion profiles centred at all TSS regions. The showing cells are passing ArchR QC thresholds for each sample (right). The line colour matches the one in a. (c) Scatter plot showing the TSS enrichment score vs unique nuclear fragments per cell. The colour of the dots represents the density of each point in the plot.



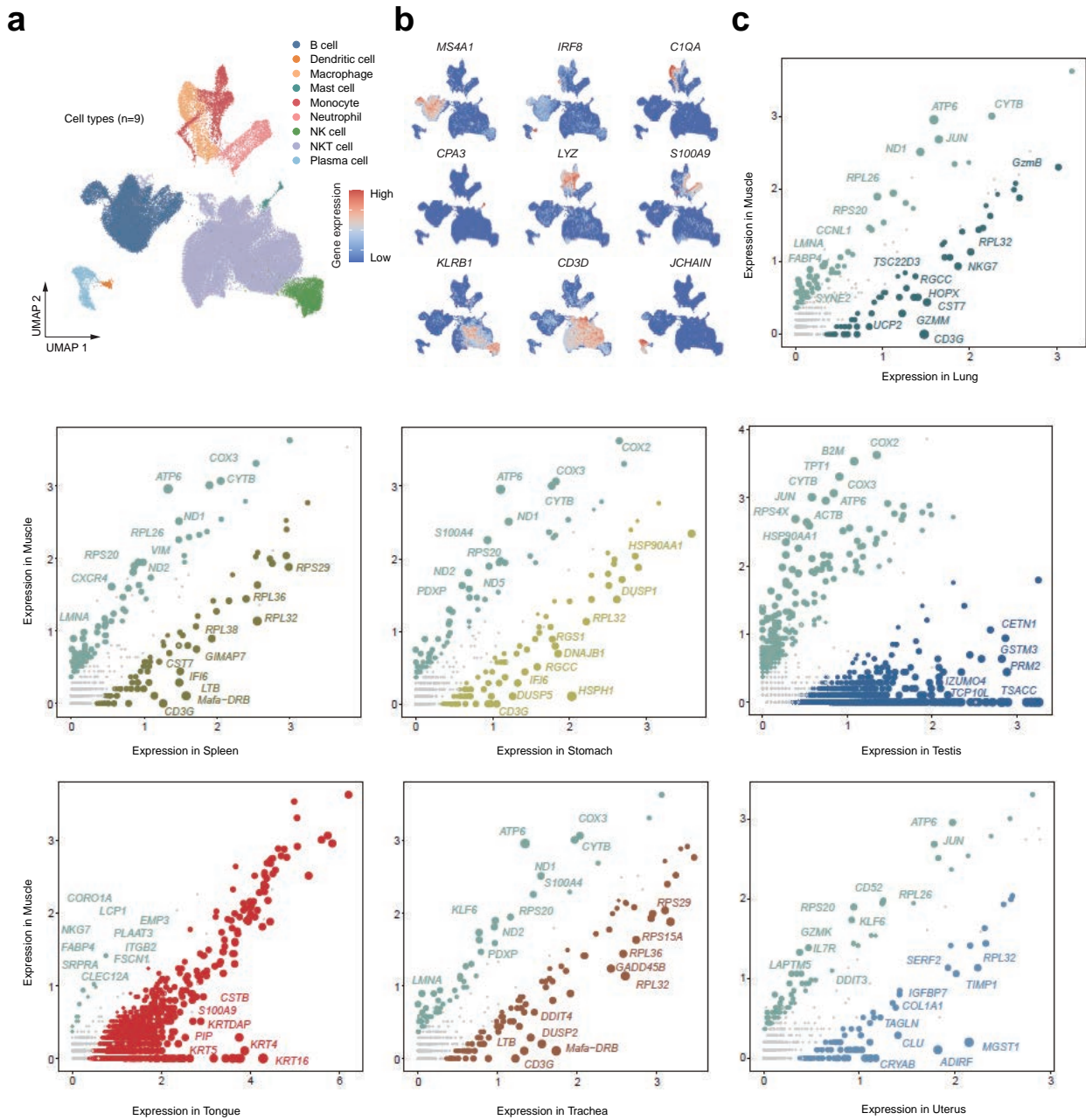
Supplementary Fig. 5: Data quality of scATAC. Related to Figure 1. (a) (left) TSNE of ArchR iterative LSI and (right) ArchR iterative LSI with Harmony-based batch correction for all samples. (b) TSNE embedding showing the subset of seven organs from scRNA-seq. Colored by cell type in Fig.1b. (c) TSNE embedding showing the subset of seven organs from scRNA-seq. Colored by organ. (d) Heatmap showing the spearman correlations between scATAC-seq and scRNA-seq samples. Correlations are calculated from the top 5,000 expressed genes for each sample. (e) UMAP co-embedding of scATAC-seq and scRNA-seq cell. 40,000 cells with scATAC-seq and scRNA-seq were sampled according to the percentage of cell types. (f) Co-embedded cells colored by scRNA-seq cell-type annotations. (g) Scatter plot showing the predicted score distribution for scATAC-seq and scRNA-seq integration. (h) Heatmap showing the spearman correlations between scATAC-seq and scRNA-seq cell types. Correlations are calculated from the top 3,000 expressed genes for each cell type.



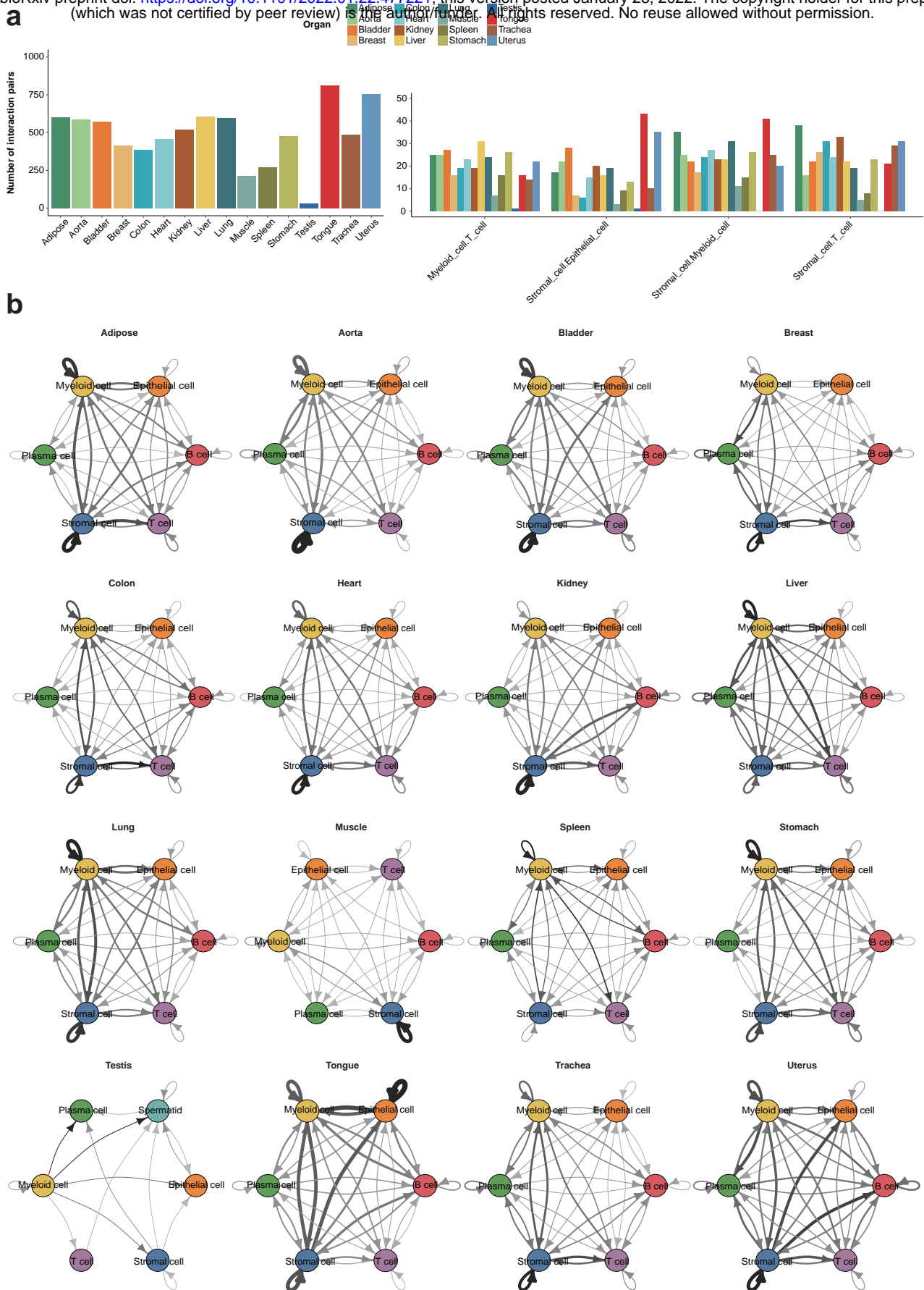
Supplementary Fig. 6: Pseudotime trajectory analysis by monocle2. Related to Figure 2. (a) Distribution of 4 epithelial cell subtypes on the UMAP. (b) Bar plot showing the percentage of cell subtypes in each organ. (c) Distribution of marker genes on branches during pseudotime differentiation trajectory. The shade of the color represents high or low gene scores. (d) Fitting curve showing the dynamic expression levels of CST6, PDZK1IP1, KRT19, PSCA in cell states. The dots represent the distribution of cells along pseudotime. (e) Diagrams showing the density distribution of 14 cell subtypes (16 clusters) during pseudotime differentiation trajectory.

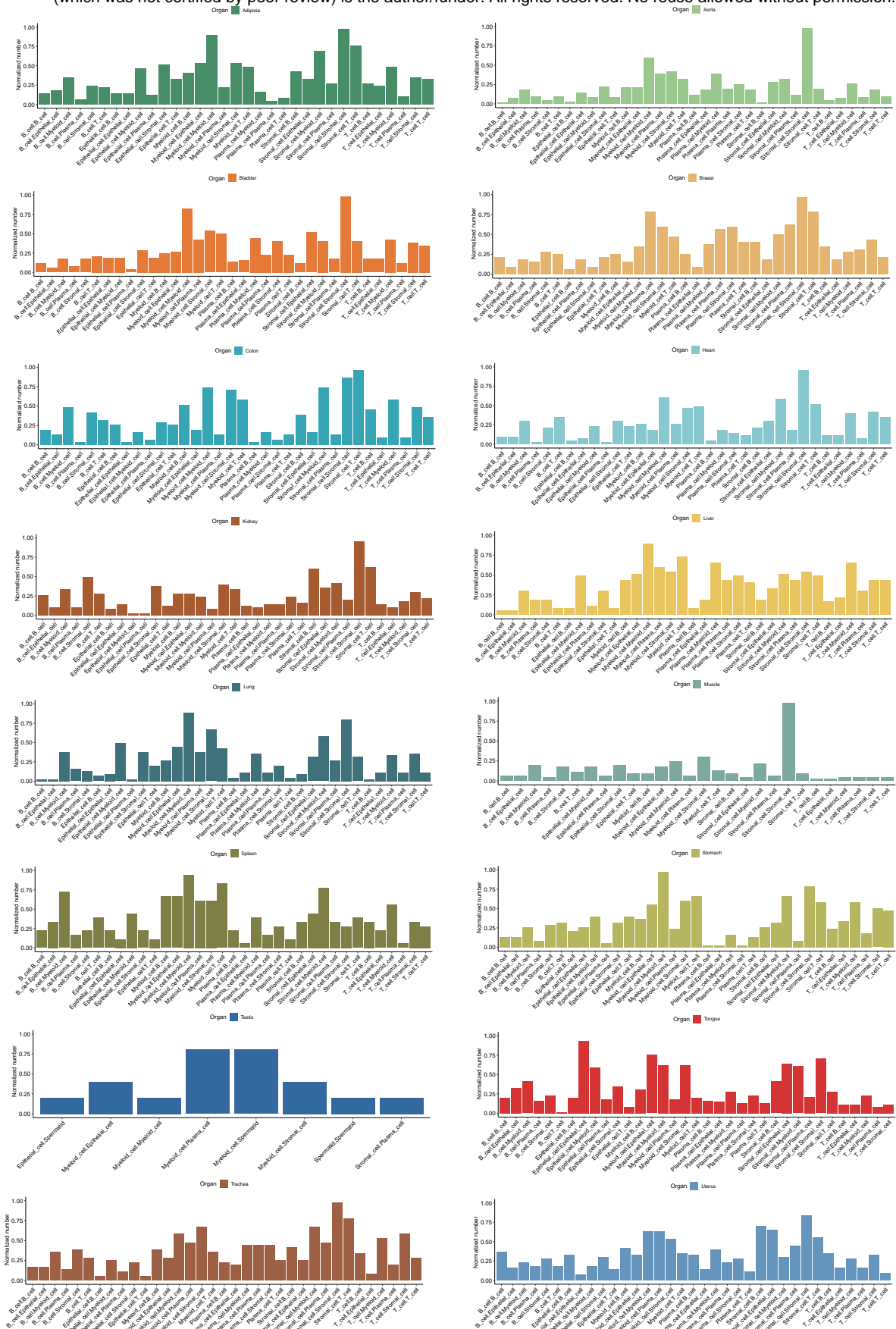


Supplementary Fig. 7: Pseudotime trajectory analysis by RNA velocity. Related to Figure 3. (a) Distribution of 4 stromal cell subtypes on the UMAP. (b) Bar plot showing the percentage of cell subtypes in each organ. (c) Pie chart showing the proportion of stromal cells that are spliced versus un-spliced. Bar plot showing the proportion of stromal cell subtypes that are spliced versus un-spliced. (d) UMAPs diagram showing the velocity length and velocity confidence of stromal cells by RNA Velocity analysis. (e) Distribution of marker genes on the UMAPs by running RNA Velocity.

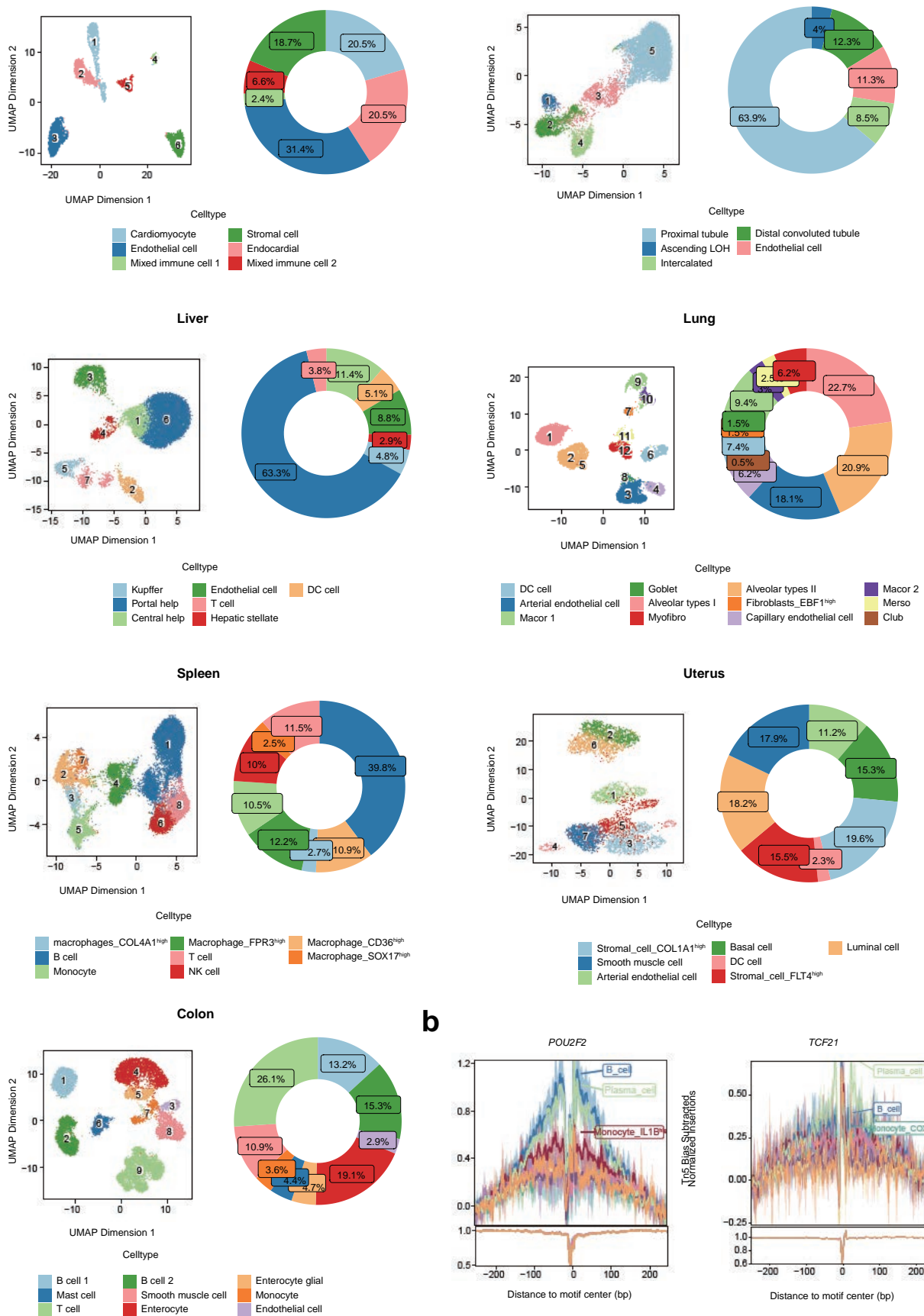


Supplementary Fig. 8: DEGs analysis of immune cell subsets. Related to Figure 4. (a) Distribution of 9 immune cell subtypes on the UMAP. (b) Feature Plot showing the expression of marker genes. (c) Scatter plots shows pairings of gene expression between muscle and other organs. Each point represents a DEG, and its size is proportionate to the fold change.

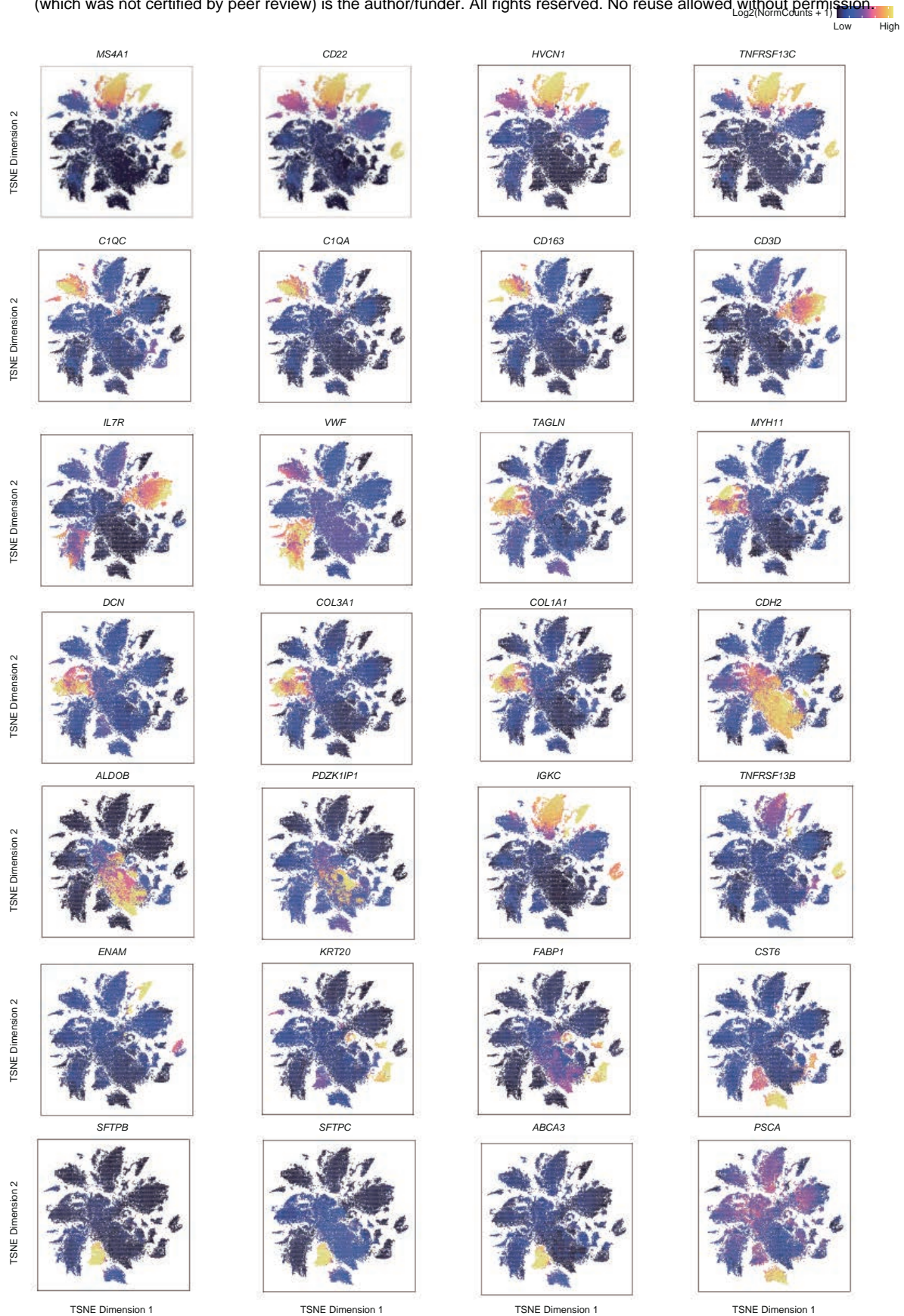




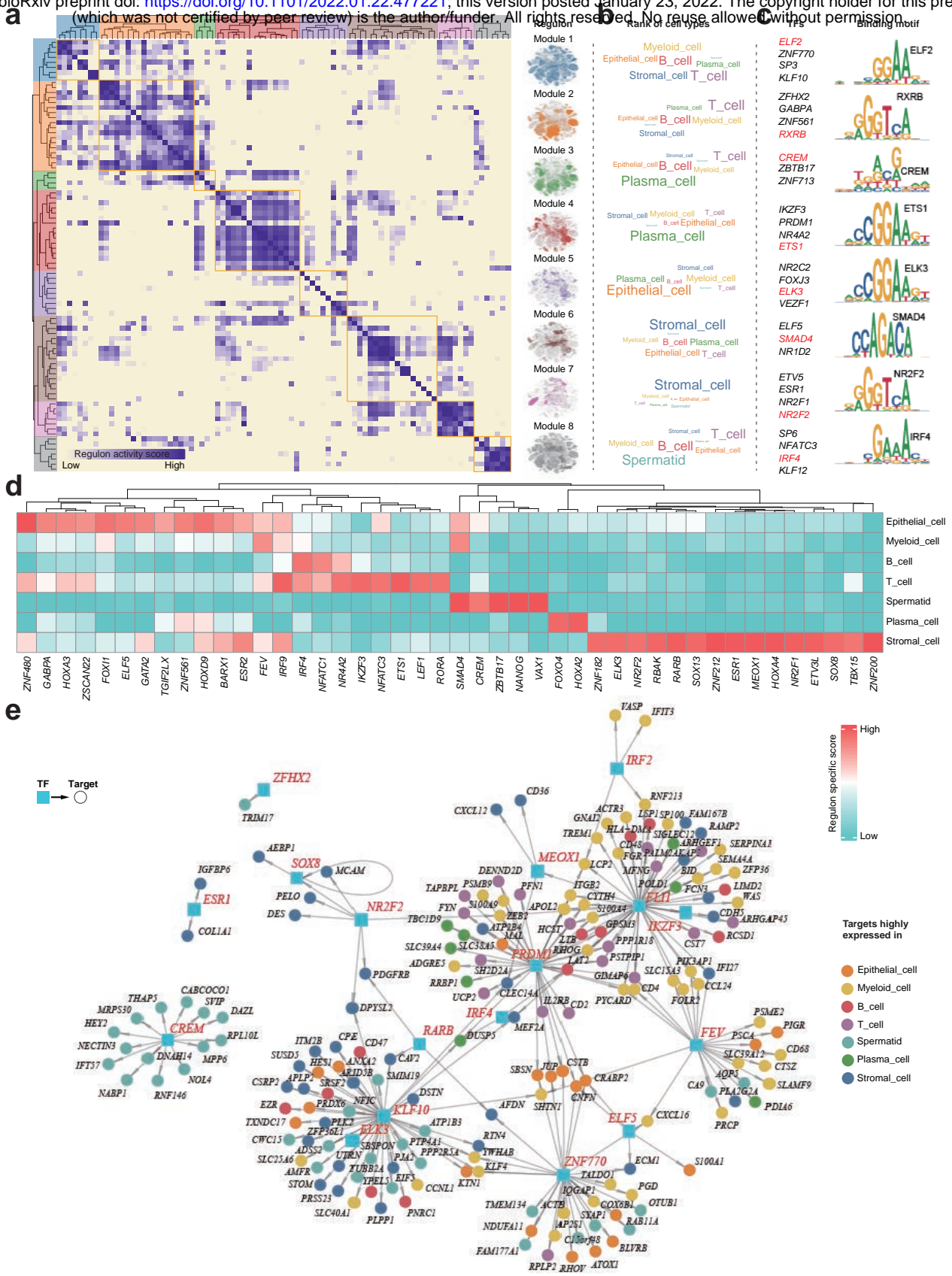
Supplementary Fig. 10: Cell-cell interactions. Related to Figure 5. Bar chart showing the normalized number of interaction pairs of different cell-cell interaction types in 16 organs.



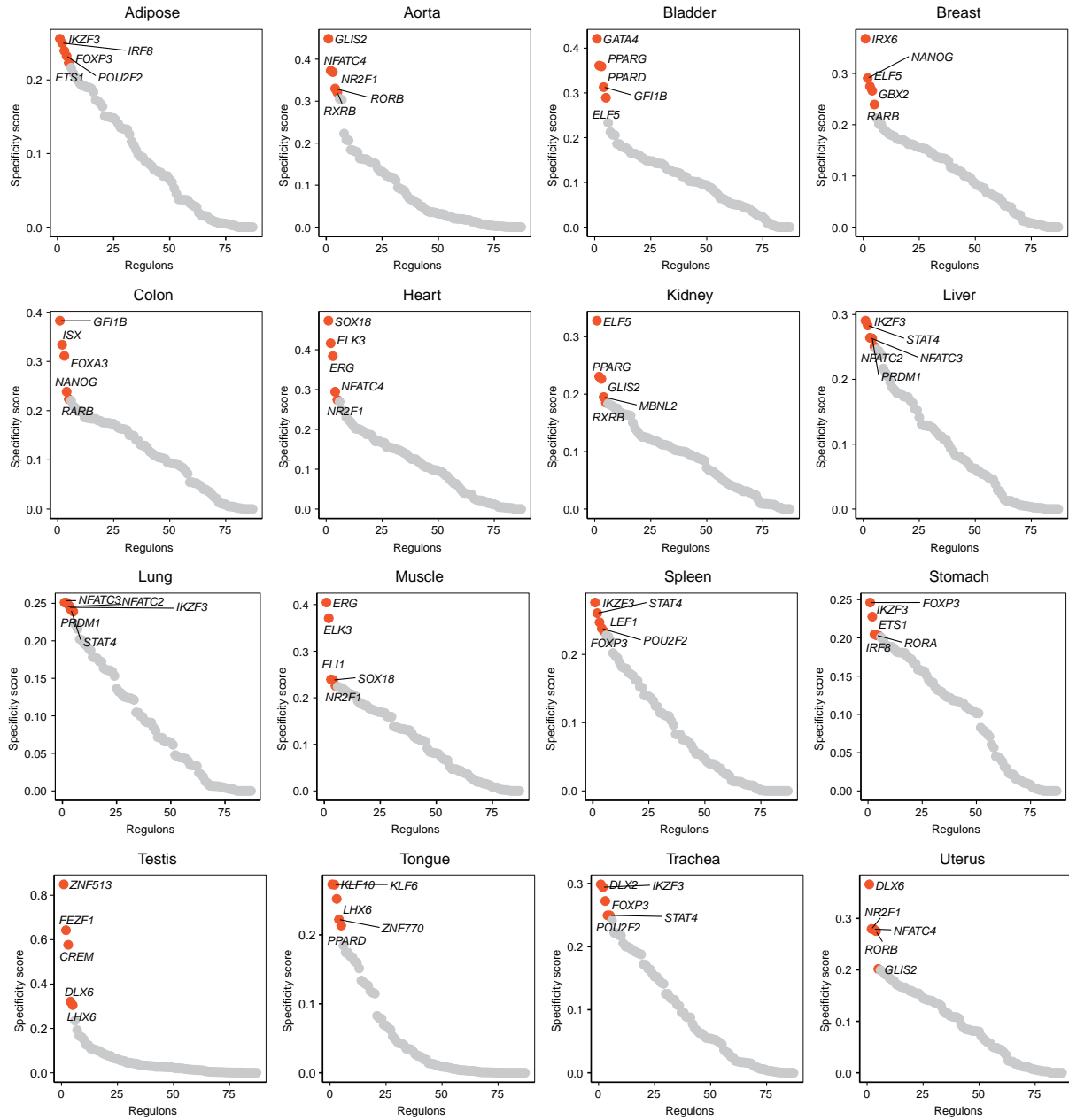
Supplementary Fig. 11: The UMAPs of all organs and Donut charts of the proportion of different cell types (scATAC). Related to Figure 6. (a) Distribution of different clusters on the UMAP in 7 organs (scATAC-seq), and donut chart shows the cell proportion of 7 organs of cell types. (b) Tn5 bias-adjusted TF footprint analysis of the *POU2F2* and *TCF21* transcription factors.



Supplementary Fig. 12: The TSNEs of marker genes expression (scATAC). Related to Figure 6. Distribution of marker genes in cell types such as B cells (*MS4A1*, *CD22*, *HVCN1*, *TNFRSF13C*, *IGKC*, *TNFRSF13B*, *ENAM*), macrophages (*C1QC*, *C1QA*, *CD163*), T cells (*CD3D*, *IL7R*), endothelial cells (*VWF*), fibroblasts (*TAGLN*, *MYH11*, *DCN*, *COL3A1*, *COL1A1*), secretory cells (*CDH2*, *ALDOB*, *PDZK1IP1*), epithelial cells (*KRT20*, *FABP1*) and ciliated cells (*CST6*, *SFTPB*, *SFTPC*, *ABCA3*, *PSCA*) on the TSNEs.

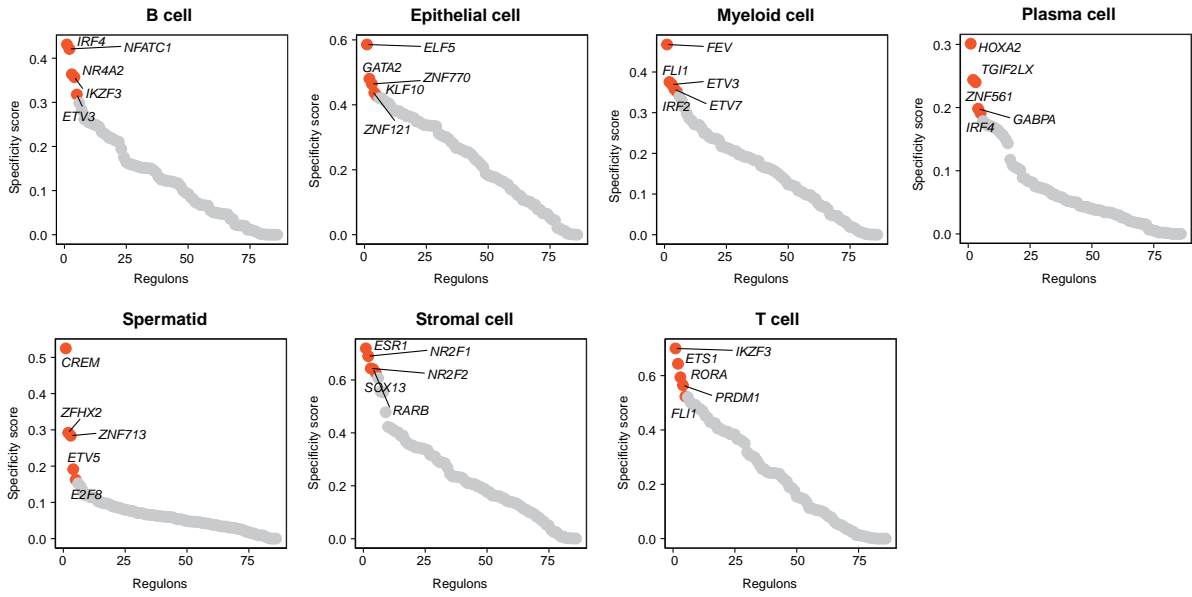


Supplementary Fig. 13: Transcriptional regulatory network of cells from cynomolgus monkey. Related to Figure 7. (a) Identification of regulon modules using SCENIC. Heatmap (left) shows the similarity of different regulons ($n=86$) based on the AUCell score. Eight regulon modules were identified based on regulon similarity. UMAPs (right) illustrate the average AUCell score distribution for different regulon modules (in different colors). (b) Wordcloud plots shows enrichment of cell types in different regulon modules. (c) Representative transcription factors and corresponding binding motifs in different regulon modules. (d) Heatmap showing transcription factors enriched in different cell types. Color depth represents the level of regulon-specific score. (e) Integrated gene-regulatory networks of the regulons from (a). Regulon-associated TFs are highlighted in blue rectangles and target genes in circles. Target genes (in circles) are colored according to their highly expressed cell types.

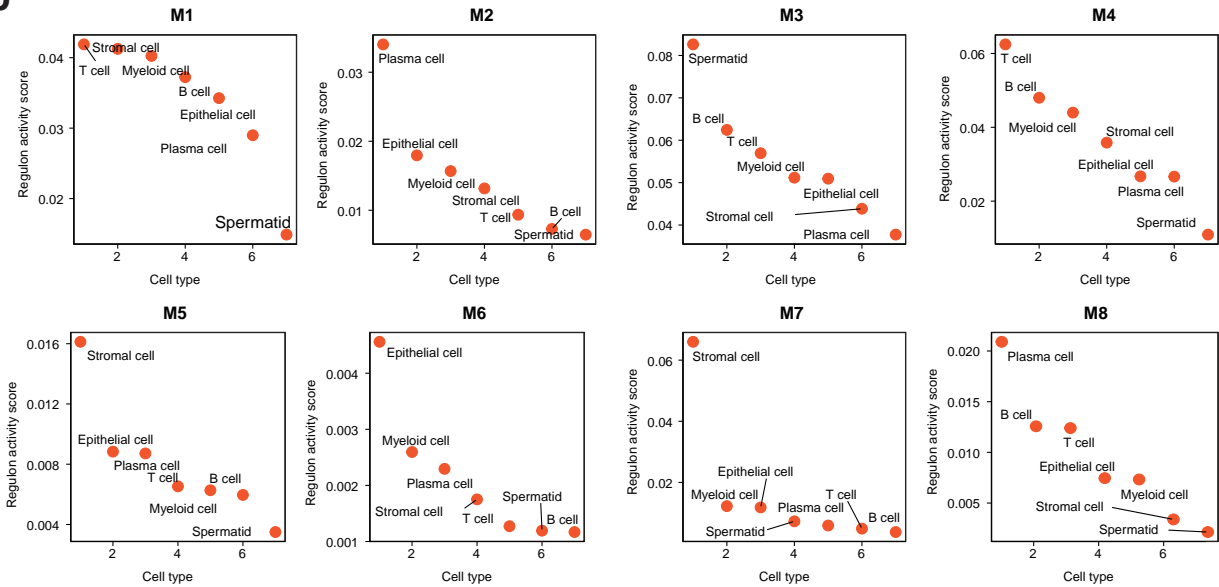


Supplementary Fig. 14: Regulons and modules. Related to Figure 7. Scatter plot showing top regulons of 16 organs ordered by regulon specific score and red dots represent the highest activity.

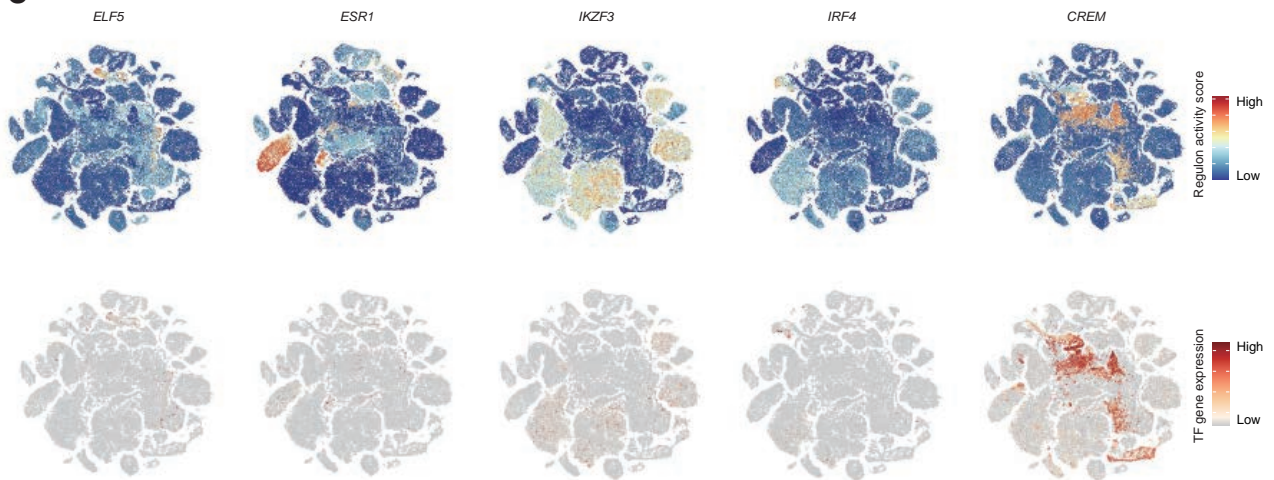
a



b



c



Supplementary Fig. 15: Regulons and modules. Related to Figure 7. (a) Scatter plot showing top regulons of 7 major cell types ordered by regulon specific score and red dots represent the highest activity. (b) Scatter plot showing major cell types of 8 regulon modules ordered by regulon activity score. (c) TSNEs displaying the AUCell score distribution for the selected regulons (above) and gene expression patterns for the corresponding TFs (below).

