Cell transcriptomic atlas of the non-human primate *Macaca fascicularis*

Lei Han[1,2,3,35], Xiaoyu Wei[1,4,35], Chuanyu Liu[1,3,35], Giacomo Volpe[5,35], Zhenkun Zhuang[1,6,35], Xuanxuan Zou[1,4,35], Zhifeng Wang[1,2,35], Taotao Pan[1], Yue Yuan[1,4], Xiao Zhang[7], Peng Fan[7], Pengcheng Guo[7], Yiwei Lai[8], Ying Lei[1,2,3], Xingyuan Liu[7], Feng Yu[8], Shuncheng Shangguan[9], Guangyao Lai[9], Qiuting Deng[1,4], Ya Liu[1,2], Liang Wu[1,2,4], Quan Shi[1,10], Hao Yu[1], Yunting Huang[1,11], Mengnan Cheng[1,4], Jiangshan Xu[1,4], Yang Liu[1,4], Mingyue Wang[1], Chunqing Wang[1,4], Yuanhang Zhang[1,4], Duo Xie[1,4], Yunzhi Yang[12], Yeya Yu[12], Huiwen Zheng[12], Yanrong Wei[12], Fubaoqian Huang[1,6], Junjie Lei[1,4], Waidong Huang[1,4], Zhiyong Zhu[1,4], Haorong Lu[1,11], Bo Wang[1,11], Xiaofeng Wei[1,11], Fengzhen Chen[1,11], Tao Yang[1,11], Wensi Du[1,11], Jing Chen[1,11], Shibo Xu[13], Juan An[8,14], Carl Ward[8], Zongren Wang[15], Zhong Pei[16], Chi-WaiWong[17], Xiaolei Liu[7], Huafeng Zhang[18], Mingyuan Liu[7], Baoming Qin[19], Axel Schambach[20,21], Joan Isern[22], Liqiang Feng[23], Yan Liu[13], Xiangyu Guo[24], Zhen Liu[25], Qiang Sun[25], Patrick H. Maxwell[26], Nick Barker[27], Pura Muñoz-Cánoves[28], Ying Gu[1], Jan Mulder[29,30], Mathias Uhlen[29,30], Tao Tan[31], Shiping Liu[1,2,3], Huanming Yang[1,32], Jian Wang[1,32], Yong Hou[1,2,3,12,✉], Xun Xu[1,33,12,✉], Miguel A. Esteban[7,8,34✉], Longqi Liu[1,3,12✉]

[1]BGI-ShenZhen, Shenzhen 518103, China.

[2]Shenzhen Key Laboratory of Single-Cell Omics, BGI-Shenzhen, Shenzhen 518120, China.

[3]Shenzhen Bay Laboratory, Shenzhen 518000, China.

[4]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China.

[5]Hematology and Cell Therapy Unit, IRCCS-Istituto Tumori 'Giovanni Paolo II', Bari 70124, Italy.

[6]School of Biology and Biological Engineering, South China University of Technology,

Guangzhou 510006, China.

[7]State Key Laboratory for Zoonotic Diseases, Key Laboratory for Zoonosis Research of Ministry of Education, Institute of Zoonosis, College of Veterinary Medicine, Jilin University, Changchun 130062, China.

[8]Laboratory of Integrative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China.

[9]Joint School of Life Sciences, Guangzhou Institutes of Biomedicine and Health and Guangzhou Medical University, Guangzhou 510530, China.

[10]Department of Biology, University of Copenhagen, Copenhagen DK-2200, Denmark.

[11]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China.

[12]BGI College and Henan Institute of Medical and Pharmaceutical Sciences, Zhengzhou University, Zhengzhou 450000, China.

[13]Institute for Stem Cells and Neural Regeneration, School of Pharmacy, State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing 211166, China.

[14]University of Science and Technology of China, Hefei 230026, China.

[15]Department of Urology, First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510000, China.

[16]Department of Neurology, First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510000, China.

[17]Huazhen Biosciences, Guangzhou 510900, China.

[18]Department of Orthopedics, Tianjin Medical University General Hospital, Tianjin 300052, China.

[19]Laboratory of Metabolism and Cell Fate, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China.

[20]Institute of Experimental Hematology, Hannover Medical School, Hannover 30625, Germany.

[21]Division of Hematology/Oncology, Harvard Medical School, Boston MA 02115, USA.

[22]Spanish National Center on Cardiovascular Research (CNIC), Madrid E-28029, Spain.

[23]State Key Laboratory of Respiratory Diseases, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China.

[24]Jinan University, Guangzhou 510632, China.

[25]Institute of Neuroscience, State Key Laboratory of Neuroscience, CAS Key Laboratory of Primate Neurobiology, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China.

[26]Cambridge Institute for Medical Research, Department of Medicine, University of Cambridge, Cambridge CB2 0XY, United Kingdom.

[27]A*STAR Institute of Molecular and Cell Biology, Singapore 138648, Singapore.

[28]Department of Experimental and Health Sciences, Pompeu Fabra University (UPF), ICREA and CIBERNED, Barcelona E-08003, Spain.

[29]Department of Protein Science, Science for Life Laboratory, KTH-Royal Institute of Technology, Stockholm 17121, Sweden.

[30]Department of Neuroscience, Karolinska Institute, Stockholm 17177, Sweden.

[31]State Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming 650500, China.

[32]James D. Watson Institute of Genome Sciences, Hangzhou 310058, China.

[33]Guangdong Provincial Key Laboratory of Genome Read and Write, Shenzhen 518120, China.

[34]Institute of Stem Cells and Regeneration, Chinese Academy of Sciences, Beijing 100101, China.

[35]These authors contributed equally to this work.

✉Corresponding authors. Email: houyong@genomics.cn (Y.H.), xuxun@genomics.cn (X.X.), miguel@gibh.ac.cn (M.A.E.), liulongqi@genomics.cn (L.L.).

Studying tissue composition and function in non-human primates (NHP) is crucial to understand the nature of our own species. Here, we present a large-scale single-cell and single-nucleus transcriptomic atlas encompassing over one million cells from 43 tissues from the adult NHP *Macaca fascicularis*. This dataset provides a vast, carefully annotated, resource to study a species phylogenetically close to humans. As proof of principle, we have reconstructed the cell-cell interaction networks driving Wnt signalling across the body, mapped the distribution of receptors and co-receptors for viruses causing human infectious diseases and intersected our data with human genetic disease orthologous coordinates to identify both expected and unexpected associations. Our *Macaca fascicularis* cell atlas constitutes an essential reference for future single-cell studies in human and NHP.

MAIN TEXT

Global initiatives such as the Human Cell Atlas are aiming to chart the cell types and cell states of all tissues in the human body using high-throughput single-cell/nucleus RNA-sequencing (sc/snRNA-seq) and other technologies[1-5]. The ultimate goal of these efforts is to create complete reference maps across different ethnic groups, ages, environmental conditions and pathologies. A major obstacle in this endeavour is that accessing a wide range of 'high quality' human samples and obtaining enough sample size is complicated by relevant practical and ethical considerations. Model animals (e.g., mouse and rat) are a useful resource to fill knowledge gaps[6-8], in particular the effects of experimental perturbation, but due to profound phylogenetic differences many developmental, physiological and pathological aspects are not mimicked in humans. Given the evolutionary proximity, NHP present an excellent alternative (the nearest-to-human) when no other suitable models exist. Generating a NHP cell atlas will produce an extensive catalogue of human disease and age-related features that can be modelled in NHP. It will also provide unique insights into the evolutionary and adaptative mechanisms underlying changes in body function between the two species. In this regard, it could for example discover tissue regenerative capacities selectively maintained in NHP and potential ways to boost them in human.

NHP encompass a large and very diverse group of species with major ecological, dietary, locomotor and behavioural differences[9-11]. Because of their close evolutionary proximity to humans among NHP, overall characteristics and wider availability, macaques are primarily employed for research purposes worldwide including human disease modelling and preclinical safety assessment studies[12,13]. Here, we have used adult *Macaca fascicularis* (cynomolgus monkey) to generate the largest single-cell transcriptomic NHP dataset to date, encompassing over 1 million individual cells/nuclei from 43 tissues covering all major systems (nervous, immune, endocrine, cardiovascular, respiratory, digestive, skeletal, reproductive and urinary), all performed

with the same droplet-based approach[14]. To facilitate the exploration of this dataset, we have created the first version of the Non-Human Primate Cell Atlas or NHPCA, an open and interactive database (https://db.cngb.org/nhpca/) that will be regularly updated with subsequent sc/snRNA-seq *Macaca fascicularis* datasets focused on development, aging, disease and drug responses, as well as other omics datasets and data from other NHP species.

Generation of an adult monkey single-cell transcriptomic atlas

We isolated cells/nuclei from 43 different tissue samples from three male and three female six-year-old *Macaca fascicularis* monkeys (**Fig. 1a and Supplementary Table 1a**). Bladder (two), cerebellum (two), diaphragm (two), gallbladder (two), kidney (two), liver (three), lung (two), salivary gland (two), subcutaneous (two) and visceral adipose tissue (two) were analyzed as biological replicates to assess individual and gender variability, observing good overlap in all cases (**Extended Data Fig. 1**). Most of the tissues were profiled by snRNA-seq[15-17], which allows both to circumvent complications associated with stressful dissociation protocols that can alter the cell transcriptome and to profile cells from frozen tissues for removing the need of sample processing immediately after tissue acquisition. However, due to technical limitations in obtaining high quality nuclei, scRNA-seq was performed for colon, duodenum, spleen, stomach, lymph node and bone marrow. Peripheral blood mononuclear cells (PBMC) were also profiled using scRNA-seq. All experiments used the DNBelab C4 droplet-based platform for library generation[14]. To ensure quality, all cells with a gene count lower than 500 and/or mitochondrial content higher than 10% were excluded. We also applied DoubletFinder to detect and remove doublets, which accounted for roughly 5% of the estimated total cell/nuclei. Overall, we retained transcriptomic data for a total of 1,084,164 cells/nuclei (**Fig. 1a**), with numbers ranging from 99,123 in the cerebellum to 2,039 in the duodenum (**Supplementary Table 1a**). Global visualization of cell clustering using Uniform Manifold Approximation and Projection (UMAP) showed

that each tissue clusters separately, with tissues from the same system generally clustering closer (**Fig. 1a, b and Extended Data Fig. 2-6**). We then performed individual UMAP representations for each tissue and applied unbiased graph-based Seurat clustering, which identified 463 cell clusters among all tissues (**Extended Data Fig. 7-10**). Based on the expression levels of cell type-specific markers (**Extended Data Fig. 11**), we identified 106 cell types in the global UMAP view of all tissues (**Fig. 1c and Supplementary Table 1b, c**). These were roughly categorized into epithelial cells (40 clusters), immune cells (13 clusters), endocrine cells (11 clusters), muscle cells (9 clusters), stromal cells (7 clusters), endothelial cells (7 clusters), neurons (7 clusters), glia (7 clusters), mesothelial cells (3 clusters), adipocytes (1 cluster) and unknown cells (1 cluster from carotid). On average, we detected 1,368 genes and 3,024 unique molecular identifiers (UMI) per cell. The median gene count per tissue varied between 3,016 in the neocortex and 736 in the case of PBMC, while UMI ranged between 8,015 for the neocortex and 1,313 for the prostate (**Extended Data Fig. 12**). The number of cells for each of these 106 cell types ranged from 87,890 granule cells in the cerebellum to 37 bone marrow stromal cells (**Extended Data Fig. 13**). Reassuringly, many of the 106 clusters were largely composed of a cell type belonging to a specific tissue, such as cerebellar granule cells in cluster 45, hepatocytes in clusters 87 and 88, epididymis stereociliated cells in cluster 29 and salivary acinar cells in cluster 83 (**Fig. 1c and Extended Data Fig. 14a**). However, cell types such as endothelial, stromal and various immune cells were shared between different tissues, as expected (**Extended Data Fig. 14b**). A detailed annotation of all cell populations detected in every tissue is provided in **Extended Data Figure 7-10 and Supplementary Table 1d, e**. Our *Macaca fascicularis* atlas is the largest NHP single-cell transcriptome dataset to date and can be explored interactively by tissue, cell type and gene through our NHPCA database.

Common cell types across monkey tissues

We inspected whether common cell types distributed throughout different tissues in the monkey body display tissue-specific transcriptional programs[3,18-20]. First, we selectively clustered stromal cells, macrophages (including microglia), endothelial cells and smooth muscle cells from all sequenced tissues. While observing a considerable diversity, many cell clusters grouped together on the basis of tissue origin, such as stromal cells from the female reproductive system, microglia from the central nervous system, endothelial cells from the respiratory system and smooth muscle cells from the male reproductive system (**Extended Data Fig. 15a-d**). We also performed differentially expressed gene (DEG) analysis to obtain tissue-specific signatures, revealing a substantial heterogeneity among these common cell types across all tissues (**Extended Data Fig. 15e-h and Supplementary Table 2a-d**).

Our transcriptomic profiling of single nuclei offers the possibility of studying cell populations that cannot be characterized by conventional scRNA-seq analysis, such as myonuclei from multinucleated skeletal muscle fibers. We grouped and re-clustered cells from tissues in our atlas known to contain skeletal muscle cells (diaphragm, tongue, esophagus and abdominal wall). This showed two distant populations in abdominal wall and diaphragm, whereas nuclei from esophagus and tongue where more concentrated (**Fig. 2a**). The separation of nuclei in abdominal wall and diaphragm corresponded to *MYH7*$^{+}$ type I (slow-twitch) and *MYH2*$^{+}$ type II (fast-twitch) myofibers[21] (**Fig. 2b, c and Supplementary Table 2e-g**). In contrast, type I and type II tongue myonuclei were in close vicinity, which may be related to the tongue being a highly innervated muscle[22]. Differential threshold of *MYH2* and *GPD2* further subdivided type II myonuclei into type IIa (*MYH2*^{high}) and type IIb (*MYH2*^{low} *GPD2*$^{+}$). In addition, we discriminated, albeit at low proportions, *NAV3*$^{+}$ neuromuscular junction (NMJ) nuclei in the diaphragm and *ETV5*$^{+}$ myotendinous junction (MTJ) nuclei in both tongue and diaphragm (**Fig. 2b-d**). Moreover, we detected *PAX7*$^{+}$ nuclei from satellite cells (the stem cells from the skeletal muscle lineage), and a small cluster of *LVRN*$^{+}$ fibroadipogenic progenitors (FAP) could be annotated in the diaphragm. Skeletal

muscle nuclei displayed subtype-specific and tissue-specific gene expression signatures and gene ontology (GO) terms (**Fig. 2e, f and Extended Data Fig 16a-c**). We also noticed substantial myonuclei heterogeneity within the same subtype and tissue (**Fig. 2f**).

Next, to explore the heterogeneity between different types of adipocytes, we grouped and re-clustered cells from subcutaneous and visceral adipose tissues, resulting in 10 major clusters (**Extended Data Fig. 17a**). We observed a marked distinction between mature adipocytes and adipocyte progenitors, as reflected by the differential expression of *ADIPOQ* and *CD34* (**Extended Data Fig. 17b**). Visceral mature adipocytes and adipocyte progenitors displayed enriched expression of *ITLN1*, in agreement with visceral adipocytes having mesothelial origin[23], and also high mitochondrial activity exemplified by high expression of *ND4*, *ATP6* and *COX3* [24,25] (**Extended Data Fig. 17c, d**). In contrast, subcutaneous mature adipocytes and adipocyte progenitors were enriched in *FOS*. Likewise, *SLC11A1* and *SPOCK3* marked mature subcutaneous and visceral adipocytes, respectively. Adipocyte progenitors contained two populations for visceral tissue (*WT1*$^{+}$ and *CFD*^{high}), three for subcutaneous tissue (*ESR1*$^{+}$, *CXCL14*$^{+}$*APOD*$^{+}$ and *DPP4*$^{+}$) and one shared between both tissues (*NOX4*$^{+}$) (**Extended Data Fig. 17a, c and d**). Within the subcutaneous *CXCL14*$^{+}$*APOD*$^{+}$ progenitor cluster, we observed a population of *CFD*^{high} cells that also co-expressed *DPP4*, a marker of highly proliferative adipocyte progenitors in both mouse and human[26]. However, we did not detect significant proliferation in any of the monkey adipocyte progenitor populations based on the expression of the pan-cycling marker *MKI67*[27] (**Extended Data Fig. 17c**). *NOX4*$^{+}$ is an NAPDH oxidase that acts as a switch from insulin-induced proliferation to adipocyte differentiation, suggesting that the shared cluster is a converging route for both adipose tissues towards adipocytic maturation[28].

Finally, we grouped and re-clustered all tissues that contain mesothelial cells, a type of specialized epithelial cells. Mesothelial cells from bladder, ovary and fallopian tube

were in close proximity while those from other tissues clustered more separately (**Fig. 2g**). We also observed intra-tissue heterogeneity, in particular for visceral adipose tissue and ovary. In the former, we observed a cluster of immune-like mesothelial cells that, aside from the expression of the typical mesothelial markers (*MSLN*, *ITLN1* and *PKHD1L1*), express high levels of immune cell markers (e.g., *PTPRC*, *IL7R* and *TRAC*) (**Fig. 2h**). This is in agreement with the emerging concept that structural cells display immune cell properties[3,18] and the known immunomodulatory role of visceral adipose tissue in responses to bacteria in the gut[29]. Interestingly, in the ovary, we identified a classical mesothelial population and two close *PAX8*$^{+}$ epithelial-like populations (one mature and one progenitor-like) of mesothelial origin[30] (**Fig. 2i-k**). Progenitor-like ovarian epithelial cells expressed well-known stem cell markers such as *LGR5*, *MECOM* and *CD44*[31].

These findings add up to the growing understanding of common cell type heterogeneity and tissue-specific molecular signatures[3,18-20]. Our data provide a new resource for further dissecting these differences, clarifying the underlying mechanisms and studying interspecies differences[32].

Analysis of Wnt signaling components identifies potential stem cell populations

A single-cell body atlas of large dimensions like ours is ideal for the systematic investigation of multifaceted cell-cell interactions including those occurring in cytokine or growth factor-mediated signaling pathways such as the Wnt (wingless-related integration site) pathway[33,34]. Besides playing essential roles in embryonic development, Wnt factors control growth and maintenance of numerous tissues throughout life. Consistently, Wnt signaling effects are associated with the regulation of adult stem cell function[35]. To exert this role, Wnt factors bind to specific receptors (FZD, frizzled) and co-receptors (LRP, low-density lipoprotein receptor related protein). In addition, LGR (leucine rich repeat containing G protein-coupled receptor) proteins (LGR4, 5 and 6) act as amplifiers of Wnt signals by inhibiting negative regulators[36]. Accordingly,

LGR5 and 6 often mark and regulate adult homeostatic and facultative stem cells, mostly of epithelial origin, in multiple mammalian tissues, whereas LGR4 has a widespread distribution and less clear function. We thus performed a survey of LGR proteins throughout the monkey body to thoroughly dissect cells targeted by the Wnt pathway and identify previously unappreciated stem cell populations. In this regard, it is worth noting that the majority of reports of LGR5-expressing cells to date have been performed with genetically engineered mouse models due to the lack of specific tools and reagents to study other mammals[36].

LGR5 was detected across several monkey tissues, unexpectedly with the highest expression in type I skeletal muscle myonuclei, epithelial cells of the uterus and fallopian tube, oligodendrocyte progenitor cells (OPC) and renal distal convoluted tubule cells (DCTC) (**Fig. 3a**). With the exception of epithelial cells in the uterus and fallopian tube[36], these tissues have not previously been reported to contain LGR5$^+$ cells in mammalian adulthood. The expression of *LGR6* appeared to be more restricted (**Extended Data Fig. 18a**), with higher abundance in cardiomyocytes, thyroid follicular cells, folliculostellate cells of the pituitary gland and the previously reported smooth muscle cells[37] **(Extended Data Fig. 19-22)**. We also detected *LGR5*$^+$ or *LGR6*$^+$ cells in selected cell populations of numerous other tissues including both previously reported (e.g., ovary epithelial cells[31], hepatocytes[38] and colon enterocytes[39]) and unreported (e.g., *LGR5*$^+$ cells in bipolar cells of the retina[40]) (**Fig. 3a, Extended Data Fig. 18a and 19-22**). In general, *LGR5* and *LGR6* did not overlap, apart from fallopian tube epithelial cells and vagina smooth muscle cells (**Extended Data Fig. 18b)**. Moreover, we observed little overlap between *LGR5*$^+$ or *LGR6*$^+$ cells with those expressing *MKI67*, apart from epithelial cells of the fallopian tube and uterus and basal cells of the salivary gland **(Extended Data Fig. 19-22 and Supplementary Table 3a-c)**. In contrast to *LGR5* and *6*, *LGR4* was ubiquitously expressed across most tissues, with the highest expression in pancreatic acinar, beta and ductal cells, Müller cells of the retina and adipocytes (**Extended Data Fig. 18c**).

In the kidney, *LGR5*$^{+}$ cells were mostly enriched in the DCTC and to a lesser extent in the descending and ascending loop of Henle (**Fig. 3a and Extended Data Fig. 20**). To support this observation, we performed single-cell Assay for Transposase Accessible Chromatin sequencing (scATAC-seq) of monkey kidney and integrated the results with our kidney snRNA-seq data dataset (N = 6,879) (**Fig. 3b, c and Extended Data Fig. 23a, b**). The analysis showed peaks of open chromatin at the *LGR5* promoter and a putative enhancer open in the same cell types expressing *LGR5* (**Fig. 3d**). As validation, we performed single-molecule fluorescence *in-situ* hybridization (smFISH) for *LGR5*, which showed strong expression in selected kidney tubules (**Fig. 3e**). Moreover, GO analysis of DEG comparing the *LGR5*$^{+}$ fractions of DCTC, ascending and descending loop of Henle revealed the enrichment of pathways involved in kidney development in DCTC (**Fig. 3f**), suggesting the possibility that these are progenitor cells. This was strengthened by the observation that DCTC *LGR5*$^{+}$ cells co-express renal progenitor cell markers such as *PAX2*, *LHX1* and *TNFRSF19*[41,42]. We also integrated our data with available human[43] and mouse[44] kidney snRNA-seq datasets. Despite observing good integration, we noticed very little, or no, *LGR5* expression in those adult human or mouse kidney datasets[45] (**Extended Data Fig. 24a-c**).

In the neocortex, integration of available human[46] and our own mouse snRNA-seq data with our monkey data pointed as well at differential *LGR5* expression patterns between species. *LGR5* expression was highest in OPC in monkey and in oligodendrocytes in human, whereas in mouse it was higher in inhibitory neurons than OPC and oligodendrocytes (**Extended Data Fig. 25a-c**). Pseudotime ordered by Monocle 2 of the OPC maturation trajectory towards oligodendrocyte showed concentration of *LGR5* in monkey OPC (**Extended Data Fig. 25d, e**). Likewise, double immunofluorescence for the OPC marker PDGFRA and LGR5 confirmed their co-expression in OPC from monkey neocortex (**Extended Data Fig. 25f**). The observation that type I skeletal myonuclei and cardiomyocytes ranked first in expression of *LGR5* and *LGR6* in monkey tissues, respectively, was intriguing (**Fig. 3a and Extended Data**

Fig. 18a). To inspect this further, we grouped and re-clustered all types of muscle cells (skeletal, smooth and cardiac) in our atlas (**Fig. 3g**). *LGR5* was more enriched in *MYH7*$^{+}$ slow-twitch myonuclei of the abdominal wall and diaphragm (**Fig. 3h**), whereas *LGR6* was higher in cardiomyocytes and smooth muscle cells (aorta, ovary, carotid and vagina) (**Extended Data Fig. 26a**). *LGR5* and *LGR6* expression in slow-twitch skeletal myonuclei and in cardiomyocytes, respectively, were validated by smFISH (**Fig. 3i and Extended Data Fig. 26b**). In mouse, *LGR5* is known to be expressed in NMJ myonuclei[47] and a subset of satellite cells activated upon injury[48], but we did not detect *LGR5* enrichment in either cell type in our monkey dataset (**Extended Data Fig. 19**). The lack of *LGR5* enrichment in monkey satellite cells is unsurprising given that we did not apply any injury to the skeletal muscle tissues profiled. Yet, we could detect *LGR6* in cardiomyocytes using previously reported mouse and human snRNA-seq datasets[49,50] (**Extended Data Fig. 26c, d**). Similarly, *LGR6* was enriched in several monkey pituitary cell populations, being most highly expressed in folliculostellate cells, which have been reported to be pituitary gland stem cells[51] (**Extended Data Fig. 26e**). Consistently, those cells also expressed other progenitor markers such as *SOX2*, *PAX6*, *CD44* and *CXCR4* (**Extended Data Fig. 26f**). Moreover, GO analysis of DEG specific to this *LGR5*$^{+}$ population compared to other pituitary cells showed enrichment of terms related to development (**Extended Data Fig. 26g**).

Next, we profiled the genes encoding Wnt factors and the R-spondin family (RSPO1-4) of ligands for LGR proteins[35,36] in a panel of monkey tissues containing cells with high *LGR5* (kidney, epididymis, fallopian tube, liver, ovary, neocortex and diaphragm) and *LGR6* (heart and pituitary gland) expression (**Extended Data Fig. 27a, b and 28-31**). This allowed us to dissect the potential cell-cell interaction networks driving Wnt signalling throughout the monkey body. Notably, RSPO cytokines were widely distributed but displayed higher expression in mesenchymal-like cells (e.g., smooth muscle cells of epididymis, hepatic stellate cells and folliculostellate cells of the pituitary gland) and mesothelial cells (e.g., of diaphragm, fallopian tube and ovary)

of different tissues. Interestingly, *RSPO2* was also high in inhibitory neurons of the neocortex (**Extended Data Fig. 30**). The expression of Wnt factors was more limited and in general lower than RSPO cytokines but we noticed high levels of *WNT9B* in principal cells of the collecting duct in kidney (**Extended Data Fig. 27a, c**), *WNT2B* in mesothelial cells of the fallopian tube (**Extended Data Fig. 29a**) and ovary (**Extended Data Fig. 30c**), and as expected *WNT2* in endothelial cells of the liver[52] (**Extended Data Fig. 29c**). Wnt9b is an essential regulator of kidney embryonic development in multiple species and of kidney regeneration in lower vertebrates[53]. Supporting the snRNA-seq data, scATAC-seq analysis of the *WNT9B* locus revealed increased enhancer accessibility in monkey principal cells compared to other kidney cell types (**Extended Data Fig. 27d**). In contrast, we detected low *WNT9B* expression in available mouse[44] and human[46] snRNA-seq datasets (**Extended Data Fig. 27e**). WNT9B may be responsible for inducing *LGR5* (a Wnt pathway target) in a fraction of DCTC, potentially creating a feedback loop that amplifies WNT9B signals to keep those cells in a progenitor state. In fact, Wnt factors are known to act predominantly on neighbouring cells[33,35], and cells of the collecting duct and DCTC are in closer proximity than other nephron structures (**Extended Data Fig. 27f**). We further included Wnt receptors and other co-receptors[54] in the analysis, and also the TCF family of transcription factors bound by β-catenin[55], as a resource for additional exploration in these tissues (**Extended Data Fig. 27a,b and 28-31**).

Therefore, we have reconstructed the Wnt signaling network in monkey tissues and identified cell types with potential progenitor or homeostatic characteristics. Additional signaling pathways and/or ligand-receptor interactions can be explored through our NHPCA database.

Prediction of viral infection vulnerability in monkey tissues

To demonstrate the utility of our atlas for advancing the knowledge of disease pathogenesis, we first mapped the expression of the main viral receptors/co-receptors

for a panel of 126 viruses including respiratory ones across all monkey tissues. As expected, *NCAM1* (cytomegalovirus receptor) was enriched in astrocytes, oligodendrocytes and neurons, consistent with the knowledge of this virus attacking the central nervous system[56]. In contrast, *CD46*[57] (receptor for Measles and Herpes viruses) was enriched in epithelial cells from bladder, female and male reproductive system, and liver endothelial cells (**Fig. 4a, Extended Data Fig. 32 and Supplementary Table 4a**). Given the emergency state of the current COVID-19 pandemic caused by SARS-CoV-2[58], we focused on its receptor *ACE2* and co-receptor *TMPRSS2*[59] to assess how widespread and homogeneous their expression is in monkey tissues. This offers the major advantage of studying COVID-19 pathogenesis in a species phylogenetically close to humans[60], and also provides the possibility of profiling cell types and/or tissues that have not been studied in human. In this regard, although the lung is the predominantly affected tissue in COVID-19, it is important to clarify what other tissues are targeted to better understand the disease course and its transmissibility[61]. *TMPRSS2* displayed a broad expression across multiple monkey tissues, whereas *ACE2* had a more restricted pattern. The highest *ACE2* expression was found in epithelial cells from gallbladder (glandular cells), kidney (mostly proximal tubule cells), lung (ciliated, club and alveolar type 2 [AT2] cells) and liver (hepatocytes and cholangiocytes) (**Fig. 4b, Extended Data Fig. 33, 34 and Supplementary Table 4b**). *ACE2* in these tissues was remarkably heterogeneous, suggesting that regulatory mechanisms fine-tune its expression levels. Notably, double positive (*ACE2*$^{+}$ *TMPRSS2*$^{+}$) cells have a higher risk of infection by SARS-CoV-2[59] but it remains unclear what tissues and cell types throughout the human body co-express these genes. We noticed the largest overlap between *ACE2* and *TMPRSS2* in monkey gallbladder cells in agreement with reports of COVID-19 patients developing acute cholecystitis[62]. Significant co-expression was also observed in ciliated and club cells of the lung, as expected[63,64], and, interestingly, proximal and connecting tubule cells of the kidney. A smaller overlap was observed in hepatocytes, bladder epithelial cells and pancreatic beta and ductal cells (**Fig. 4c**). Next,

we performed a comparative analysis of *ACE2* and *TMPRSS2* distribution in human[3,6,43] and monkey. A similar distribution was seen in both the gallbladder and liver in the two species, while distinct patterns were observed for proximal tubule cells of the kidney and for ciliated and AT2 cells of the lung (**Extended Data Fig. 35a**). This is important because it implys a mechanism by which the infection with SARS-CoV-2 in the two species could have different consequences.

As a representative tissue with high but heterogeneous *ACE2* expression and a significant proportion of *ACE2*$^{+}$ *TMPRSS2*$^{+}$ cells, we studied the kidney in more detail by looking at the integration of snRNA-seq and scATAC-seq data. Analysis of open chromatin regions revealed discrete peaks in the *ACE2* locus with the highest signal detected in a population of proximal tubule cells that also contains the highest proportion of *ACE2*-expressing cells (**Fig. 4d**). Motif analysis demonstrated that *ACE2* promoter and enhancer regions are enriched in *STAT1* and *3*, *FOXA1*, *JUNB* and several *IRF* (interferon response factor) binding sites (**Fig. 4e**). These transcription factors have important immune functions and are targets of tissue protective and innate immune responses such as those mediated by interleukin-6 (IL6), interleukin-1 (IL1) and interferons[65]. In this regard, dysregulation of both IL6 and IL1β has been implicated in the pathogenesis of severe COVID-19[66]. Thus, we investigated the co-expression of their receptors (*IL6R*, *IL1R1* and *IL1RAP*) with *ACE2* in monkey kidney, only observing good correlation with *ACE2* in proximal tubule cells for *IL6R* (**Extended Data Fig. 35b**). These observations imply a potential link between IL6, STAT transcription factors and enhanced *ACE2* expression in specific tissues such as the kidney that can either facilitate the existence of viral reservoirs or exacerbate COVID-19 disease progression due to increased viral dissemination (**Extended Data Fig. 35c)**. In addition to ACE2 and TMPRSS2, numerous other molecules have been implicated in facilitating SARS-CoV-2 binding to the cell surface or in COVID-19 pathogenesis[67,68]. Their expression or co-expression in monkey tissues, as well as other

447 potential associations and other virus-host interactions can be explored using our
448 NHPCA database.
449
450 **Investigation of common human traits and genetic diseases in monkey**
451 We next assessed the effect of genetic variation linked to complex human traits and
452 diseases by applying Genome Wide Association Studies (GWAS) to our monkey
453 dataset. We linked human single-nucleotide polymorphisms from 163 GWAS taken
454 from the UK Biobank to orthologous coordinates in the monkey single-cell
455 transcriptome to calculate the enrichment of traits across the genes expressed in each
456 cell cluster annotated in our dataset. As a general trend, we observed enriched
457 heritability for neurological traits such as 'schizophrenia', 'depression' or 'autism' in
458 clusters corresponding to neuronal and glial cells (**Fig. 5a, Extended Data Fig. 36 and**
459 **Supplementary Table 5a**). Similarly, we observed enrichment of Alzheimer's disease
460 traits in immune cells, in line with the knowledge that immune dysfunction contributes
461 to the pathogenesis of this disease[69]. Consistent with expectations, we also noticed
462 enrichment of immunological-related traits ('lymphocyte count', 'monocyte count' and
463 traits related to immune disorders) in myeloid cells and B and T lymphocytes. Likewise,
464 blood related traits such as 'mean sphered cell volume' and 'red blood cell distribution
465 width' were enriched in erythrocytes and bone marrow progenitor cells. Interestingly,
466 however, we observed some unexpected trends for traits like 'body mass index' or
467 'waste ratio'. Despite showing the expected highest enrichment in adipocytes, these
468 trends additionally revealed an enrichment in smooth muscle cells, melanocytes and
469 stromal cells. Similarly, type 2 diabetes and cholesterol-related traits revealed not only
470 the expected association with hepatocytes but also with several kidney cell
471 populations[70]. Our analysis also pointed at the enrichment of attention deficit and
472 hyperactive disorder (ADHD) in skeletal muscle type I and type II myonuclei but not
473 in neuronal cell types, suggesting an intriguing link between this pathology and motor
474 abnormalities (**Fig. 5a**). In this regard for example, ocular muscle hyperactivity is an

accompanying sign of ADHD and might be a major trigger for the disease rather than a consequence[71].

Besides the association of complex human traits to cell types stated above, we also generated a correlation map of mutant genes causing human genetic diseases with all cell types annotated in our monkey dataset (**Extended Data Fig. 37 and Supplementary Table 5b**). As expected, genes related to retinitis pigmentosa were specifically expressed in monkey photoreceptors, while genes related to porphyria were found associated to erythroblasts. This shows that our dataset can predict cell types that are directly affected in human genetic diseases. In addition, we compared the interspecies distribution of a panel of genes related to human neurological diseases using snRNA-seq data for mouse, monkey and human neocortex[46]. Notably, for most genes, we observed a generally higher correlation of the expression in specific cell types between human and monkey than between human and mouse (**Fig. 5b**). However, some diseases also appeared to be related to different cell types in monkey compared to human. For instance, distal neuropathy caused by mutations in *HSPB8*[72] was enriched in *CNR1*$^{+}$ inhibitory neurons in human while being enriched in astrocytes in monkey and mouse. Similarly, ataxia telangiectasia caused by mutations in *ATM* was mostly enriched in oligodendrocytes[73] in human while in monkey and mouse it was enriched in *PVALB*$^{+}$ and *LAMP5*$^{+}$ inhibitory neurons, respectively.

Our analysis thus highlights the potential for modelling human diseases in species phylogenetically closer to humans and underlines that differences will still exist. Further scrutiny of GWAS datasets and gene mutations and wider comparisons between species will provide additional relevant observations.

DISCUSSION

Despite the enormous potential, few NHP tissues have been profiled to date at the single-cell level and the use of different species, experimental conditions and platforms makes comparisons challenging[20,74,75]. To address this, we have generated the first

version of a large single-cell transcriptomic atlas for a NHP widely used in research studies, *Macaca fascicularis*, and an expandable and interactive database (https://db.cngb.org/nhpca/) to facilitate its exploration. The current version of our atlas provides a comprehensive and integrated overview of gene expression in 106 cell types extracted from 43 tissue types. Specialized tissues such as skin, thymus, testis and some parts of the gastrointestinal tract, as well as increased cell numbers for some of the already profiled ones, will be added in future releases. Cell type identification relied on previously reported markers and gene expression profiles. Therefore, although we identified most (if not all) known cell types in these tissues, our current annotations are likely to benefit from deeper sub-clustering and further revision.

We provide a detailed description of individual tissue single-cell composition and a comparison of common cell types across all sequenced tissues. This information will be particularly valuable for understanding tissues that have either not been profiled at all at the single-cell level in human (e.g., diaphragm, tongue and salivary gland) or lack enough cell numbers (e.g., liver, gallbladder and substantia nigra), and for prediction of human disease susceptibilities. Regarding the latter, we have identified an unexpected link between ADHD and muscle function. ADHD is a polygenic and multifactorial disorder associated with hyperactivity and motor coordination abnormalities that are thought to have a neurological origin[76]. Our data support the possibility that skeletal muscle rather than the nervous system may be a direct driver of ADHD pathogenesis[77]. Similarly, as part of the analysis for virus receptors and co-receptors, we provide a comprehensive map of *ACE2*$^{+}$/*TMPRSS2*$^{+}$ double positive cells throughout the monkey body that may be useful to understand COVID-19 pathogenesis in human[59,61]. In particular, the link between IL6, STAT transcription factors and ACE2 expression could explain the reported positive effects of tocilizumab, a humanized monoclonal antibody against IL6R for the treatment of patients with severe COVID-19[78]. On the other hand, our study shows significant interspecies differences in cell type-specific gene expression with potentially important functional consequences. For

example, the distribution of *ACE2* and *TMPRSS2* across different cell types is not identical between monkey and human and this could influence the disease course. Moreover, in the context of the survey of Wnt pathway components we have identified *LGR5*$^{+}$ renal cells with progenitor characteristics that are seemingly absent in human and mouse based on analysis of reported datasets. This is relevant because the kidney has limited regenerative capacity in mammals[79]. During embryonic development *LGR5*$^{+}$ cells located at the junction between the ureteric bud (source of the collecting tubule and connecting tubule) and the metanephric blastema are responsible for nephrogenesis, but they quickly disappear after birth[45]. Their persistence in adult monkey kidney suggests a higher regenerative capacity compared to other species, which if true raises the hope of activating a similar mechanism in human[80]. Similarly, *LGR5*$^{+}$ cells in the neocortex correspond mainly to OPC in monkey and to oligodendrocytes and to a lesser extent OPC in human, whereas in mouse inhibitory neurons are more highly enriched. This finding is consistent with the knowledge that Wnt activity regulates OPC and oligodendrocyte function and differentiation[81] but suggest interspecies differences in the mode of action. Likewise, the expression of *LGR5* in skeletal slow-twitch myofibers, and *LGR6* in the pituitary gland and heart, is intriguing. During development, Wnt activity regulates skeletal myogenesis and myofiber typing[82], cardiomyocyte proliferation[83] and pituitary gland growth[84], but little is known about the adult. The functional implications of these and other related findings and the extent to which the patterns differ between monkey and other mammalian species will require further study. Finally, interspecies comparison of single-cell gene expression in neocortex highlights the problems associated with modelling neurological diseases in rodents and suggests that a cautious approach should also be taken when studying NHP. Additional comparisons with other human and mouse single-cell/nuclei datasets will provide a more comprehensive, body-wide picture of differences in disease vulnerability among the three species.

In the future, with efforts from us and scientists worldwide, the NHPCA database will be extended with additional single-cell datasets generated from disease modelling studies, spontaneously developed diseases (e.g., diabetes or cardiomyopathy) and aging. Adding other layers of single-cell -omics studies, in particular scATAC-seq and spatially resolved transcriptomics[85] for all tissues presented here, will help characterize cell states and the interactions between different cell types more accurately. Proof of principle is the kidney scATAC-seq dataset included here. In addition, it will be important to compare our *Macaca fascicularis* atlas with datasets from other non-endangered NHP species such as *Macaca mulatta* (rhesus monkey), *Callithrix jacchus* (marmoset monkey)[86] and *Microcebus murinus* (mouse lemur)[10,13]. Altogether, this information will be instrumental for understanding primate evolution and human disease.

FIGURE LEGENDS

Figure 1. Generation of a single-cell atlas across 43 tissues of *Macaca fascicularis* monkey.

(a) Schematic representation of monkey tissues analyzed in this study (top left panel). A total of 43 tissues were collected from three male and three female 6-year-old monkeys. UMAP visualization of the global clustering indicating all single cells from the dataset colored by tissue (top middle panel) and bar plot showing the number of cells/nuclei profiled for every tissue after passing the quality control (top right panel). N = 1,084,164 individual nuclei/cells analyzed.

(b) UMAP visualization of tissues grouped by specific systems such as immune system (bone marrow, peripheral blood, spleen, tonsil and lymph node), digestive system (colon, duodenum, esophagus, gallbladder, liver, stomach and

585 tongue) and female reproductive system (fallopian tube, ovary, uterus and
586 vagina).
587 **(c)** UMAP visualization of all clusters colored by major cell types. A total of 106
588 cell clusters were identified in the dataset. Cell type annotation for all major
589 clusters is provided in the right-hand side legend. *SERPINE1* was used to
590 discriminate two distinct cluster of hepatocytes.
591
592 **Figure 2. Characterization of skeletal myofibers and mesothelial cells.**
593 **(a)** UMAP visualization of the global clustering of skeletal muscle cells annotated
594 in our dataset. Clusters are colored by tissue (abdominal wall, diaphragm,
595 esophagus and tongue). Due to their low number, fallopian tube, vagina and
596 tonsil skeletal cells were excluded from this analysis. Endothelial and immune
597 cells were not included in this analysis.
598 **(b)** UMAP representation of all re-clustered skeletal muscle cells colored by
599 subtype.
600 **(c)** UMAP visualization of specific markers used to identify type I (*MYH7*), type
601 IIa (*MYH2*) and type IIb myonuclei (*GPD2*), FAP (*LVRN*), MTJ (*NAV3* and
602 *COL22A1*), NMJ (*ETV5* and *MUSK*) and satellite cells (*PAX7*), as shown in **b**.
603 Due to their small proportions, the latter three populations are highlighted by a
604 red arrow.
605 **(d)** Stacked bar plot representing the proportion of skeletal muscle nuclei
606 (myonulcei subtypes type I, type IIa, type IIb, MTJ and NMJ, and also satellite
607 cells and FAP) in the indicated tissues.
608 **(e)** Heatmap showing DEG among the skeletal muscle populations highlighted in
609 **d**.
610 **(f)** Bubble plot showing DEG for each of the myonuclei subtypes comparing
611 different tissues.

(g) UMAP visualization of mesothelial cells from the selected tissues (bladder, diaphragm, fallopian tube, lung, ovary and visceral adipose tissue). Two different clusters of mesothelial cells in visceral adipose tissue are indicated by the red dotted line.

(h) Violin plot showing the differential expression of mesothelial and immune markers in the two visceral adipose tissue clusters highlighted by the red dotted line in panel **g**.

(i) UMAP visualization of three different clusters of mesothelial cells from the ovary (left panel). Mesothelial cells (Meso), surface epithelial (Surface epi) and progenitor-like epithelial (Prog-like epi) cells are highlighted in red, blue and yellow, respectively.

(j) UMAP visualization of *LGR5* expression in ovarian cells.

(k) Violin plot showing the DEG among the three populations of ovarian cells highlighted in the UMAP.

Figure 3. Analysis of *LGR5*$^{+}$ cells across all monkey tissues.

(a) UMAP visualization of *LGR5* expression across all tissues profiled in this study. The bubble plot on the right shows the *LGR5* expression ratio in the indicated cell types.

(b) Co-embedding of kidney snRNA-seq (highlighted in blue) and scATAC-seq (highlighted in red) datasets.

(c) UMAP visualization of integrated kidney snRNA- and scATAC-seq data. Cell clusters are colored according to cell identity. Abbreviations: DCTC, distal convoluted tubule cells; Endo, endothelial cells; LOH, loop of Henle; mDC, myeloid-derived dendritic cells; Myofibro, myofibroblasts.

(d) UMAP visualization of *LGR5* across kidney cell types and ArchR track visualization of aggregate scATAC-seq signals on the *LGR5* locus in each cell

type annotated in **c**. The bar plot on the right side indicates the ratio (%) of *LGR5*$^+$ cells in each cell type of kidney.

(e) Representative image of smFISH detection for *LGR5* expression in DCTCs (scale bar 20 µm). The bottom panel represents a magnification of the area indicated by the white box in the top panel.

(f) GO analysis showing the pathways associated to the DEGs obtained by comparing *LGR5*$^+$ cells from DCTC, ascending and descending LOH. The UMAP and the barplot on the right highlight the presence and the percentage of *LGR5*$^+$ cells co-expressing the progenitor markers *PAX2*, *TNFRSF19* and *LHX2*.

(g) UMAP visualization of all muscle cell types annotated in our dataset clustered by tissue (abdominal wall, aorta bladder, carotid, diaphragm, esophagus, fallopian tube, heart, ovary, prostate, spermaduct, tongue, uterus and vagina). The dotted lines group clusters of cells belonging to a specific muscle type (cardiac, skeletal and smooth muscle).

(h) UMAP visualization of *LGR5*, *MYH2* and *MYH7* across all skeletal muscle cell types. The blue dotted line in the left panel indicates all clusters belonging to the diaphragm while the one in the right panel indicates *LGR5*$^+$ cells.

(i) Representative image of smFISH detection for *LGR5*, *MYH7* and their co-expression in skeletal myonuclei of the diaphragm (scale bar 20 µm). The panel of the right is a magnification of the area indicated by the white box.

Figure 4. Global analysis of *ACE2* and *TMPRSS2* across monkey tissues.

(a) Heatmap showing the expression of entry receptors for a selection of the most common viruses (indicated on the left) in all cell clusters annotated in our dataset (indicated at the bottom).

(b) UMAP visualization of *ACE2* (top) and *TMPRSS2* (bottom) expression in all single cells from our dataset. The bubble plot next to each UMAP shows the expression levels of *ACE2* and *TMPRSS2* in the indicated cell types. The color

of each bubble represents the levels of expression and the size indicates the proportion of expressing cells.

(c) UMAP projection of *ACE2*$^{+}$/*TMPRSS2*$^{+}$ cells (highlighted in yellow). The bar plot on the right represents the ratio of cells that co-express both genes.

(d) UMAP visualization of *ACE2* in the integrated scATAC-seq and snRNA-seq from monkey kidney.

(e) ArchR track visualization of aggregate scATAC-seq signals on the *ACE2* locus in each on the annotated cell types of the kidney. Predicted binding of human transcription factor predicted based on DNA sequence is shown in the corresponding open chromatin regions of *ACE2*. The bar plot on the right indicates the ratio (%) of *ACE2*$^{+}$ cells in each annotated cell type of the monkey kidney.

Figure 5. Association of monkey transcriptomic profiles with human common traits and genetic diseases.

(a) Heatmap showing the association of selected common human traits and diseases (indicated on the right) with the cell types (indicated at the bottom) annotated in our dataset. The colored boxes indicate enriched specific patterns related to human traits/diseases subtypes.

(b) Heatmap showing the enrichment of genetic diseases related to the central nervous system in human, monkey and mouse neocortex snRNA-seq datasets. The black boxes indicated specific patterns associated with cell types annotated in the neocortex dataset.

Extended Data Figure 1. Quality control analysis of gender and individual effect.

UMAP visualization of single-cell profiles for selected tissues to calculate the batch effect between tissues from different individuals and genders. Two individuals were analyzed for bladder (F1 and F3), cerebellum (F3 and M1),

diaphragm (F1 and M2), gallbladder (F1 and F3), kidney (F1 and F2), lung (F1 and F3), salivary gland (F1 and F3), subcutaneous (F1 and M2) and visceral adipose (F1 and M2) tissues, and three for liver (F1, F2 and F3).

Extended Data Figure 2. Global clustering of different systems.

UMAP visualization of cell clusters in selected tissues grouped by system: cardiovascular (aorta, carotid and heart), endocrine (adrenal, pancreas, pineal, pituitary and thyroid glands), skeletal (abdominal wall and diaphragm), central nervous (cerebellum, neocortex, pigmentary epithelium choroid plexus, retina and spinal cord), respiratory (bronchus, lung and trachea) and urinary (bladder and kidney). Adipose tissues (subcutaneous and visceral) are also shown grouped. Clusters shown in every plot are colored by tissue. Abbreviation: pigmentary epi, pigmentary epithelium and choroid plexus.

Extended Data Figure 3. Global profiling of individual monkey tissues – 1.

UMAP projection of the global clustering indicating the distribution of all single cells (highlighted in yellow) from individual tissues for abdominal wall, adrenal gland, aorta, bladder, bone marrow, bronchus, carotid, cerebellum, colon, diaphragm, duodenum and epididymis.

Extended Data Figure 4. Global profiling of individual monkey tissues – 2.

UMAP projection of the global clustering indicating the distribution of all single cells (highlighted in yellow) from individual tissues for esophagus, fallopian tube, gallbladder, heart, kidney, liver, lung, lymph node, neocortex, ovary, pancreas and PBMC.

Extended Data Figure 5. Global profiling of individual monkey tissues – 3.

UMAP projection of the global clustering indicating the distribution of all single cells (highlighted in yellow) from individual tissues for pigmentary epithelium choroid plexus, pineal gland, pituitary gland, prostate, retina, salivary gland, spermaduct, spinal cord, spleen, stomach, subcutaneous adipose tissue and substantia nigra.

Extended Data Figure 6. Global profiling of individual monkey tissues – 4.

UMAP projection of the global clustering indicating the distribution of all single cells (highlighted in yellow) from individual tissues for thyroid, tongue, tonsil, trachea, uterus, vagina and visceral adipose tissue.

Extended Data Figure 7. Cluster annotations – 1.

UMAP visualization of cell clusters in the abdominal wall, adrenal gland, aorta, bladder, bone marrow, bronchus, carotid, cerebellum, colon, diaphragm, duodenum and epididymis. The name of the population in each cluster and the total number of cells profiled for every tissue are indicated in every plot. Abbreviations: Adipo, adipocytes; Astro, astrocytes; AT1, alveolar type 1 cells; AT2, alveolar type 2 cells; BC, B cells; CLP, common lymphoid progenitors; CMP, common myeloid progenitors; Endo, endothelial cells; Epi, epithelial cells; Ery, erythroblasts; FAP, fibroadipogenic progenitors; GMP, granulocyte monocyte progenitors; Macro, macrophages; mDC, myeloid derived dendritic cells; MEP, megakaryocyte erythrocyte progenitors; Meso, mesothelial cells; Mol interneu, molecular interneurons; Mono, monocytes; MTJ, myotendinous junction; Myofibro, myofibroblasts; NK, natural killers; NKT, natural killer T cells; NMJ, neuromuscular junction; Oligo, oligodendrocytes; OPC, oligodendrocyte progenitor cells; SMC, smooth muscle cells; TC, T cells.

Extended Data Figure 8. Cluster annotations – 2.

UMAP visualization of cell clusters in the esophagus, fallopian tube, gallbladder, heart, kidney, liver, lung, lymph node, neocortex, ovary, pancreas and PBMC. The name of the population in each cluster and the total number of cells profiled for every tissue are indicated in every plot. Abbreviations: Adipo, adipocytes; Astro, astrocytes; AT1, alveolar type 1 cells; AT2, alveolar type 2 cells; BC, B cells; Endo, endothelial cells; Epi, epithelial cells; EX, excitatory neurons; Hep, hepatocytes; IN, inhibitory neurons; lDC, lymphoid derived dendritic cells; LOH, loop of Henle cells; Lymph prog, lymphoid progenitors; Macro, macrophages; mDC, myeloid derived dendritic cells; Meg, megakaryocytes; Meso, mesothelial cells; Mono, monocytes; Myofibro, myofibroblasts; NK, natural killers; NKT, natural killer T cells; NMJ, neuromuscular junction; Oligo, oligodendrocytes; OPC, oligodendrocyte progenitor cells; Prog-like epi, progenitor-like epithelial cells; SMC, smooth muscle cells; TC, T cells.

Extended Data Figure 9. Cluster annotations – 3.

UMAP visualization of cell clusters in the pigmentary epithelium choroid plexus, pineal gland, pituitary gland, prostate, retina, salivary gland, spermaduct, spinal cord, spleen, stomach, subcutaneous adipose tissue and substantia nigra. The name of the population in each cluster and the total number of cells profiled for every tissue are indicated in every plot. Abbreviations: Adipo, adipocytes; Astro, astrocytes; BC, B cells; DAN, dopaminergic neurons; DC, conventional dendritic cells; Endo, endothelial cells; Epi, epithelial cells; EX, excitatory neurons; IN, inhibitory neurons; Macro, macrophages; Mono, monocytes; Myofibro, myofibroblasts; Neutro, neutrophils; NK, natural killers; NKT, natural killer T cells; Oligo, oligodendrocytes; OPC, oligodendrocyte progenitor cells; SMC, smooth muscle cells; TC, T cells.

Extended Data Figure 10. Cluster annotations – 4.

UMAP visualization of cell clusters in the thyroid, tongue, tonsil, trachea, uterus, vagina and visceral adipose tissue. The name of the population in each cluster and the total number of cells profiled for every tissue are indicated in every plot. Abbreviations: Adipo, adipocytes; BC, B cells; Endo, endothelial cells; lDC, lymphoid derived dendritic cells; LOH, loop of Henle; Macro, macrophages; mDC, myeloid derived dendritic cells; Meso, mesothelial cells; Mono, monocytes; NK, natural killers; NMJ, neuromuscular junction; SMC, smooth muscle cells; TC, T cells.

Extended Data Figure 11. Selected markers for cell cluster annotations.

Heatmap showing the expression of the marker genes used to manually annotate all cell clusters identified in every tissue of this dataset.

Extended Data Figure 12. UMI and gene numbers of the sequenced tissues and annotated cell types.

(a) Boxplot indicating the number of UMI (top) and genes (bottom) in each tissue of the dataset.

(b) Boxplot indicating the number of UMI (top) and genes (bottom) detected in each of the major annotated cell types shown in **Figure 1c**.

Extended Data Figure 13. Cell numbers and proportions among the sequenced tissues.

Bar plot representation of the number of cells analyzed for each cell type described in main **Figure 1c**. The stacked bar plot at the bottom indicates the ratio of each cell type detected in every tissue.

Extended Data Figure 14. Unique and shared cell populations.

(a) UMAP projection of the global clustering showing the expression of specific markers for cerebellum granule cells (*GABRA6*), hepatocytes (*ALB*), salivary gland acinar cells (*PRR27*), epididymis stereociliated cells (*ROS1*), pancreatic alpha cells (*GCG*) and fasciculata cells of the adrenal gland (*CYP11A1*).

(b) UMAP projection of the global clustering showing the expression of pan-markers of endothelial (*FLT1*), stromal (*DCN*), immune (*PTPRC*), skeletal myonuclei (*TTN*), adipocytes (*ADIPOQ*) and mesothelial cells (*ITLN1*) that are shared across tissues.

Extended Data Figure 15. Global analysis of common cell types.

UMAP visualization of (**a**) stromal cells (n = 35,415), **(b)** macrophages (n = 10,929), **(c)** endothelial cells (n = 37,640) and **(d)** smooth muscle cells (n = 24,175) from all analyzed monkey tissues. Tissues with low numbers of the selected cell types were excluded. Cell clusters are colored by tissue. The heatmap on the right shows tissue-specific DEG for **(e)** stromal cells, **(f)** macrophages, **(g)** endothelial cells and **(h)** smooth muscle cells.

Extended Data Figure 16. Analysis of skeletal myonuclei molecular signatures.

(a) Bubble plot indicating tissue-specific enriched GO terms in type I myonuclei from abdominal wall, diaphragm and tongue.

(b) Bubble plot indicating tissue-specific enriched GO terms in type IIa myonuclei from diaphragm, esophagus and tongue.

(c) Bubble plot indicating tissue-specific enriched GO terms in type IIb myonuclei from abdominal wall and diaphragm.

Extended Data Figure 17. Global analysis of adipocyte populations.

(a) UMAP visualization of mature adipocyte and adipocyte progenitors from visceral (VAT) and subcutaneous (SAT) adipose tissues. Data were grouped

together and re-clustered either by tissue type (on the left) or by cell type (on the right).

(b) UMAP visualization of specific markers for mature adipocytes (*ADIPOQ*) or adipocyte progenitors (*CD34*).

(c) UMAP visualization of markers for tissue-specific (*ITLN1* and *FOS*), cell-type specific (*SLC11A1*, *SPOCK3*, *WT1, ESR1, CXCL14, APOD, CFD, DPP4* and *NOX4*) or cycling markers (*MKI67*).

(d) Heatmap indicating the DEG in all clusters identified in **a**.

Extended Data Figure 18. Global analysis of *LGR4, LGR6* and *LGR5/LGR6* co-expression across monkey tissues.

(a) UMAP visualization of *LGR6* across all tissues profiled in this study. The bubble plot on the right shows the *LGR6* expression ratio in the indicated cell types.

(b) UMAP visualization of *LGR5* and *LGR6* co-expression across all tissues profiled in this study. The barplot on the right shows the co-expression ratio in the indicated cell types.

(c) UMAP visualization of *LGR4* across all tissues profiled in this study. The bubble plot on the right shows the *LGR4* expression ratio in the indicated cell types.

Extended Data Figure 19. Global analysis of *LGR5* and *LGR6* across monkey tissues – 1.

Bubble plot (left) showing the ratio of $LGR5^{+}$, $LGR6^{+}$ and $MKI67^{+}$ cells in the annotated cell types for each tissue and UMAP visualization (right) of *LGR5*, *LGR6* and *MKI67* in abdominal wall, adrenal gland, aorta, bladder, bone marrow, bronchus, carotid, cerebellum, colon, diaphragm and duodenum.

Extended Data Figure 20. Global analysis of *LGR5* and *LGR6* across monkey tissues – 2.

Bubble plot (left) showing the ratio of *LGR5*$^+$, *LGR6*$^+$ and *MKI67*$^+$ cells in the annotated cell types for each tissue and UMAP visualization (right) of *LGR5*, *LGR6* and *MKI67* in epididymis, esophagus, fallopian tube, gallbladder, heart, kidney, liver, lung, lymph node and ovary.

Extended Data Figure 21. Global analysis of *LGR5* and *LGR6* across monkey tissues – 3.

Bubble plot (left) showing the ratio of *LGR5*$^+$, *LGR6*$^+$ and *MKI67*$^+$ cells in the annotated cell types for each tissue and UMAP visualization (right) of *LGR5*, *LGR6* and *MKI67* in pancreas, PBMCs, pigmentary epithelium choroid plexus (indicated as pigmentary epi), pineal gland, pituitary gland, prostate, retina, salivary gland, spermaduct, spinal cord and spleen.

Extended Data Figure 22. Global analysis of *LGR5* and *LGR6* across monkey tissues – 4.

Bubble plot (left) showing the ratio of *LGR5*$^+$, *LGR6*$^+$ and *MKI67*$^+$ cells in the annotated cell types for each tissue and UMAP visualization (right) of *LGR5*, *LGR6* and *MKI67* in stomach, subcutaneous adipose tissue, substantia nigra, thyroid, tongue, tonsil, trachea, uterus, vagina and visceral adipose tissue.

Extended Data Figure 23. Kidney snRNA-seq and scATAC-seq dataset integration.

(a) Violin plot showing the expression of selected markers used to annotate the kidney cell clusters from snRNA-seq data.

(b) ArchR track visualization of aggregate scATAC-seq signals on the locus of the selected marker genes indicated in **a**. Abbreviations: DCTC, distal convoluted

tubule cells; Endo, endothelial cells; LOH, loop of Henle; Myofibro, myofibroblasts.

Extended Data Figure 24. *LGR5* analysis in integrated human, monkey and mouse kidney data.

(a) UMAP visualization of cell clusters in human (left), monkey (middle) and mouse (right) kidney snRNA-seq datasets. The annotation of each cluster in provided in the legend at the bottom. Abbreviations: Endo, endothelial cells; LOH, loop of Henle; mDC, myeloid dendritic cells; Myofibro, myofibroblasts.

(b) UMAP visualization of *LGR5* in human (left), monkey (middle) and mouse (right) kidney.

(c) Bubble plot showing the ratio and expression levels of *LGR5* and DCTC marker *SLC12A3* in human, monkey and mouse kidney datasets. The color of each bubble represents the level of expression and the size indicates the proportion of expressing cells.

Extended Data Figure 25. *LGR5* analysis in integrated human, monkey and mouse neocortex data.

(a) UMAP visualization of cell clusters in human (left), monkey (middle) and mouse (right) neocortex snRNA-seq datasets. The annotation of each cluster is provided in the legend at the bottom. Abbreviations: Astro, astrocytes; Endo, endothelial cells; IN, inhibitory neurons; OPC, oligodendrocyte progenitor cells; EX, excitatory neurons; Oligo, oligodendrocytes.

(b) UMAP visualization of *LGR5* in human (left), monkey (middle) and mouse (right) neocortex. OPC and oligodendrocytes are indicated by a red and yellow dotted circle, respectively.

(c) Bubble plot showing the ratio and expression levels of *LGR5* and *PDGFRA* in human, monkey and mouse neocortex. The color of each bubble represents the level of expression and the size indicates the proportion of expressing cells.

(d) Monocle 2 pseudotime-ordered trajectory of OPC (labelled in orange) maturation towards mature oligodendrocytes (labelled in blue).

(e) Monocle 2 pseudotime analysis showing the expression of OPC markers (*LGR5*, *OLIG2* and *PDGFRA*) and the oligodendrocytes marker *PLP1*.

(f) Representative image of immunofluorescence staining for PDGFRA (red) and LGR5 (green), respectively, and their co-expression in OPC of monkey neocortex (scale bar 20 μm). The smaller panel at the bottom is a magnification of the area indicated by the green box.

Extended Data Figure 26. Analysis of *LGR6* expression in monkey heart and pituitary gland.

(a) UMAP visualization of *LGR6* across all muscle cell types annotated in our dataset, as displayed in **Figure 3g**. The dotted red line indicates a cluster of muscle cells belonging to the heart. The red arrows indicate *LGR6*$^{+}$ cells in aorta, carotid, ovary and vagina.

(b) Representative image of smFISH detection for *LGR6* in heart myonuclei (scale bar 40 μm). The bottom right panel is a magnification of the area indicated by the white box.

(c) UMAP visualization of cell clusters in human (left), monkey (middle) and mouse (right) heart snRNA-seq datasets. The annotation of each cluster is provided in the legend at the bottom. Abbreviations: Endo, endothelial cells; Macro, macrophages; SMC, smooth muscle cells.

(d) UMAP visualization of *LGR6* in human (left), monkey (middle) and mouse (right) heart.

941 **(e)** UMAP visualization of *LGR6* expression in pituitary gland highlighting the
942 highest expression in folliculostellate cells.
943 **(f)** UMAP visualization of *SOX2*, *PAX6*, *CD44* and *CXCR4* in folliculostellate cells
944 as indicated by the black box.
945 **(g)** Barplot showing GO terms associated to the DEGs in folliculostellate cells of
946 pituitary gland.
947
948 **Extended Data Figure 27. Analysis of *WNT9B* and Wnt pathway gene module in**
949 **monkey kidney.**
950 **(a)** Heatmap showing the expression of all receptors and ligands of the Wnt
951 pathway in the annotated cell populations of the kidney.
952 **(b)** Network plots showing cell-cell communications based on ligand-receptor
953 interactions calculated by CellphoneDB.
954 **(c)** UMAP visualization of *WNT9B* expression in monkey kidney.
955 **(d)** ArchR track visualization of aggregate scATAC-seq signals on the *WNT9B*
956 locus in each on the annotated cell types. The bar plot at the bottom indicates
957 the ratio (%) of *WNT9B*$^{+}$ cells in each cell type of kidney.
958 **(e)** Bubble plot showing the ratio and expression levels of *WNT9B* and principal
959 tubule cell marker *FXYD4* in human, monkey and mouse kidney datasets. The
960 color of each bubble represents the level of expression and the size indicates the
961 proportion of expressing cells.
962 **(f)** Schematic representation of a kidney nephron illustrating Wnt pathway ligand-
963 receptor interactions.
964
965 **Extended Data Figure 28. Global analysis of the Wnt pathway gene module in**
966 **monkey diaphragm and epididymis.**
967 **(a)** Heatmap showing the expression of all receptors and ligands of the Wnt
968 pathway in the annotated cell populations of the diaphragm.

(b) Network plots showing cell-cell communication based on ligand-receptor interactions calculated by CellphoneDB in the diaphragm dataset. Abbreviations: Adipo, adipocytes; Endo, endothelial cells; Macro, macrophages; Meso, mesothelial cells; NMJ, neuromuscular junctions.

(c) Heatmap showing the expression of all receptors and ligands of the Wnt pathway in the annotated cell populations of the epididymis.

(d) Network plots showing cell-cell communication based on ligand-receptor interactions calculated by CellphoneDB in the epididymis dataset. Abbreviations: Adipo, adipocytes; Endo, endothelial cells; SMC, smooth muscle cells; TC, T cells.

Extended Data Figure 29. Global analysis of the Wnt pathway gene module in monkey fallopian tube and liver.

(a) Heatmap showing the expression of all receptors and ligands of the Wnt pathway in the annotated cell populations of the fallopian tube.

(b) Network plots showing cell-cell communication based on ligand-receptor interactions calculated by CellphoneDB in the fallopian tube dataset. Abbreviations: Endo, endothelial cells; epi, epithelial cells; Meso, mesothelial cells; Mono, monocytes; SMC, smooth muscle cells; TC, T cells.

(c) Heatmap showing the expression of all receptors and ligands of the Wnt pathway in the annotated cell populations of the liver.

(d) Network plots showing cell-cell communication based on ligand-receptor interactions calculated by CellphoneDB in the liver dataset. Abbreviations: BC, B cells; Endo, endothelial cells; hep, hepatocytes; mDC, myeloid derived dendritic cells; Mono, monocytes; TC, T cells.

Extended Data Figure 30. Global analysis of the Wnt pathway gene module in monkey neocortex and ovary.

(a) Heatmap showing the expression of all receptors and ligands of the Wnt pathway in the annotated cell populations of the neocortex.

(b) Network plots showing cell-cell communication based on ligand-receptor interactions calculated by CellphoneDB in the neocortex dataset. Abbreviations: Astro, astrocytes; Endo, endothelial cells; EX, excitatory neurons; IN, inhibitory neurons; Oligo, oligodendrocytes; OPC, oligodendrocyte progenitor cells.

(c) Heatmap showing the expression of all receptors and ligands of the Wnt pathway in the annotated cell populations of the ovary.

(d) Network plots showing cell-cell communication based on ligand-receptor interactions calculated by CellphoneDB in the ovary dataset. Abbreviations: Endo, endothelial cells; epi, epithelial cells; Meso, mesothelial cells; Mono, monocytes; Myofibro, myofibroblasts; Prog-like epi; progenitor-like epithelial cells; SMC, smooth muscle cells.

Extended Data Figure 31. Global analysis of the Wnt pathway gene module in other monkey heart and pituitary gland.

(a) Heatmap showing the expression of all receptors and ligands of the Wnt pathway in the annotated cell populations of the heart.

(b) Network plots showing cell-cell communication based on ligand-receptor interactions calculated by CellphoneDB in the pituitary gland dataset. Abbreviations: Endo, endothelial cells; Myofibro, myofibroblasts.

(c) Heatmap showing the expression of all receptors and ligands of the Wnt pathway in the annotated cell populations of the pituitary gland.

(d) Network plots showing cell-cell communication based on ligand-receptor interactions calculated by CellphoneDB in the pituitary gland dataset. Abbreviations: Endo, endothelial cells.

Extended Data Figure 32. Global analysis of virus entry receptors across monkey tissues.

Heatmap showing the expression of entry receptor for most common viruses (shown on the right) in the indicated cell types (shown at the bottom).

Extended Data Figure 33. Analysis of *ACE2* and *TMPRSS2* expression across monkey tissues – 1.

UMAP visualization of *ACE2* (left), *TMPRSS2* (middle) and $ACE2^{+}/TMPRSS2^{+}$ (right) in abdominal wall, adrenal gland, aorta, bladder, bone marrow, bronchus, carotid, cerebellum, colon, diaphragm, duodenum, epididymis, esophagus, fallopian tube, gallbladder, heart, kidney, liver, lung, lymph node and ovary.

Extended Data Figure 34. Analysis of *ACE2* and *TMPRSS2* expression across monkey tissues – 2.

UMAP visualization of *ACE2* (left), *TMPRSS2* (middle) and $ACE2^{+}/TMPRSS2^{+}$ (right) in pancreas, PBMC, pigmentary epithelium choroid plexus (indicated as pigmentary epi), pineal gland, pituitary gland, prostate, retina, salivary gland, spermaduct, spinal cord, spleen, stomach, subcutaneous adipose tissue, substantia nigra, thyroid, tongue, tonsil, trachea, uterus, vagina and visceral adipose tissue.

Extended Data Figure 35. Comparative analysis of *ACE2* and *TMPRSS2* expression in monkey and human.

(a) Bubble plot showing the ratio and expression levels of *ACE2* and *TMPRSS2* in gallbladder, kidney, liver and lung in monkey and human. The color of each bubble represents the level of expression and the size indicates the proportion of expressing cells.

(b) UMAP visualization of *IL6R*, *IL1R1* and *IL1RAP* expressing in monkey kidney (top). The UMAP in the bottom represent the co-expression of *ACE2* and *IL6R*, *IL1R1* and *IL1RAP* in monkey kidney. Double positive cells are indicated in yellow.

(c) Schematic diagram of the potential mechanism for SARS-CoV-2 spreading through gallbladder, kidney, liver and lung. Kidney proximal tubule cells within the nephron are among the highest ACE2 expressing cells. After virus contact, IL6R stimulates an immune response that, through the activation of STAT transcription factors, potentiates a paracrine positive feedback loop that enhances ACE2 expression and facilitates virus spreading. IL6 expression, which is higher in elderly patients and those with inflammatory conditions, is effectively targeted by anti-IL6R monoclonal antibodies leading to a more favourable disease course.

Extended Data Figure 36. Expression of genes associated with human common traits in monkey cell types.

Heatmap showing the association of common human traits and diseases from the UK Biobank (indicated on the right) with the cell types (indicated at the bottom) annotated in our dataset.

Extended Data Figure 37. Association of monkey cell type-specific transcriptomic profiles with human genetic diseases.

Heatmap showing the association of human genetic diseases (indicated on the right) with the cell types (indicated at the bottom) annotated in our dataset.

Supplementary Table 1. Description of all tissues profiled, cell types and markers used for cluster annotation

Supplementary Table 2. Global analysis of common cell types and tissue-specific signatures

Supplementary Table 3. Global distribution of LGR5, LGR6 and MKI67 expression

Supplementary Table 4. Analysis of the expression of common virus and SARS-Cov-2 receptors

Supplementary Table 5. Correlation of GWAS traits and human genetic diseases with monkey cell types

METHODS

Ethics statement

This study was approved by the Institutional Review Board on Ethics Committee of BGI (permit no. BGI-IRB19125).

Collection of monkey tissues

A total of three females and three males, approximately 6-year-old, cynomolgus monkeys were obtained from Huazhen Laboratory Animal Breeding Centre and Hubei Topgene Biotechnology (Guangzhou, China). Animals were anesthetized with ketamine hydrochloride (10 mg/kg) and sodium pantabarbital (40 mg/kg) injection before being euthanized by exsanguination. Tissues were isolated and placed on the ice-cold board for dissection. A total of 43 whole tissues were isolated: abdominal wall, adrenal gland, aorta and carotid arteries, bladder, bone marrow, bronchia, cerebellum, colon, diaphragm, duodenum, epididymis, esophagus, fallopian tube, gallbladder, heart,

kidney, liver, lung, lymph node, neocortex, ovary, pancreas, PBMC, pigmentary epithelium choroid plexus, pineal gland, pituitary gland, prostate, retina, salivary gland, spermaduct, spinal cord, spleen, stomach, subcutaneous adipose tissue, substantia nigra, thyroid gland, tongue, tonsil, trachea, uterus, vagina and visceral adipose tissue. Each tissue (except for bone marrow, peripheral blood and tissues on which enzymatic digestion was performed) was cut into 5-10 pieces of roughly 50-200 mg each. Samples were transferred to cryogenic vials (Corning, #430488), then quickly frozen in liquid nitrogen and finally stored until nuclear extraction was performed. PBMC and bone marrow cells were isolated from heparinized venous blood using a Lymphoprep™ medium (STEMCELL Technologies, #07851) according to standard density gradient centrifugation methods. Cells from those two tissues were resuspended in 90% FBS, 10% DMSO (Sigma Aldrich, #D2650) freezing media and frozen using a Nalgene® Mr. Frosty® Cryo 1°C Freezing Container (Thermo Fisher Scientific, #5100-0001) in a -80°C freezer for 24 hours before being transferred to liquid nitrogen for long-term storage.

Single-nucleus/cell suspension preparation

Single nucleus isolation was performed as described previously[87]. Briefly, tissues were thawed, minced and transferred to a 1 ml Dounce homogenizer (TIANDZ) with 1 ml of homogenization buffer A containing 250 mM sucrose (Ambion), 10 mg/ml BSA (Ambion), 5mM $MgCl_2$ (Ambion), 0.12 U/µl RNasin Plus (Promega, #N2115), 0.12 U/µl RNasein (Promega, #N2115) and 1x Protease Inhibitor (Roche, #11697498001). Tissues were kept in an ice box and homogenized by 25-50 strokes of the loose pestle (Pestle A) after which the mixture was filtered using a 100 µm cell strainer in to a 1.5 ml tube (Eppendorf). The mixture was then transferred to a clean 1 ml dounce homogenizer to which 750 ul of buffer A containing 1% Igepal (Sigma, #CA630) was added and the tissue was further homogenized by 25 strokes of the tight pestle (Pestle B). After this, the mixture was filtered through a 40 µm strainer in a 1.5 ml tube and

centrifuged at 500 g for five minutes at 4°C to pellet nuclei. At this stage, the pellet was resuspended in 1 ml of buffer B containing 320 mM Sucrose, 10 mg/ml BSA, 3 mM $CaCl_2$, 2 mM $MgAc_2$, 0.1 mM EDTA, 10 mM Tris-HCl, 1 mM DTT, 1x Protease Inhibitor and 0.12 U/µl RNasein. This was followed by a centrifugation at 500 g for five minutes at 4°C to pellet nuclei. Nuclei were then resuspended with cell resuspension buffer at a concentration of 1,000 nuclei/µl for single-nucleus library preparation. Cells from lymph node, spleen, duodenum, stomach and colon were obtained from fresh tissues by enzymatic digestion. Briefly, tissues were rinsed in PBS, minced into small pieces by mechanical dissociation and incubated for 1 hour in 10 ml of DS-LT buffer (0.2 mg/ml $CaCl_2$, 5 µM $MgCl_2$, 0.2% BSA and 0.2 mg/ml Liberase in HBSS) at 37°C. After this, the tissue digestion was stopped by addition of 3 ml of FBS, followed by filtration through a 100 µm cell strainer and centrifugation for 5 minutes at 500 g at 4°C. Samples were then filtered through a 40 µm cell strainer and centrifuged for five minutes at 500 g at 4°C. Pellets were then resuspended in cell resuspension buffer at 1,000 cells/µl for single-cell library preparation.

Single-cell/single-nucleus RNA-seq (sc/snRNA-seq)

DNBelab C Series Single-Cell Library Prep Set was utilized as previously described[14]. In brief, single-nucleus/cell suspensions were used for droplet generation, emulsion breakage, beads collection, reverse transcription and cDNA amplification to generate barcoded libraries. Indexed sc/snRNA-seq libraries were constructed according to the manufacturer's protocol. The concentration of sc/snRNA-seq sequencing libraries was quantified by Qubit™ ssDNA Assay Kit (Thermo Fisher Scientific, #Q10212). The resulting libraries were sequenced using a DIPSEQ T1 or DIPSEQ T7 sequencers at the China National GeneBank (Shenzhen, China).

Single-cell ATAC-seq (scATAC-seq)

ScATAC-seq libraries were prepared using DNBelab C Series Single-Cell ATAC Library Prep Set[14]. DNA nanoballs were loaded into the patterned Nano arrays and sequenced on a BGISEQ-500 sequencer using the following read length: 50 bp for read 1, 76 bp for read 2, inclusive of 50 bp insert DNA, 10 bp cell barcode 1, 6 bp constant sequence and 10 bp cell barcode 2.

Immunofluorescence

Staining of monkey neocortex sample was conducted following standard protocol[88]. In brief, paraffin embedded sections were deparaffinized, incubated with primary antibodies for PDGFRα (Cell Signaling #3174S) and LGR5 (Abcam #ab273092) overnight at 4℃, followed by an incubation with a secondary antibody (Alexa Fluor 488 and Cy3, Jackson ImmunoResearch) for 30 minutes at room temperature. Slides were mounted with Slowfade Mountant+DAPI (Life Technologies, #S36964) and sealed.

Single-molecule fluorescence *in situ* hybridization (smFISH)

SmFISH in monkey kidney, diaphragm and heart tissues was performed using RNAScope Fluorescent Multiplex and RNAScope Multiplex Fluorescent v2 (Advanced Cell Diagnostics) according to manufacturer's instructions. The following alterations were added: the thickness of paraffin section was adjusted to 5 μm and target retrieval boiling time was adjusted to 15 minutes while the incubation time of Protease plus at 40℃ was adjusted to 30 minutes. RNA smFISH probes used: *LGR5* (C1), *LGR6* (C2), *MYH7* (C2).

Sc/snRNA-seq data processing

Raw sequencing reads from DIPSEQ-T1 or DIPSEQ-T7 were filtered and demultiplexed using PISA (version 0.2) (https://github.com/shiquan/PISA). Reads were aligned to Macaca_fascicularis_5.0 genome using STAR (version 2.7.4a)[89] and

sorted by sambamba (version 0.7.0)[90]. For tissues sequenced with scRNA-seq, reads were aligned to the exon of mRNA as normal. For tissues sequenced with snRNA-seq, a custom 'pre-mRNA' reference was created for alignment of count reads to introns as well as to exons because of large amount of unspliced pre-mRNA and mature mRNA in the cell nucleus. Thus, each gene's transcripts in snRNA-seq was counted out by including exon and intron reads together[91]. In the end, cell/nucleus versus gene UMI count matrix was generated with PISA.

Doublet removal

For each library, we performed doublet removal using DoubletFinder[92]. DoubletFinder first averages the transcriptional profile of randomly chosen cell pairs to create pseudo doublets and then predicts doublets according to each real cell's similarity in gene expression to the pseudo doublets. The doublet removal was performed according to the default parameter of DoubletFinder and the top 5% of cells most similar to the "pseudo doublets" were excluded.

Cell clustering and identification of cell types

Clustering analysis of the complete cynomolgus monkey tissue dataset was performed using Scanpy (version 1.6.0)[93] in a Python environment (version 3.6). Parameters used in each function were manually curated to portray the optimal clustering of cells. In the preprocessing, cells or nuclei were filtered based on the criteria of expressing a minimum of 500 genes and genes expressed by at least three cells or nuclei were kept for the following analysis. In addition, cells or nuclei with more than 10% mitochondrial gene counts were removed. Filtered data were ln (counts per million (CPM)/100 + 1) transformed. 3,000 highly variable genes were selected according to their average expression and dispersion. The number of UMI and the percentage of mitochondrial genes were regressed out and each gene was scaled by default options. Dimension reduction starts with principal component analysis and the number of

principal components used for UMAP depended on the importance of embeddings. Louvain method is then used to detect subgroups of cells. Distinguishing differential genes among clusters were ranked (Benjamini-Hochberg, Wilcoxon rank-sum test). Cell types were manually and iteratively assigned based on overlap of literature, curated and statistically ranked genes. Each tissue dataset was portrayed using the Seurat package (version 3.2.2)[94] in R environment (version 3.6). Data from different replicates were integrated following the standard integrated pipeline by default parameters for filtering, data normalization, dimensionality reduction, clustering and gene differential expression analysis. Finally, we annotated each cell type by extensive literature reading and searching for the specific gene expression patterns.

Differentially expressed gene (DEG) and gene ontology (GO) term enrichment analysis

In the global clustering, we performed DEG analysis using the sc.pl.rank_genes_groups function in Scanpy (V1.6.0). In other studies, we used the FindMarker or FindAllMarker function in the Seurat R package (V3.2.2). Analysis of DEG comparing specific populations was performed by calculating the fold-change of the mean expression level of genes between the selected populations. DEG were defined as those with a fold-change > 2 and adjusted $P < 0.01$. GO enrichment analysis was performed using the CompareCluster function fun = "enrichGO", pvalueCutoff = 0.1, pAdjustMethod = "BH", OrgDb = org.Hs.eg.db,ontBP") of ChIPseeker R package (v.1.22.1)[95]. Only GO terms with adjusted $P < 0.05$ were retained.

Analysis of inter-species differences

For tissue inter-species analysis, in order to get more accurate comparisons, we specifically chose three tissues with snRNA-seq data, namely kidney, neocortex and heart, and processed the raw sequencing data using our pipeline described below in the 'Sc/snRNA-seq data processing' section. Kidney[43,44], neocortex[46] and heart[49,50] data

were downloaded from NCBI Gene expression omnibus (human kidney: GSE121862, mouse kidney: GSE119531, human neocortex: GSE97942, human heart: ERP123138, mouse heart: E-MTAB-7869). For each tissue we preprocessed the UMI matrix of the three species following three steps: 1. only orthologs genes among three species were kept. 2. only genes expressed in at least one cell in one species were kept. 3. the gene names of the human and mouse UMI matrix were converted into orthologs in *Macaca fascicularis*. After preprocessing, the UMI matrices of the three species were integrated together and the clustering was performed following the standard integrated pipeline using Seurat (V3.2.2) with the addition of one additional criterion for which only cells expressing more than 500 genes were kept. Also, we downsampled the cells of human and macaque neocortex to 10,000 to get a better clustering result. The Seurat clusters were then annotated into different cell types using cell type-specific markers as described above. In addition, for the comparison presented in Extended Data Figure 35 we retrieved the publicly available single-cell data for gallbladder, liver and lung from GEO GSE134355[3], GEO GSE108098[6] and GSE124395[96], respectively. Data from the three species were integrated, clustered and annotated in the same way as described.

Common cell analysis

We performed common cell analysis for 7 cell types across all the 43 tissues, those being stromal cells, macrophages/microglia, endothelial cells, smooth muscle cells, skeletal muscle cells, mesothelial cells and adipocytes. For each cell type, we extracted those cells from all tissues in our dataset according to the cell type annotation presented in Extended Data Figure 7-10. For the downstream analysis, we excluded cell types with numbers lower than 200. Data from different replicates were integrated following the standard integrated pipeline using Seurat (V3.2.2).

Single-cell trajectory analysis

Cell lineage trajectory was inferred using Monocle2[97] following the tutorial. We used the "differentialGeneTest" function to derive DEG from each cluster and genes with $q < 0.01$ were used to order the cells in a pseudotime analysis. After the cell trajectories was constructed, DDRtree was used to visualize it in a two-dimensional space.

Cell-cell interaction network

To assess the cellular crosstalk between different cell types in each tissue, we used CellPhoneDB, a public repository of ligand-receptor interactions[98]. Cell type-specific receptor-ligand interactions between cell types were identified based on specific expression of a receptor by one cell type and a ligand by another cell type. The interaction score refers to the mean total of all individual ligand-receptor partner average expression values in the corresponding interacting pairs of cell types. For this analysis, we applied a statistical method to ensure that only receptors or ligands expressed in more than 10% of the cells in the given cluster were considered. The total mean of the individual partner average expression values in the corresponding interacting pairs of cell types was calculated. For the cell-cell interaction analysis in Extended Data Figure 27-31, we plot the figure based on the indicated genes related to *LGR5* and *LGR6*.

Association of GWAS summary data of human diseases and traits with monkey cell types

To test for the enrichment of human diseases and traits in DEG for each cluster of cells based on global clustering, we applied LD (linkage disequilibrium) score regression analysis. For this, we only considered genes with an adjusted $P < 0.05$ and fold-change > 2 in the tested cell types. For accuracy, cell types identified in a number lower than 100 were excluded from this analysis. We converted the gene coordinates of *Macaca fascicularis* into hg19 genome coordinates by downloading from Ensembl the

1302 homologous gene list. Single nucleotide polymorphisms located in gene regions of the
1303 most specific genes in each cell type were added to the baseline model independently
1304 for each cell type (one file for each cell type). We then selected the coefficient *z*-score
1305 *P* value as a measure of the association of the cell type with the traits. All plots show
1306 the $-\log_{10} P$ value of partitioned LDscore regression.
1307
1308 **ScATAC-seq data processing, clustering and cell type identification**
1309 Raw sequencing reads from BGISEQ-500 were filtered and demultiplexed using PISA
1310 (version 0.2) (https://github.com/shiquan/PISA). The fragment file of each scATAC-
1311 seq library was used for downstream analysis. TSS (transcription start site) enrichment
1312 score and fragment number of each nuclei was calculated by using ArchR software[99].
1313 Nuclei with TSS enrichment score lower than five and fragment number lower than
1314 1,000 were removed. Then, we calculated the doublet score with *addDoubletScores*
1315 function in ArchR package and filtered doublets by *filterDoublets* function with
1316 parameter filterRatio = 2. ScATAC–seq clustering analysis was performed using ArchR
1317 software by first identifying a robust set of peak regions followed by iterative LSI
1318 (latent semantic indexing) clustering. Briefly, we created 500 bp windows tiled across
1319 the genome and determined whether each cell was accessible within each window. Next,
1320 we performed an LSI dimensionality reduction on these windows with *addIterativeLSI*
1321 function in ArchR packages. We then performed Seurat clustering (*FindClusters*) on
1322 the LSI dimensions at resolutions of 0.8. Anchors between scATAC-seq and
1323 sc/snRNA-seq datasets were identified and used to transfer cell type labels identified
1324 from the sc/snRNA-seq data. We embedded the data by the TransferData function of
1325 Seurat (version 3.2.2).
1326
1327 **Transcription factor motif enrichment analysis**
1328 To predict the motif footprint in peaks within the *ACE2* promoter and enhancer
1329 sequences, we extracted genome sequences in the peak region with Seqkit (version

0.7.0)[100]. The sequences were imported into R and were matched with all *Homo sapiens* motifs form JASPAR2018 using matchMotifs function in motifmatchr packages version 1.8.0 with default parameter.

Data availability

All raw data have been deposited to CNGB Nucleotide Sequence Archive (accession code: CNP0001469; https://db.cngb.org/cnsa/project/CNP0001469/reviewlink/).

ACKNOWLEDGMENTS

We would like to thank Wei Liu and Liangzhi Xu from Huazhen Laboratory Animal Breeding Centre for helping in the collection of monkey tissues, Dahai Zhu and Hu Li from Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory) for technical help, Guoji Guo and Huiyu Sun from Zhejiang University for providing analytical advice, Guoyi Dong and Chao Liu from BGI research, Xiao Zhang, Peng Li, and Chen Qi from the Guangzhou Institutes of Biomedicine and Health for experimental advice. This work was supported by the Shenzhen Key Laboratory of Single-Cell Omics (ZDSYS20190902093613831), Shenzhen Bay Laboratory (SZBL2019062801012) and the Guangdong Provincial Key Laboratory of Genome Read and Write (2017B030301011). Additionally, Longqi Liu was supported by the National Natural Science Foundation of China (31900466) and Shenzhen Basic Research Project for Excellent Young Scholars (2020251518), Miguel A. Esteban's was supported by a Changbai Mountain Scholar award (419020201252), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16030502), and a Chinese Academy of Sciences-Japan Society for the Promotion of Science joint research project (GJHZ2093), Giacomo Volpe was supported by a Chinese Academy of Sciences President's International Fellowship for Foreign Experts (2020FSB0002),

and Mingyuan Liu was supported by the National Key Research and Development Program of China (2021YFC2600200).

AUTHOR CONTRIBUTIONS

L.H., Y.H, X.X., M.A.E. and L.L. conceived the idea; Y.H, X.X., M.A.E. and L.L. supervised the work; L.H., X.W., Y.Y., M.A.E. and L.L designed the experiments; L.H., X.W., G.V., Y.Y., X.Zhang., P.F., P.G., X.L., F.Y., S.S., G.L., J.A., Y.Lei., Y. Lai, M.C., C.W., X.G., S.L. and J.M. collected tissue samples; C.L., G.V., Z.W., Y.Y., X.Zhang., P.F., Q.D., Ya. Liu, Y.Huang, H.L., B.W., M.C., J.X., M.W., C. Wang, Y.Z., Y. Yu, H.Z., Y.W. and S.X. performed the experiments. L.H., X.W., G.V., Z.Z., X. Zou, T.P., Y. Lai, L.W., Q.S., H.Y., Yang Liu, D.X., F.H., Z.Zhu and C.Ward performed data analysis. L.H., X.W., C.L., G.V., Z.Z., X.Zou, Z.Wang, T.P., Y.Yang, J.L. and L.L. prepared the figures. H.Y., X.F.W., F.C., T.Y., W.D. and J.C. prepared the website. Zong.W., Z.P., C.W.W., B.Q, A.S., J.I., L.F., Yan Liu, Z.L., X. Liu, H. Zhang, M.L., Q.S., P.M., N.B., P.M.C., Y.G., J.M., M.U., T.T., S.L., H.Y. and J.W. provided relevant advice and reviewed the manuscript. L.H., G.V., M.A.E. and L.L. wrote the manuscript with input from all authors. All other authors contributed to the work. All authors read and approved the manuscript for submission.

COMPETING INTERESTS

Employees of BGI have stock holdings in BGI. All other authors declare no competing interests.

REFERENCES

1 Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, doi:10.7554/eLife.27041 (2017).

2 Rozenblatt-Rosen, O., Stubbington, M. J. T., Regev, A. & Teichmann, S. A. The Human Cell Atlas: from vision to reality. *Nature* **550**, 451-453, doi:10.1038/550451a (2017).

3 Han, X. *et al.* Construction of a human cell landscape at single-cell level. *Nature* **581**, 303-309, doi:10.1038/s41586-020-2157-4 (2020).

4 Sun, Y. *et al.* Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. *Cell* **184**, 404-421 e416, doi:10.1016/j.cell.2020.11.041 (2021).

5 Segerstolpe, A. *et al.* Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab* **24**, 593-607, doi:10.1016/j.cmet.2016.08.020 (2016).

6 Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091-1107 e1017, doi:10.1016/j.cell.2018.02.001 (2018).

7 Tabula Muris, C. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372, doi:10.1038/s41586-018-0590-4 (2018).

8 Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324 e1318, doi:10.1016/j.cell.2018.06.052 (2018).

9 Carbone, L. *et al.* Gibbon genome and the fast karyotype evolution of small apes. *Nature* **513**, 195-201, doi:10.1038/nature13679 (2014).

10 Ezran, C. *et al.* The Mouse Lemur, a Genetic Model Organism for Primate Biology, Behavior, and Health. *Genetics* **206**, 651-664, doi:10.1534/genetics.116.199448 (2017).

11 Mantini, D. *et al.* Interspecies activity correlations reveal functional correspondence between monkey and human brain areas. *Nat Methods* **9**, 277-282, doi:10.1038/nmeth.1868 (2012).

12 Taylor, K. Clinical veterinarian's perspective of non-human primate (NHP) use in drug safety studies. *J Immunotoxicol* **7**, 114-119, doi:10.1080/15476910903213539 (2010).

13 Estes, J. D., Wong, S. W. & Brenchley, J. M. Nonhuman primate models of human viral infections. *Nat Rev Immunol* **18**, 390-404, doi:10.1038/s41577-018-0005-7 (2018).

14 Zhu, L. *et al.* Single-Cell Sequencing of Peripheral Mononuclear Cells Reveals Distinct Immune Response Landscapes of COVID-19 and Influenza Patients. *Immunity* **53**, 685-696 e683, doi:10.1016/j.immuni.2020.07.009 (2020).

15 Slyper, M. *et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat Med* **26**, 792-802, doi:10.1038/s41591-020-0844-1 (2020).

16 Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14**, 955-958, doi:10.1038/nmeth.4407 (2017).

17 Ding, J. *et al.* Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol* **38**, 737-746, doi:10.1038/s41587-020-0465-8 (2020).

18 Krausgruber, T. *et al.* Structural cells are key regulators of organ-specific immune responses. *Nature* **583**, 296-302, doi:10.1038/s41586-020-2424-4 (2020).

19 Kalucka, J. *et al.* Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell* **180**, 764-779 e720, doi:10.1016/j.cell.2020.01.015 (2020).

20 Geirsdottir, L. *et al.* Cross-Species Single-Cell Analysis Reveals Divergence of the Primate Microglia Program. *Cell* **179**, 1609-1622 e1616, doi:10.1016/j.cell.2019.11.010 (2019).

21 Petrany, M. J. *et al.* Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. *Nat Commun* **11**, 6374, doi:10.1038/s41467-020-20063-w (2020).

22 Mu, L. & Sanders, I. Human tongue neuroanatomy: Nerve supply and motor endplates. *Clin Anat* **23**, 777-791, doi:10.1002/ca.21011 (2010).

23 Chau, Y. Y. *et al.* Visceral and subcutaneous fat have different origins and evidence supports a mesothelial source. *Nat Cell Biol* **16**, 367-375, doi:10.1038/ncb2922 (2014).

24 Vijay, J. *et al.* Single-cell analysis of human adipose tissue identifies depot and disease specific cell types. *Nat Metab* **2**, 97-109, doi:10.1038/s42255-019-0152-6 (2020).

25 Sun, W. *et al.* snRNA-seq reveals a subpopulation of adipocytes that regulates thermogenesis. *Nature* **587**, 98-102, doi:10.1038/s41586-020-2856-x (2020).

26 Ghorpade, D. S. *et al.* Hepatocyte-secreted DPP4 in obesity promotes adipose inflammation and insulin resistance. *Nature* **555**, 673-677, doi:10.1038/nature26138 (2018).

27 Kretzschmar, K. *et al.* Profiling proliferative cells and their progeny in damaged murine hearts. *Proc Natl Acad Sci U S A* **115**, E12245-E12254, doi:10.1073/pnas.1805829115 (2018).

28 Chen, X. *et al.* Adipose-derived mesenchymal stem cells promote the survival of fat grafts via crosstalk between the Nrf2 and TLR4 pathways. *Cell Death Dis* **7**, e2369, doi:10.1038/cddis.2016.261 (2016).

29 Ha, C. W. Y. *et al.* Translocation of Viable Gut Microbiota to Mesenteric Adipose Drives Formation of Creeping Fat in Humans. *Cell* **183**, 666-683 e617, doi:10.1016/j.cell.2020.09.009 (2020).

30 Adler, E., Mhawech-Fauceglia, P., Gayther, S. A. & Lawrenson, K. PAX8 expression in ovarian surface epithelial cells. *Hum Pathol* **46**, 948-956, doi:10.1016/j.humpath.2015.03.017 (2015).

31 Ng, A. *et al.* Lgr5 marks stem/progenitor cells in ovary and tubal epithelia. *Nat Cell Biol* **16**, 745-757, doi:10.1038/ncb3000 (2014).

32 Hagai, T. *et al.* Gene expression variability across cells and species shapes innate immunity. *Nature* **563**, 197-202, doi:10.1038/s41586-018-0657-2 (2018).

33 Nusse, R. & Clevers, H. Wnt/β-Catenin Signaling, Disease, and Emerging Therapeutic Modalities. *Cell* **169**, 985-999, doi:10.1016/j.cell.2017.05.016 (2017).

34 Clevers, H. & Nusse, R. Wnt/β-catenin signaling and disease. *Cell* **149**, 1192-1205, doi:10.1016/j.cell.2012.05.012 (2012).

35 Nusse, R. Wnt signaling and stem cell control. *Cell Res* **18**, 523-527, doi:10.1038/cr.2008.47 (2008).

36 Leung, C., Tan, S. H. & Barker, N. Recent Advances in Lgr5(+) Stem Cell Research. *Trends Cell Biol* **28**, 380-391, doi:10.1016/j.tcb.2018.01.010 (2018).

37 Kong, Y. *et al.* LGR6 Promotes Tumor Proliferation and Metastasis through Wnt/β-Catenin Signaling in Triple-Negative Breast Cancer. *Mol Ther Oncolytics* **18**, 351-359, doi:10.1016/j.omto.2020.06.020 (2020).

38 Huch, M. *et al.* In vitro expansion of single Lgr5+ liver stem cells induced by Wnt-driven regeneration. *Nature* **494**, 247-250, doi:10.1038/nature11826 (2013).

39 Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003-1007, doi:10.1038/nature06196 (2007).

40 Chen, M. *et al.* Lgr5(+) amacrine cells possess regenerative potential in the retina of adult mice. *Aging Cell* **14**, 635-643, doi:10.1111/acel.12346 (2015).

41 Aguilar, A. Development: Pax2 keeps nephron progenitors on track. *Nat Rev Nephrol* **13**, 444, doi:10.1038/nrneph.2017.87 (2017).

42 Schutgens, F. *et al.* Troy/TNFRSF19 marks epithelial progenitor cells during mouse kidney development that continue to contribute to turnover in adult kidney. *Proc Natl Acad Sci U S A* **114**, E11190-E11198, doi:10.1073/pnas.1714145115 (2017).

43 Lake, B. B. *et al.* A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nat Commun* **10**, 2832, doi:10.1038/s41467-019-10861-2 (2019).

44 Wu, H., Kirita, Y., Donnelly, E. L. & Humphreys, B. D. Advantages of Single-Nucleus over Single-Cell RNA Sequencing of Adult Kidney: Rare Cell Types and Novel Cell States Revealed in Fibrosis. *J Am Soc Nephrol* **30**, 23-32, doi:10.1681/ASN.2018090912 (2019).

45 Barker, N. *et al.* Lgr5(+ve) stem/progenitor cells contribute to nephron formation during kidney development. *Cell Rep* **2**, 540-552, doi:10.1016/j.celrep.2012.08.018 (2012).

46 Lake, B. B. *et al.* Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol* **36**, 70-80, doi:10.1038/nbt.4038 (2018).

47 Nakashima, H. *et al.* R-spondin 2 promotes acetylcholine receptor clustering at the neuromuscular junction via Lgr5. *Sci Rep* **6**, 28512, doi:10.1038/srep28512 (2016).

48 Leung, C. *et al.* Lgr5 Marks Adult Progenitor Cells Contributing to Skeletal Muscle Regeneration and Sarcoma Formation. *Cell Rep* **33**, 108535, doi:10.1016/j.celrep.2020.108535 (2020).

49 Litvinukova, M. *et al.* Cells of the adult human heart. *Nature* **588**, 466-472, doi:10.1038/s41586-020-2797-4 (2020).

50 Vidal, R. *et al.* Transcriptional heterogeneity of fibroblasts is a hallmark of the aging heart. *JCI Insight* **4**, doi:10.1172/jci.insight.131092 (2019).

51 Vankelecom, H. Non-hormonal cell types in the pituitary candidating for stem cell. *Semin Cell Dev Biol* **18**, 559-570, doi:10.1016/j.semcdb.2007.04.006 (2007).

52 Klein, D. *et al.* Wnt2 acts as a cell type-specific, autocrine growth factor in rat hepatic sinusoidal endothelial cells cross-stimulating the VEGF pathway. *Hepatology* **47**, 1018-1031, doi:10.1002/hep.22084 (2008).

53 Karner, C. M. *et al.* Canonical Wnt9b signaling balances progenitor cell expansion and differentiation during kidney development. *Development* **138**, 1247-1257, doi:10.1242/dev.057646 (2011).

54 Niehrs, C. The complex world of WNT receptor signalling. *Nat Rev Mol Cell Biol* **13**, 767-779, doi:10.1038/nrm3470 (2012).

55 Zhang, M. *et al.* β-Catenin safeguards the ground state of mouse pluripotency by strengthening the robustness of the transcriptional apparatus. *Sci Adv* **6**, eaba1593, doi:10.1126/sciadv.aba1593 (2020).

56 Devakumar, D. *et al.* Infectious causes of microcephaly: epidemiology, pathogenesis, diagnosis, and management. *Lancet Infect Dis* **18**, e1-e13, doi:10.1016/S1473-3099(17)30398-5 (2018).

57 Dhiman, N., Jacobson, R. M. & Poland, G. A. Measles virus receptors: SLAM and CD46. *Rev Med Virol* **14**, 217-229, doi:10.1002/rmv.430 (2004).

58 Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* **382**, 727-733, doi:10.1056/NEJMoa2001017 (2020).

59 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 e278, doi:10.1016/j.cell.2020.02.052 (2020).

60 Rockx, B. *et al.* Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. *Science* **368**, 1012-1015, doi:10.1126/science.abb7314 (2020).

61 Teichmann, S. & Regev, A. The network effect: studying COVID-19 pathology with the Human Cell Atlas. *Nat Rev Mol Cell Biol* **21**, 415-416, doi:10.1038/s41580-020-0267-3 (2020).

62 Ying, M. *et al.* COVID-19 with acute cholecystitis: a case report. *BMC Infect Dis* **20**, 437, doi:10.1186/s12879-020-05164-7 (2020).

63 Ziegler, C. G. K. *et al.* SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell* **181**, 1016-1035 e1019, doi:10.1016/j.cell.2020.04.035 (2020).

64 Chua, R. L. *et al.* COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol* **38**, 970-979, doi:10.1038/s41587-020-0602-4 (2020).

65 Tosi, M. F. Innate immune responses to infection. *J Allergy Clin Immunol* **116**, 241-249; quiz 250, doi:10.1016/j.jaci.2005.05.036 (2005).

66 Bell, L. C. K. *et al.* Transcriptional response modules characterize IL-1β and IL-6 activity in COVID-19. *iScience* **24**, 101896, doi:10.1016/j.isci.2020.101896 (2021).

67 Shen, B. *et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* **182**, 59-72 e15, doi:10.1016/j.cell.2020.05.032 (2020).

68 Nie, X. *et al.* Multi-organ proteomic landscape of COVID-19 autopsies. *Cell* **184**, 775-791 e714, doi:10.1016/j.cell.2021.01.004 (2021).

69 Gate, D. *et al.* Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease. *Nature* **577**, 399-404, doi:10.1038/s41586-019-1895-7 (2020).

70 Zhong, J., Yang, H. & Kon, V. Kidney as modulator and target of "good/bad" HDL. *Pediatr Nephrol* **34**, 1683-1695, doi:10.1007/s00467-018-4104-2 (2019).

71 Chamorro, Y., Ramirez-Duenas, M. L. & Matute, E. Anticipatory oculomotor responses in parents of children with attention deficit hyperactivity disorder. *Psychiatr Genet* **30**, 65-72, doi:10.1097/YPG.0000000000000252 (2020).

72 Ghaoui, R. *et al.* Mutations in HSPB8 causing a new phenotype of distal myopathy and motor neuropathy. *Neurology* **86**, 391-398, doi:10.1212/WNL.0000000000002324 (2016).

73 Tse, K. H. & Herrup, K. DNA damage in the oligodendrocyte lineage and its role in brain aging. *Mech Ageing Dev* **161**, 37-50, doi:10.1016/j.mad.2016.05.006 (2017).

74 Wang, S. *et al.* Single-Cell Transcriptomic Atlas of Primate Ovarian Aging. *Cell* **180**, 585-600 e519, doi:10.1016/j.cell.2020.01.009 (2020).

75 Khrameeva, E. *et al.* Single-cell-resolution transcriptome map of human, chimpanzee, bonobo, and macaque brains. *Genome Res* **30**, 776-789, doi:10.1101/gr.256958.119 (2020).

76 Stray, L. L. *et al.* Motor regulation problems and pain in adults diagnosed with ADHD. *Behav Brain Funct* **9**, 18, doi:10.1186/1744-9081-9-18 (2013).

77 Fliers, E. A. *et al.* Genome-wide association study of motor coordination problems in ADHD identifies genes for brain and muscle function. *World J Biol Psychiatry* **13**, 211-222, doi:10.3109/15622975.2011.560279 (2012).

78 Villiger, P. M. *et al.* Tocilizumab for induction and maintenance of remission in giant cell arteritis: a phase 2, randomised, double-blind, placebo-controlled trial. *Lancet* **387**, 1921-1927, doi:10.1016/S0140-6736(16)00560-2 (2016).

79 Marcheque, J., Bussolati, B., Csete, M. & Perin, L. Concise Reviews: Stem Cells and Kidney Regeneration: An Update. *Stem Cells Transl Med* **8**, 82-92, doi:10.1002/sctm.18-0115 (2019).

80 Little, M. H. & Lawlor, K. T. Recreating, expanding and using nephron progenitor populations. *Nat Rev Nephrol* **16**, 75-76, doi:10.1038/s41581-019-0238-0 (2020).

81 Chavali, M. *et al.* Wnt-Dependent Oligodendroglial-Endothelial Interactions Regulate White Matter Vascularization and Attenuate Injury. *Neuron* **108**, 1130-1145 e1135, doi:10.1016/j.neuron.2020.09.033 (2020).

82 Girardi, F. & Le Grand, F. Wnt Signaling in Skeletal Muscle Development and Regeneration. *Prog Mol Biol Transl Sci* **153**, 157-179, doi:10.1016/bs.pmbts.2017.11.026 (2018).

83 Heallen, T. *et al.* Hippo pathway inhibits Wnt signaling to restrain cardiomyocyte proliferation and heart size. *Science* **332**, 458-461, doi:10.1126/science.1199010 (2011).

84 Osmundsen, A. M., Keisler, J. L., Taketo, M. M. & Davis, S. W. Canonical WNT Signaling Regulates the Pituitary Organizer and Pituitary Gland Formation. *Endocrinology* **158**, 3339-3353, doi:10.1210/en.2017-00581 (2017).

85 Chen, A. *et al.* Large field of view-spatially resolved transcriptomics at nanoscale resolution. *bioRxiv*, 2021.2001.2017.427004, doi:10.1101/2021.01.17.427004 (2021).

86 Mitchell, J. F. & Leopold, D. A. The marmoset monkey as a model for visual neuroscience. *Neurosci Res* **93**, 20-46, doi:10.1016/j.neures.2015.01.008 (2015).

87 Bakken, T. E. *et al.* Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. *PLoS One* **13**, e0209648, doi:10.1371/journal.pone.0209648 (2018).

88 Laukoter, S. *et al.* Cell-Type Specificity of Genomic Imprinting in Cerebral Cortex. *Neuron* **107**, 1160-1179 e1169, doi:10.1016/j.neuron.2020.06.031 (2020).

89 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).

90 Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032-2034, doi:10.1093/bioinformatics/btv098 (2015).

91 Del-Aguila, J. L. *et al.* A single-nuclei RNA sequencing study of Mendelian and sporadic AD in the human brain. *Alzheimers Res Ther* **11**, 71, doi:10.1186/s13195-019-0524-x (2019).

92 McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst* **8**, 329-337 e324, doi:10.1016/j.cels.2019.03.003 (2019).

93 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).

94 Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821, doi:10.1016/j.cell.2019.05.031 (2019).

95 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).

96 Aizarani, N. *et al.* A human liver cell atlas reveals heterogeneity and epithelial progenitors. *Nature* **572**, 199-204, doi:10.1038/s41586-019-1373-2 (2019).

97 Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386, doi:10.1038/nbt.2859 (2014).

98 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, doi:10.1038/s41596-020-0292-x (2020).

99 Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet*, doi:10.1038/s41588-021-00790-6 (2021).

100 Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* **11**, e0163962, doi:10.1371/journal.pone.0163962 (2016).

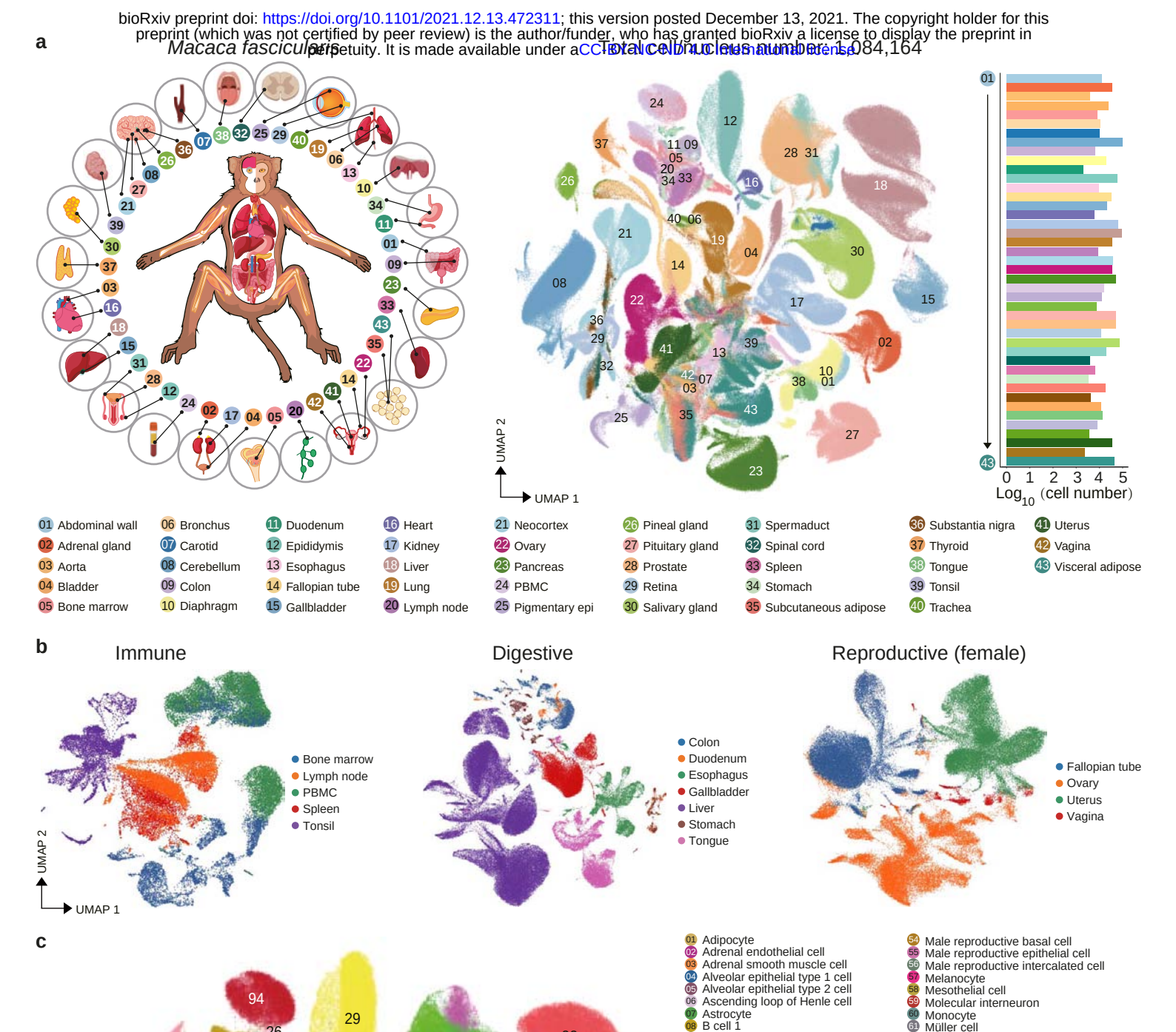


Figure 1

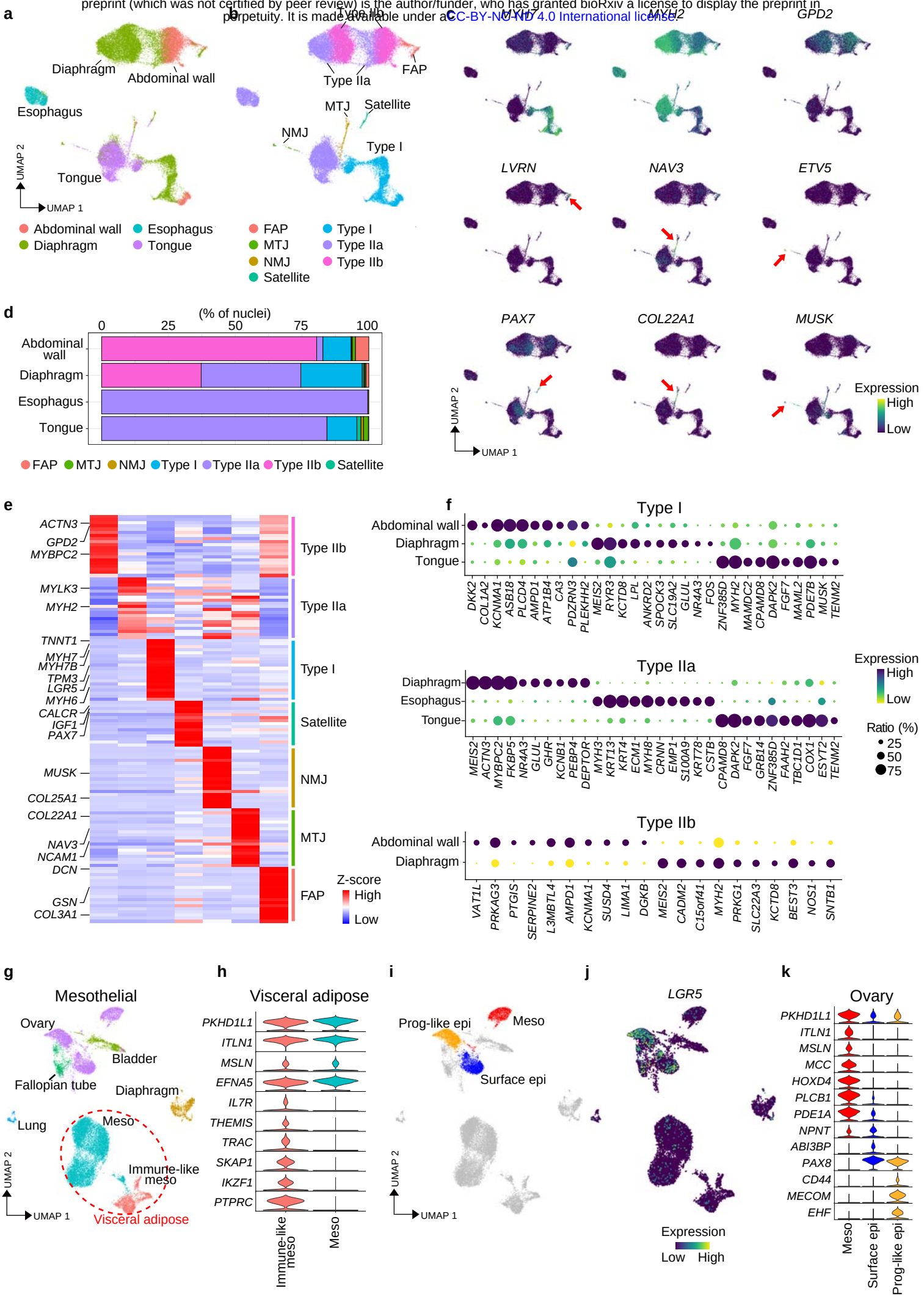


Figure 2

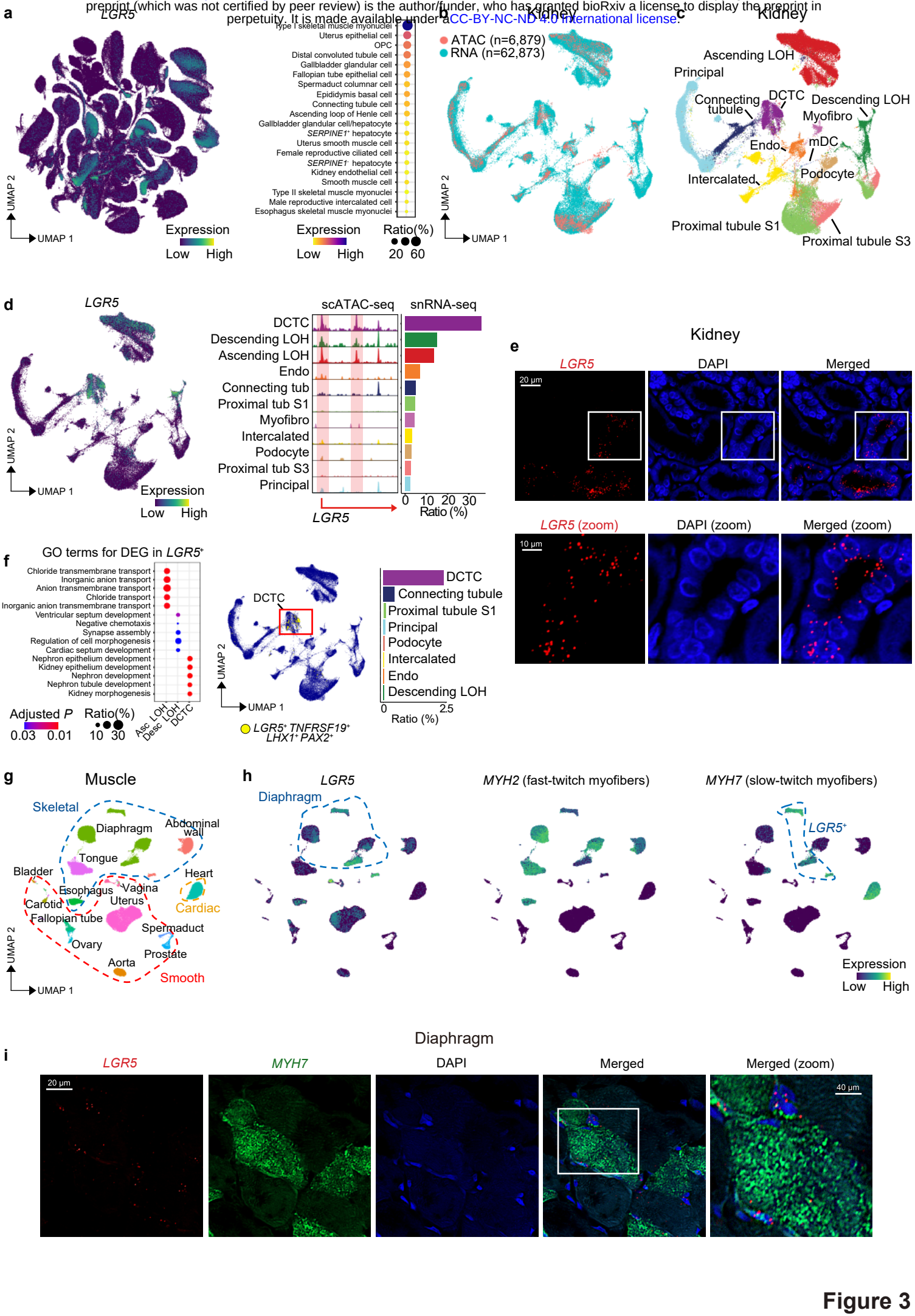


Figure 3

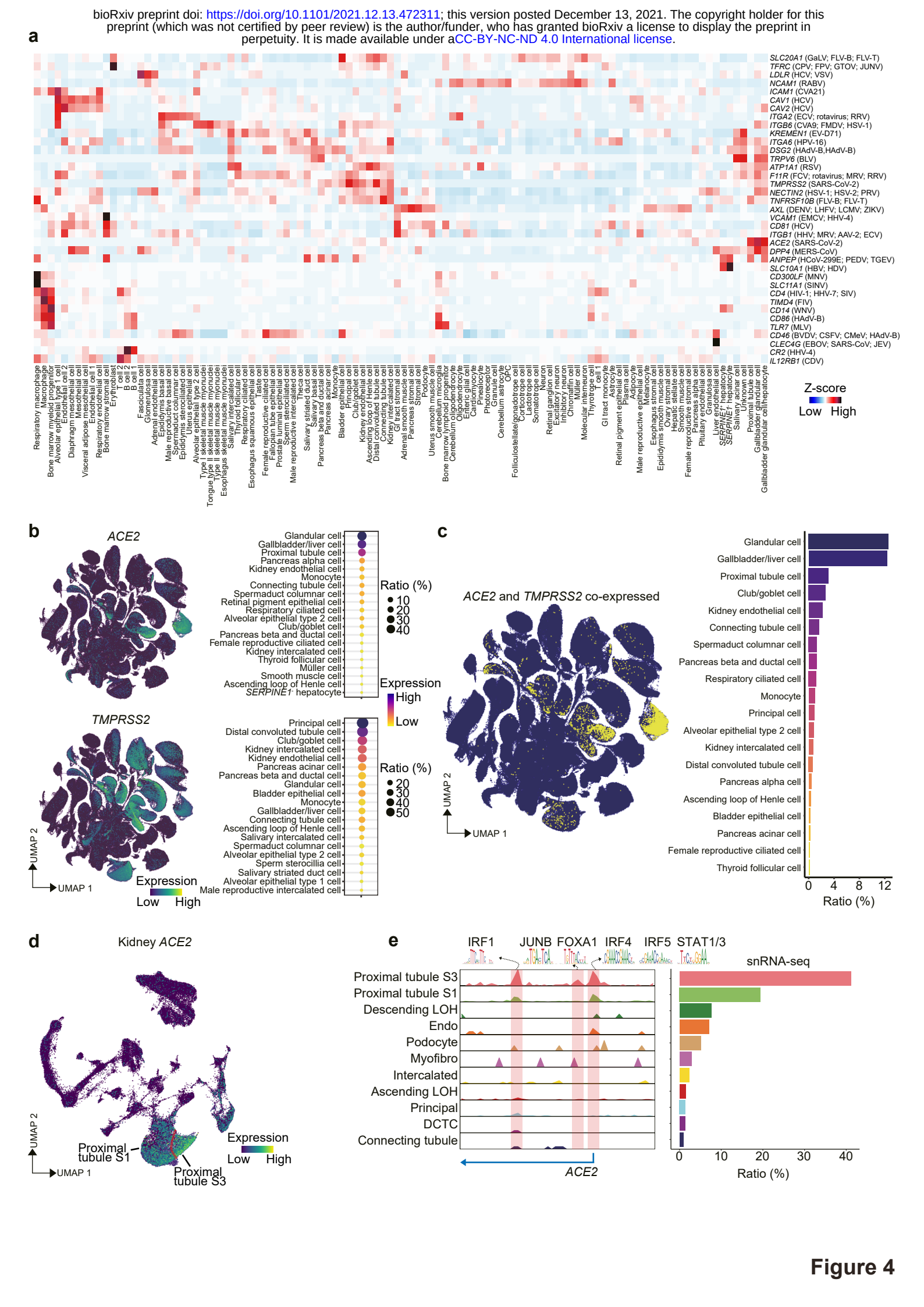


Figure 4

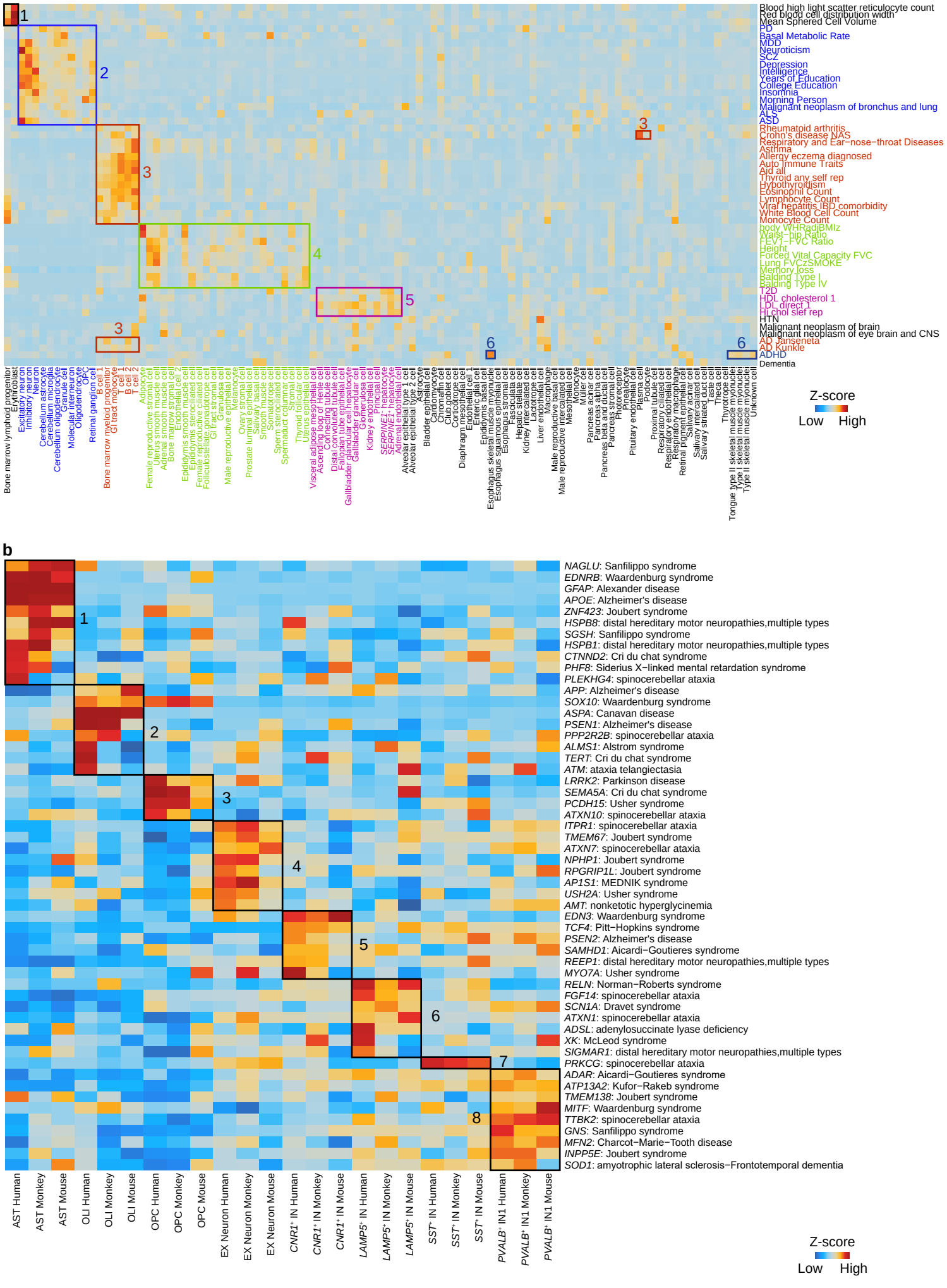
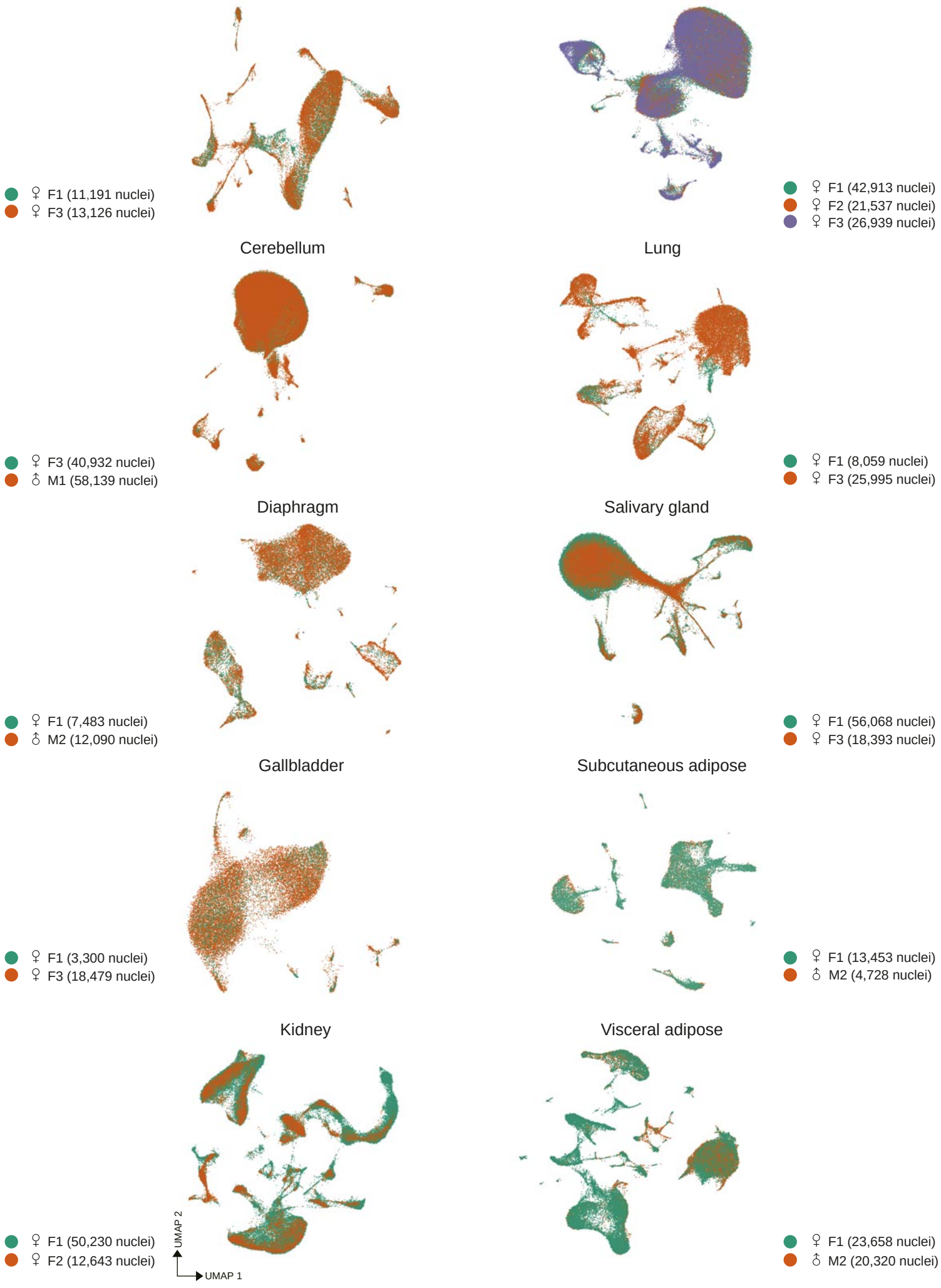
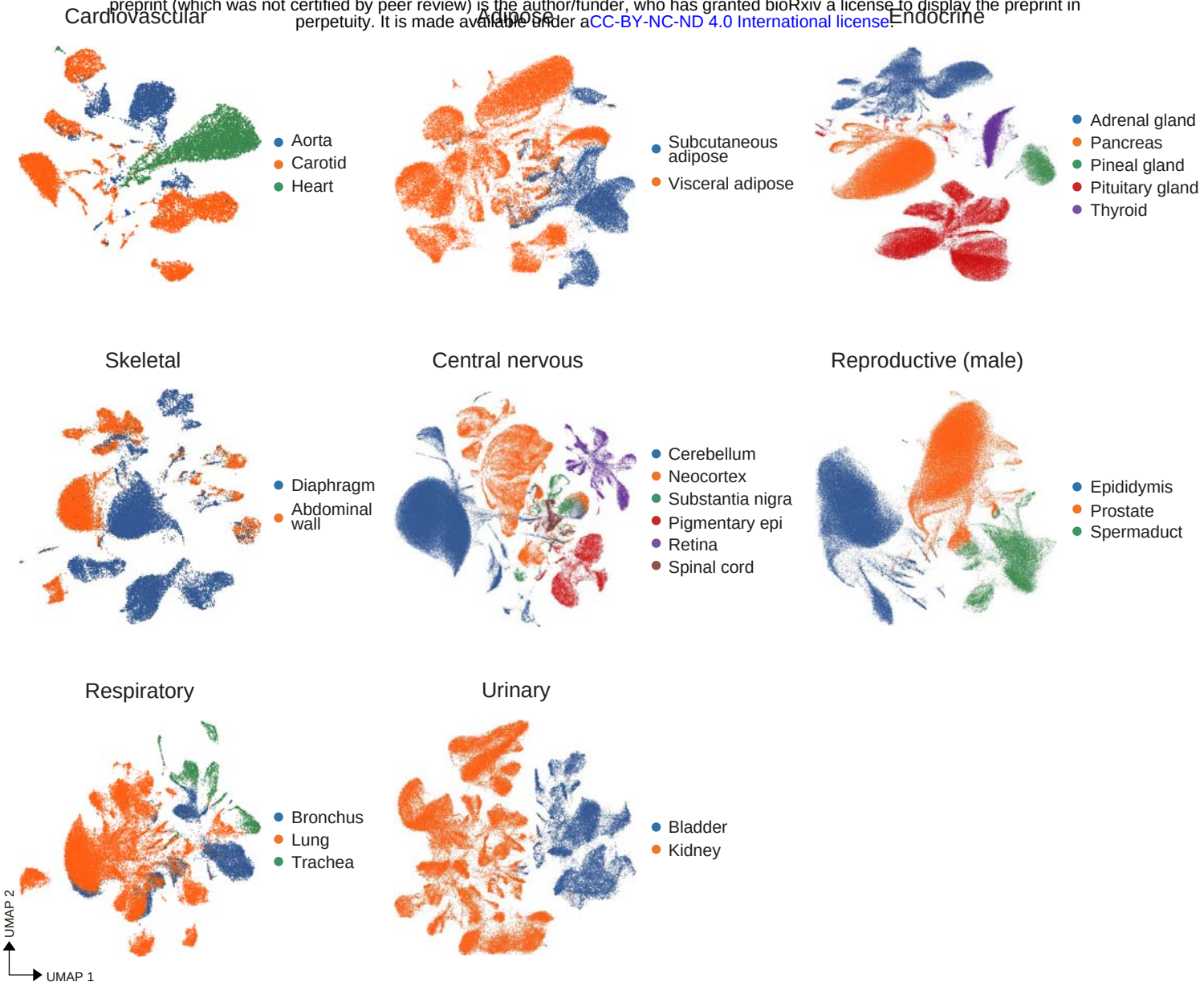


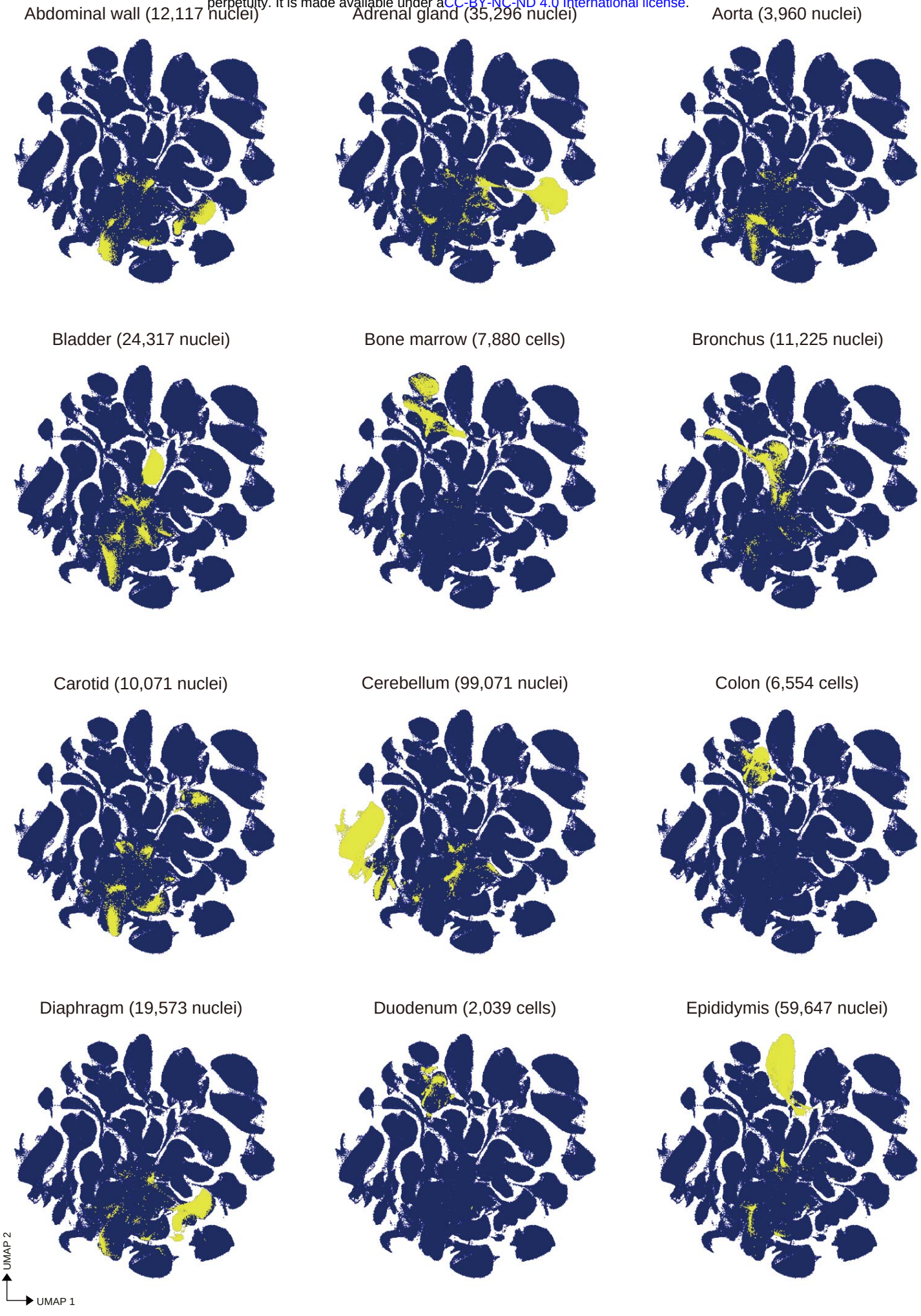
Figure 5



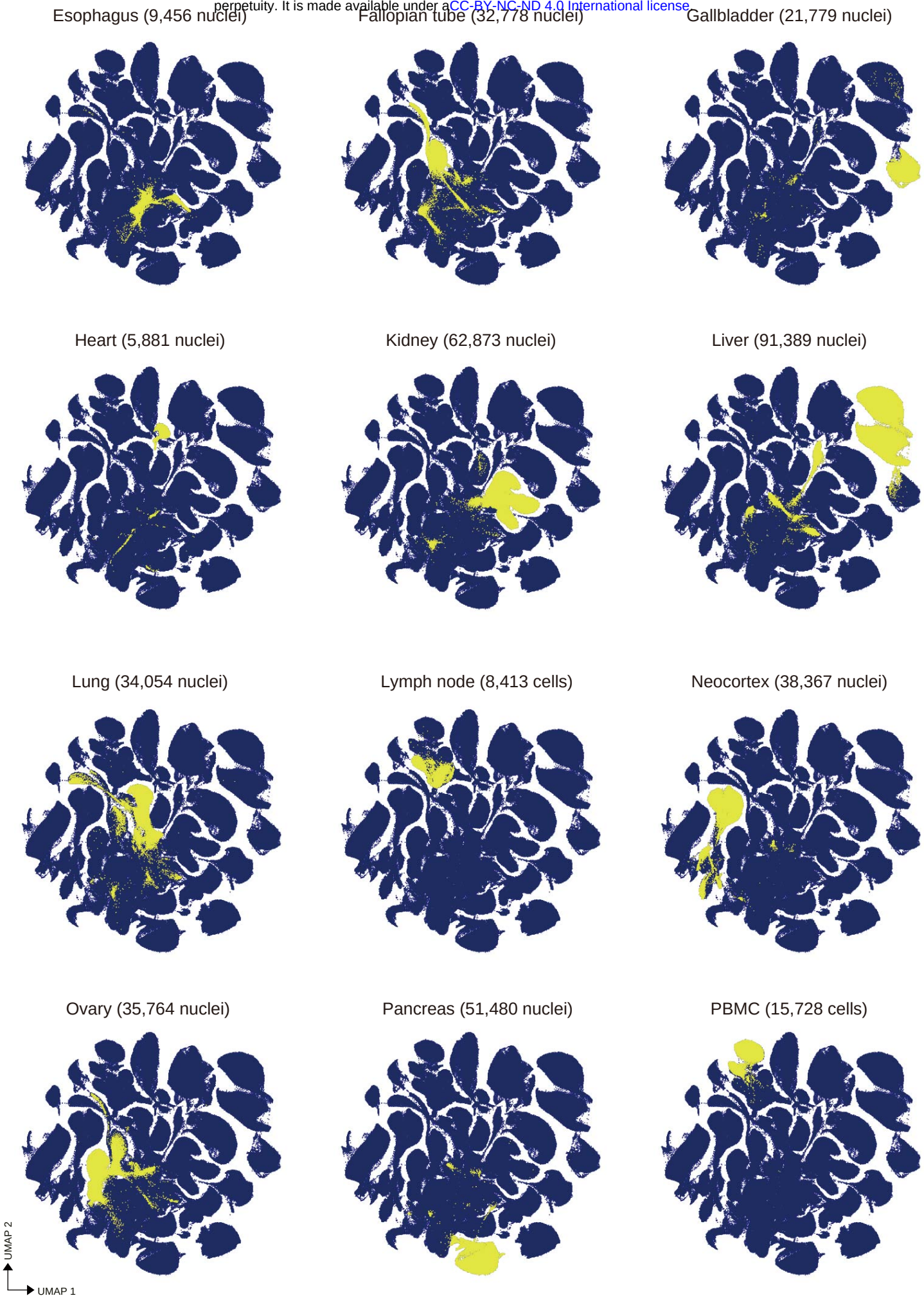
Extended Data Figure 1



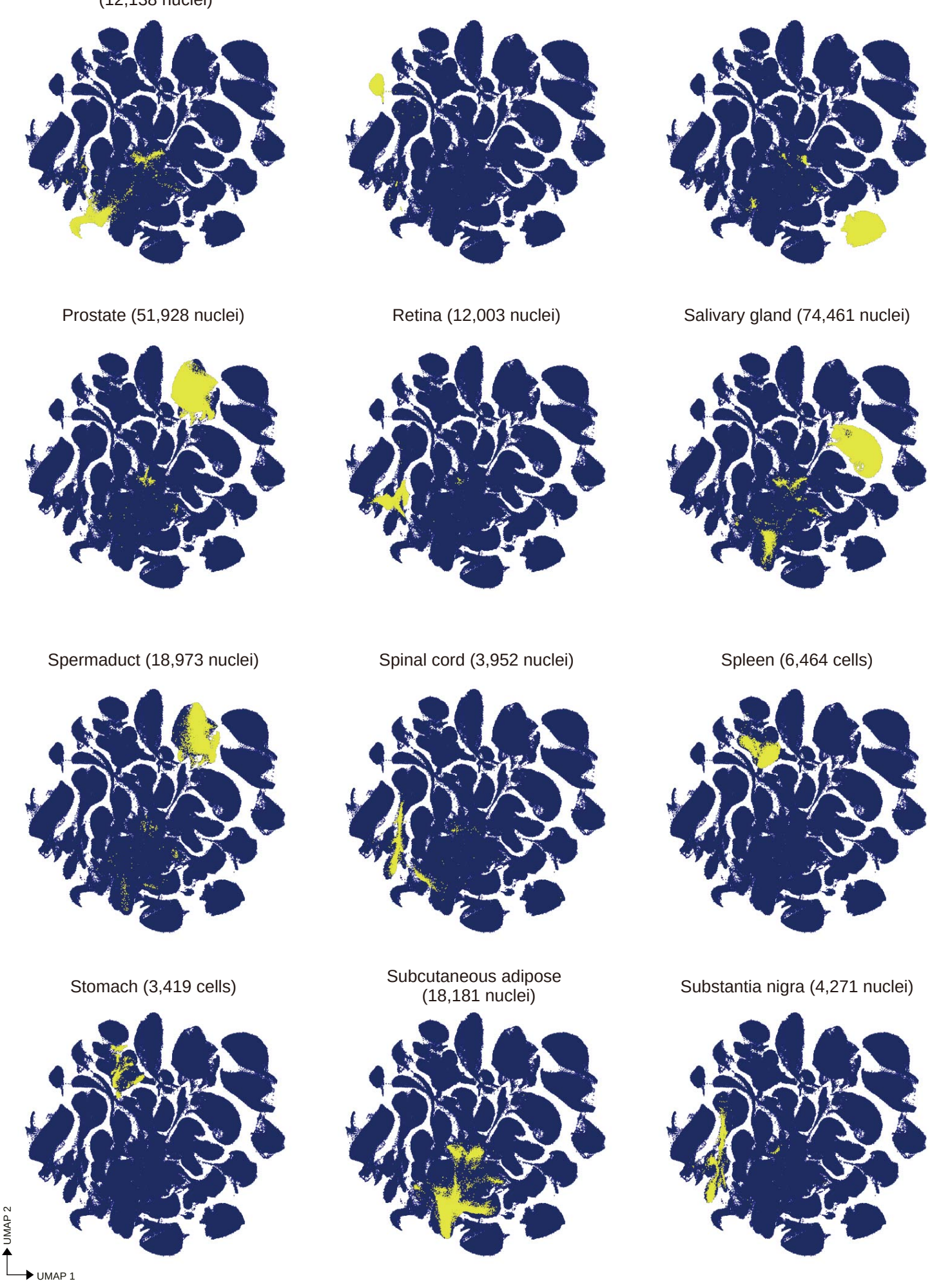
Extended Data Figure 2



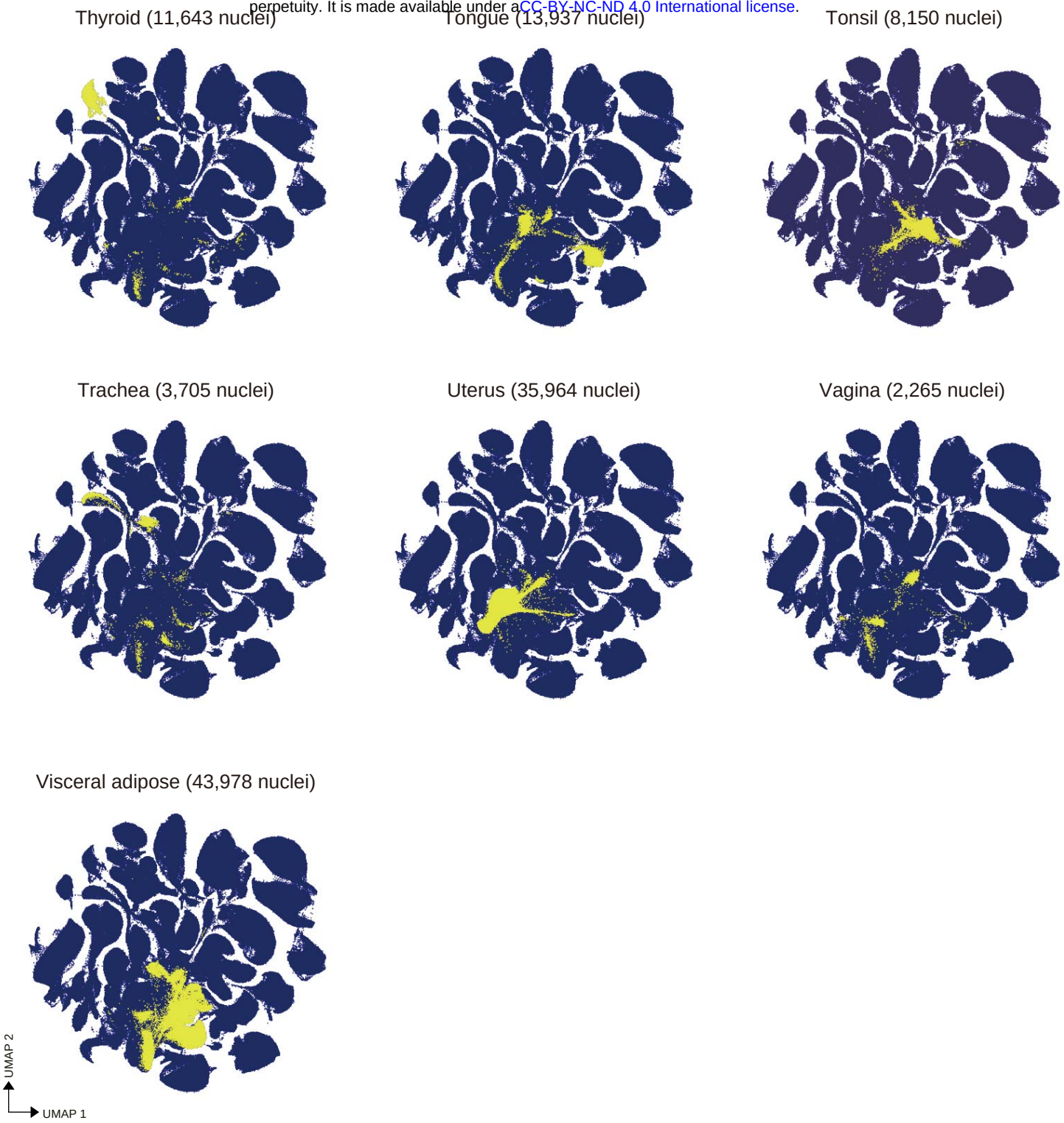
Extended Data Figure 3



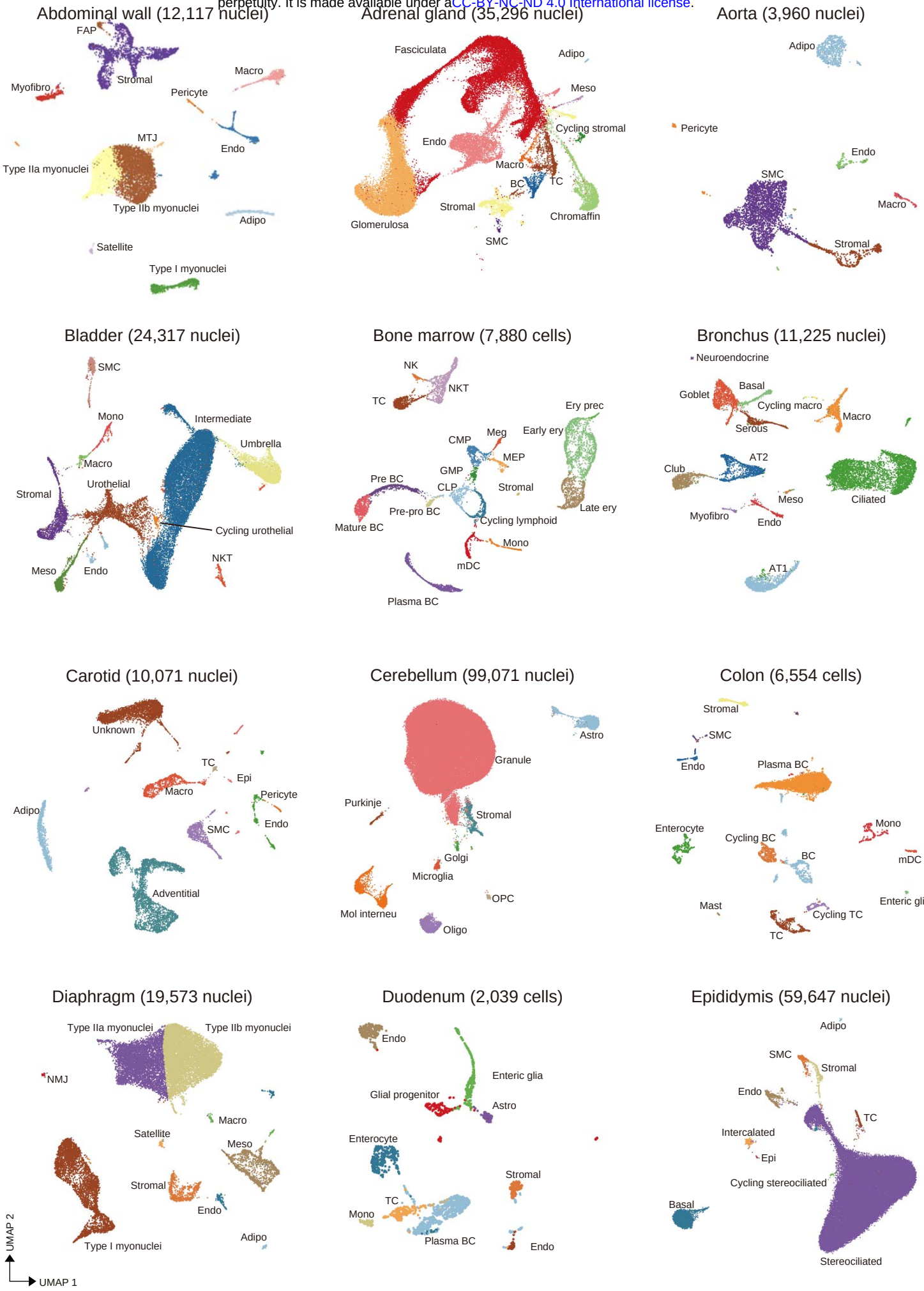
Extended Data Figure 4



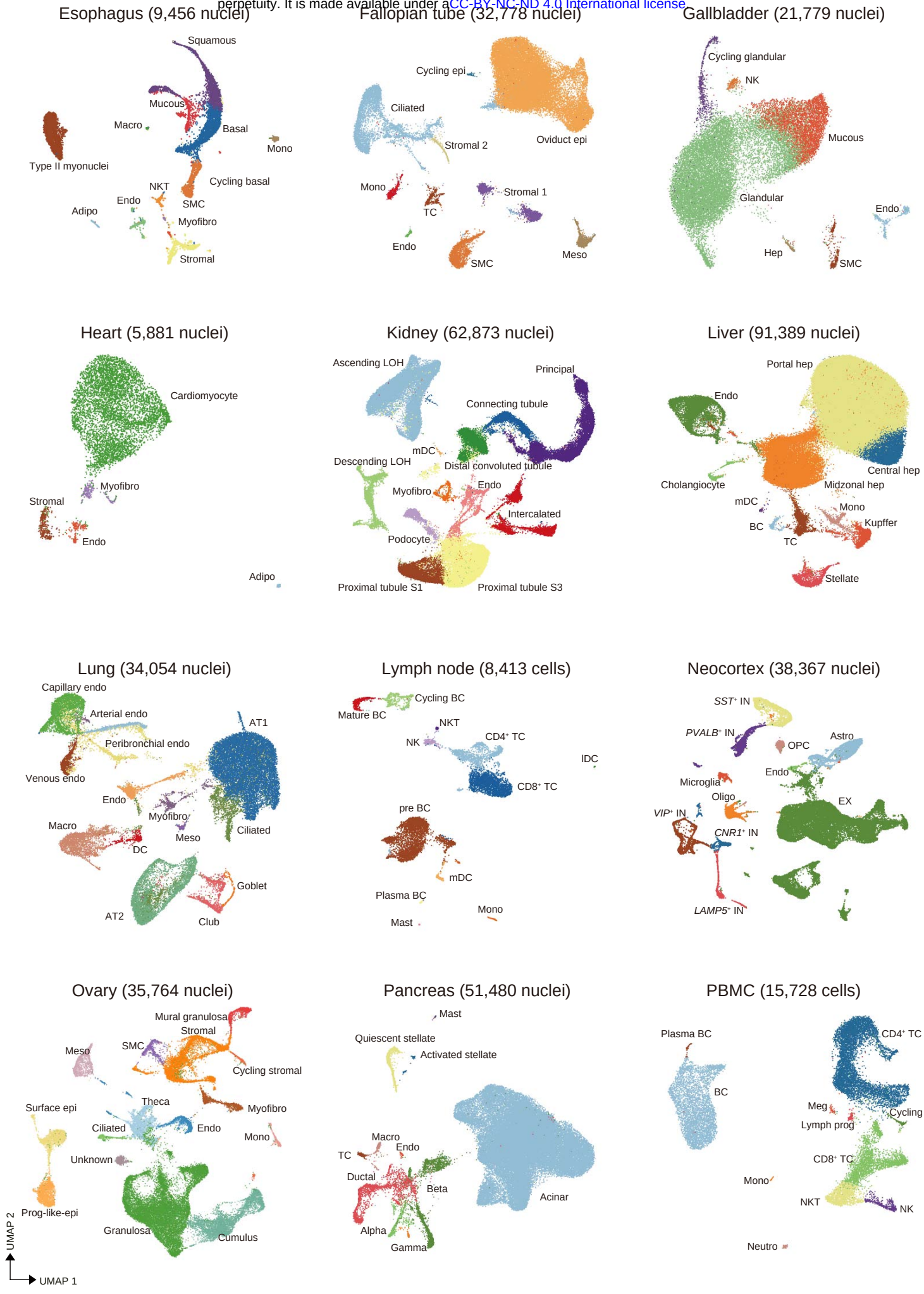
Extended Data Figure 5



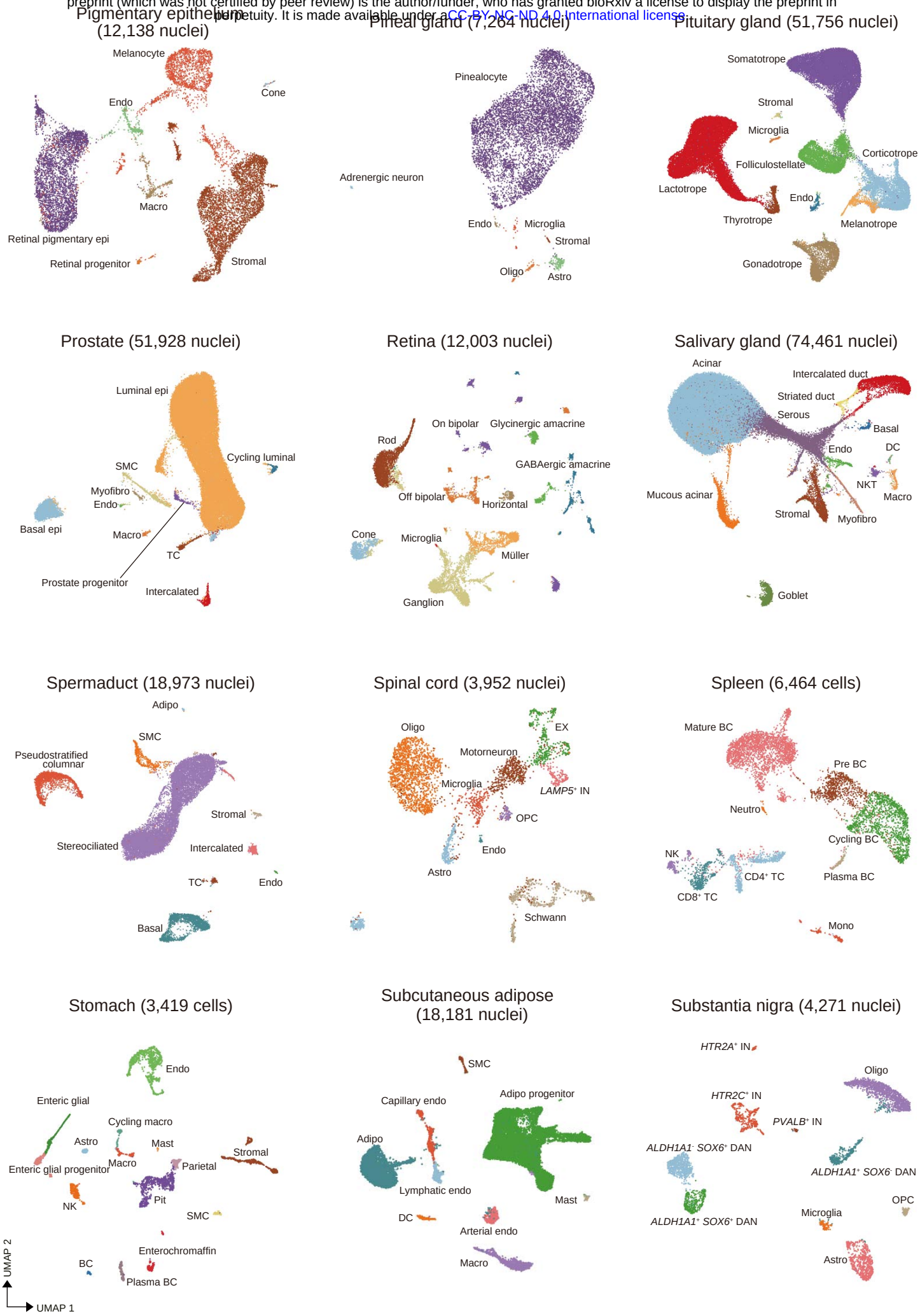
Extended Data Figure 6



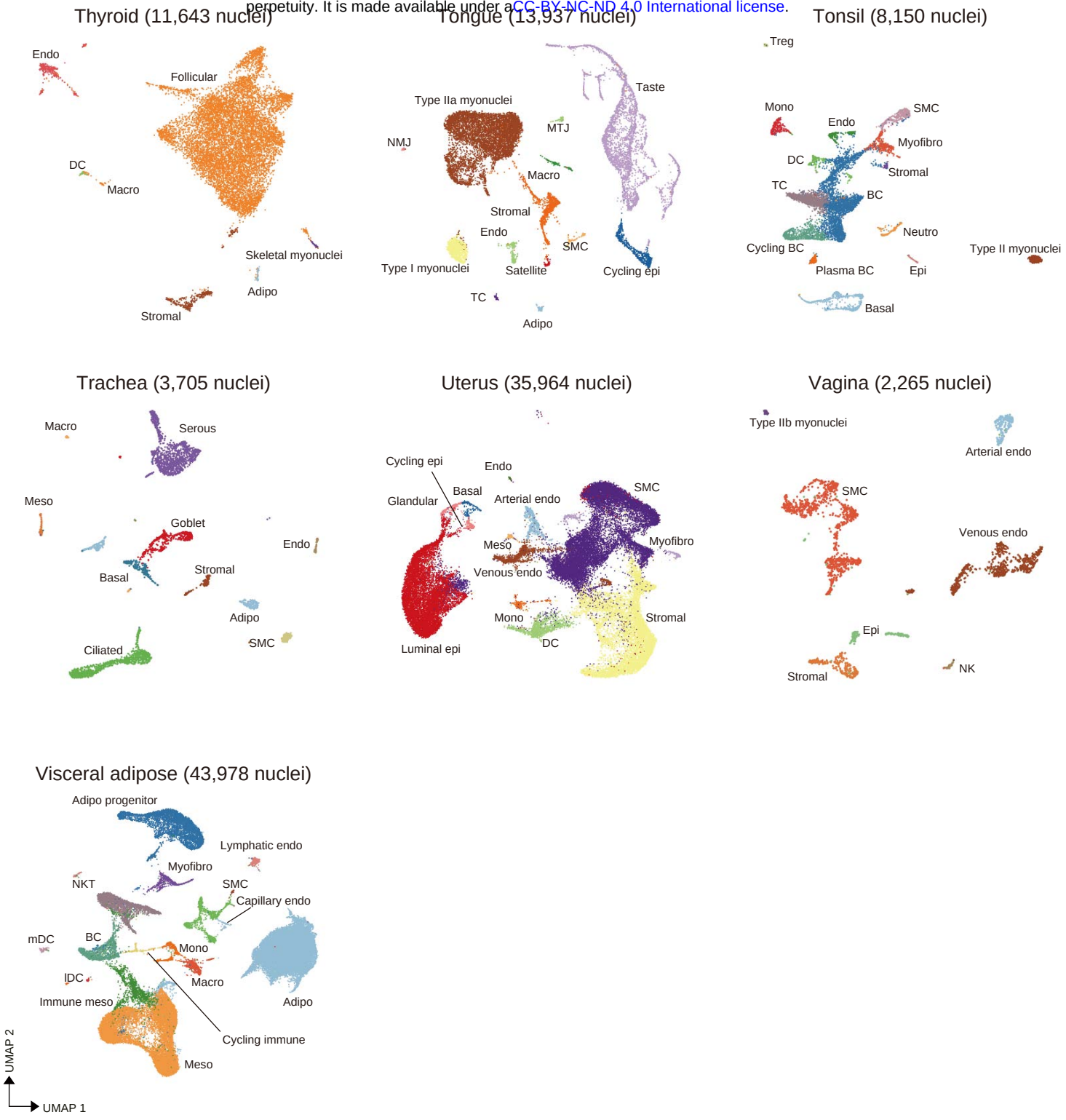
Extended Data Figure 7



Extended Data Figure 8



Extended Data Figure 9



Extended Data Figure 10

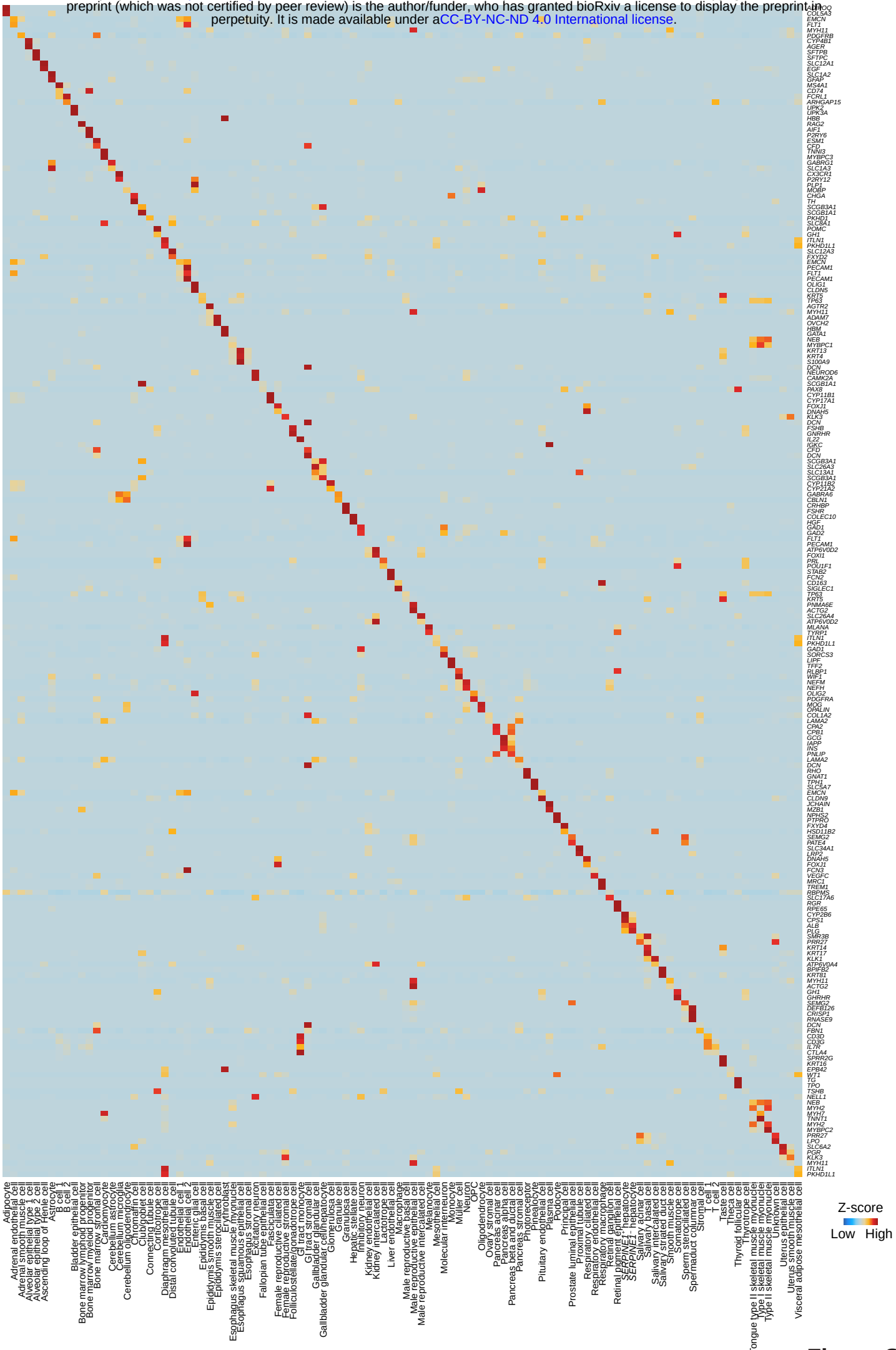
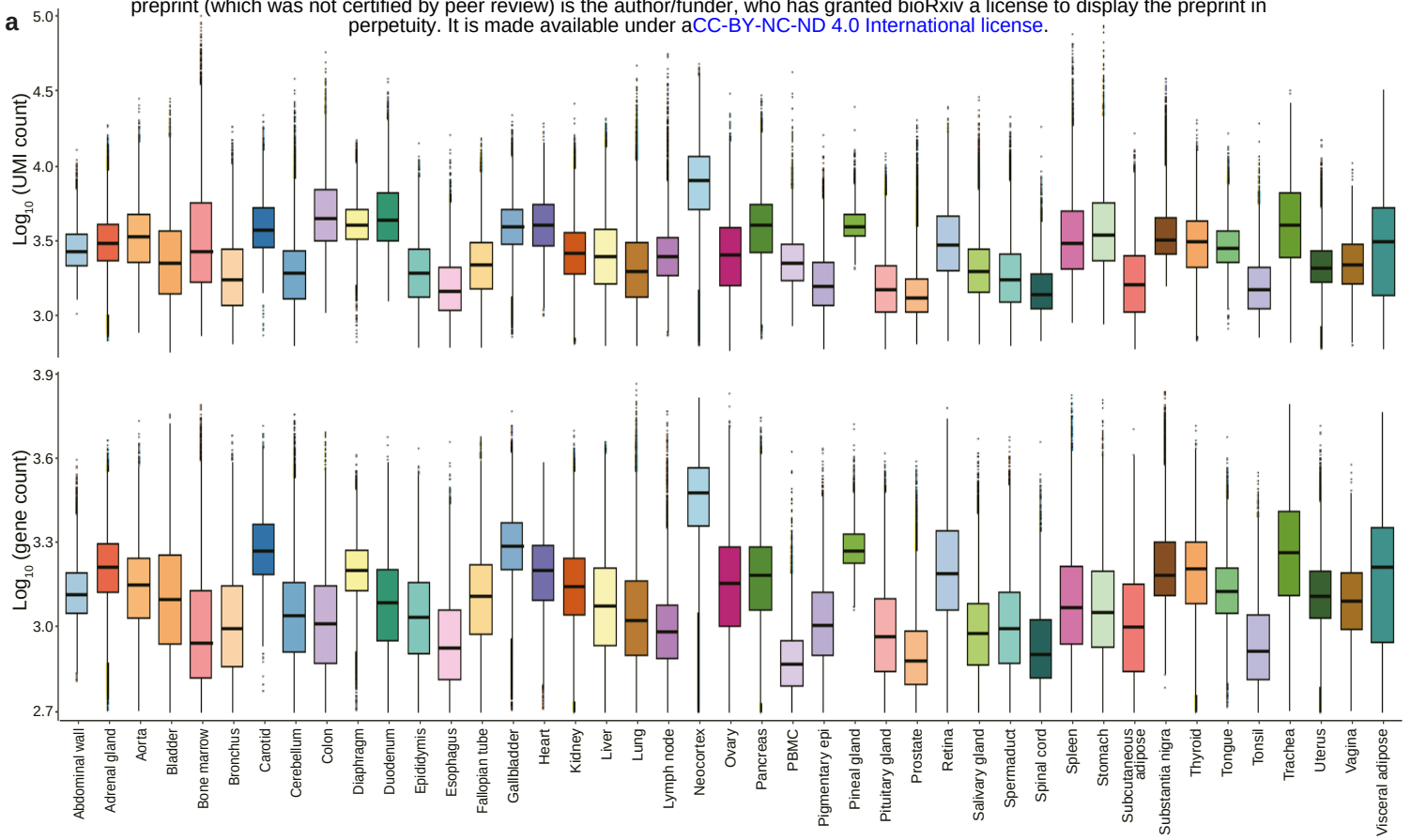
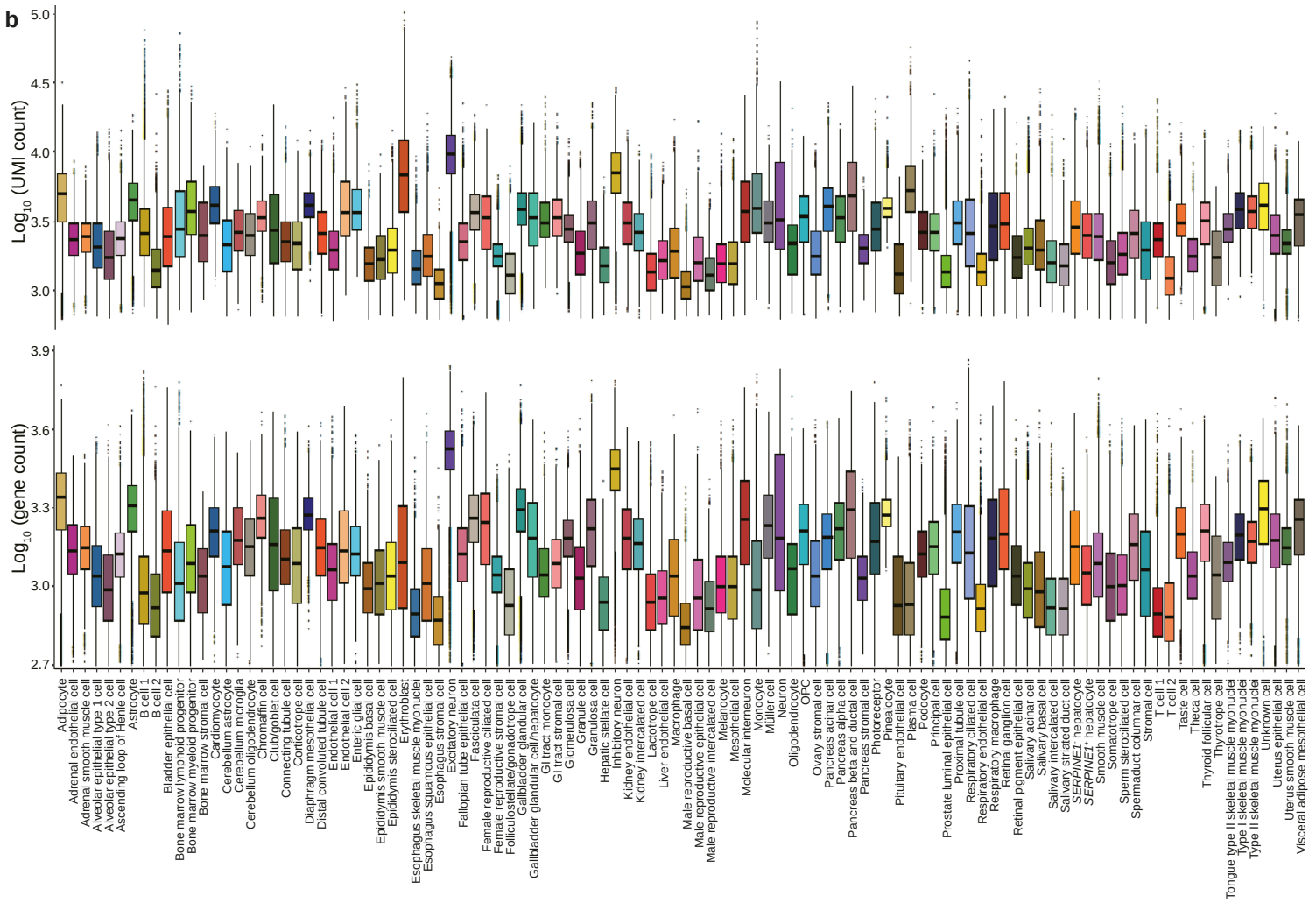
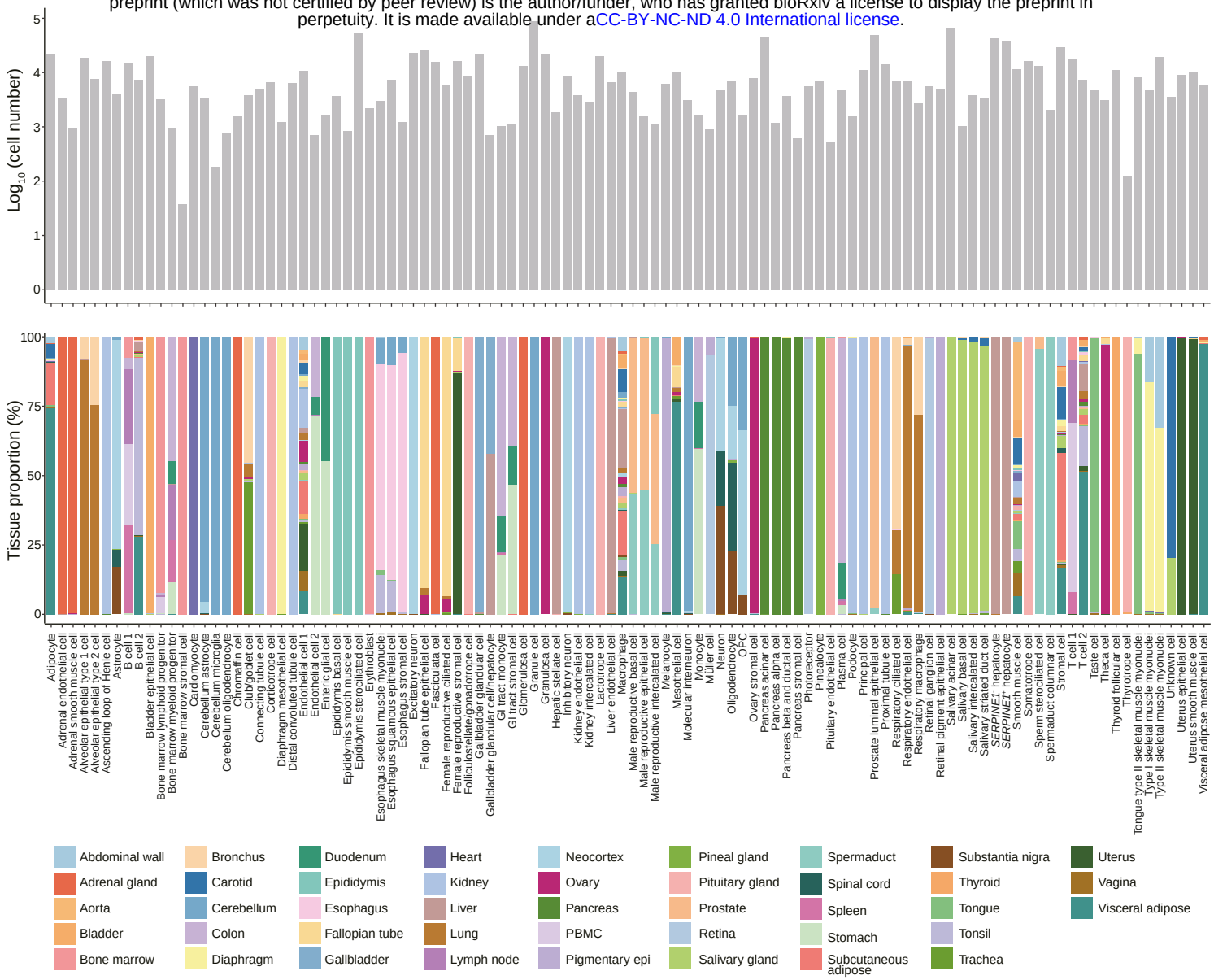


Figure S11



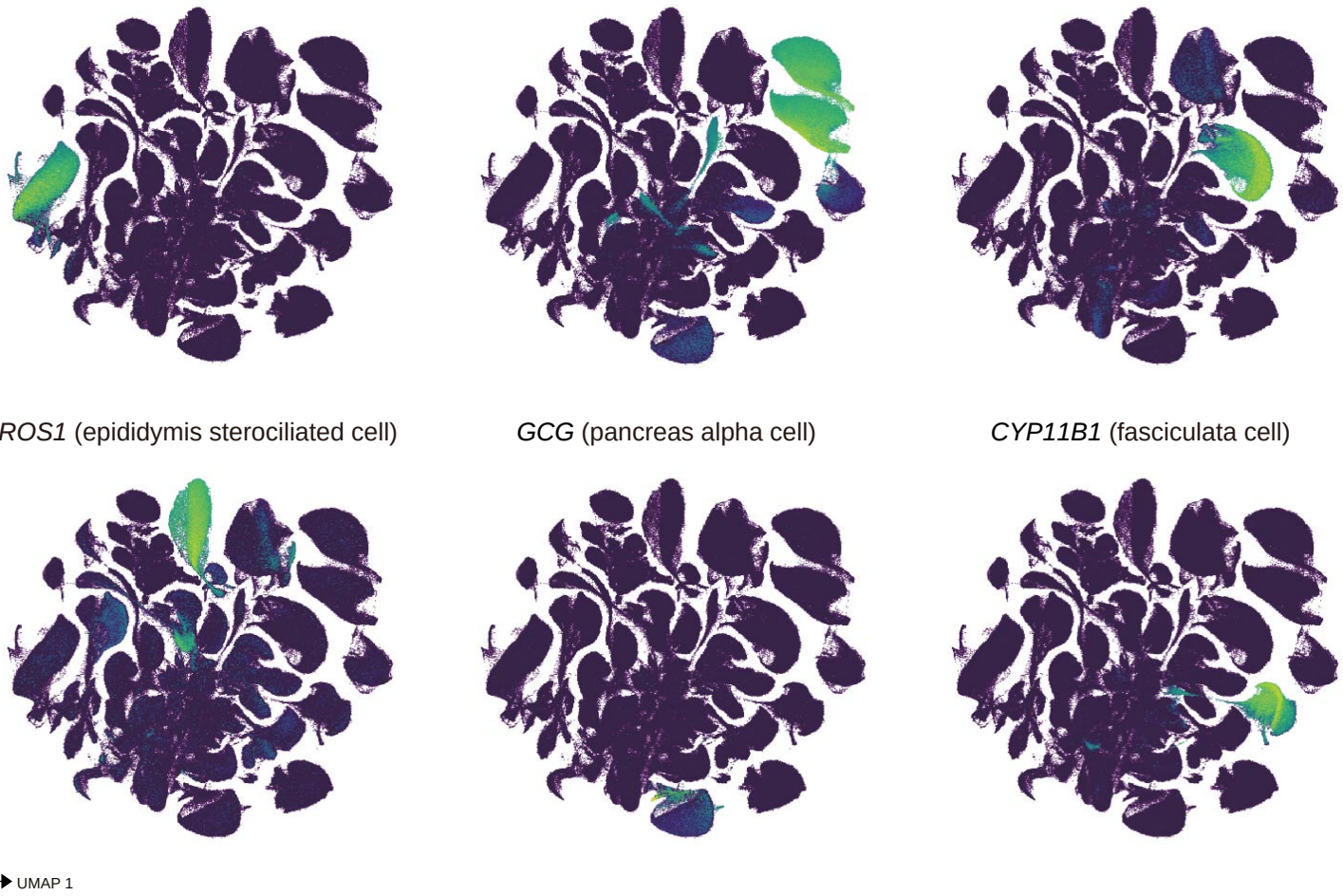


Extended Data Figure 12

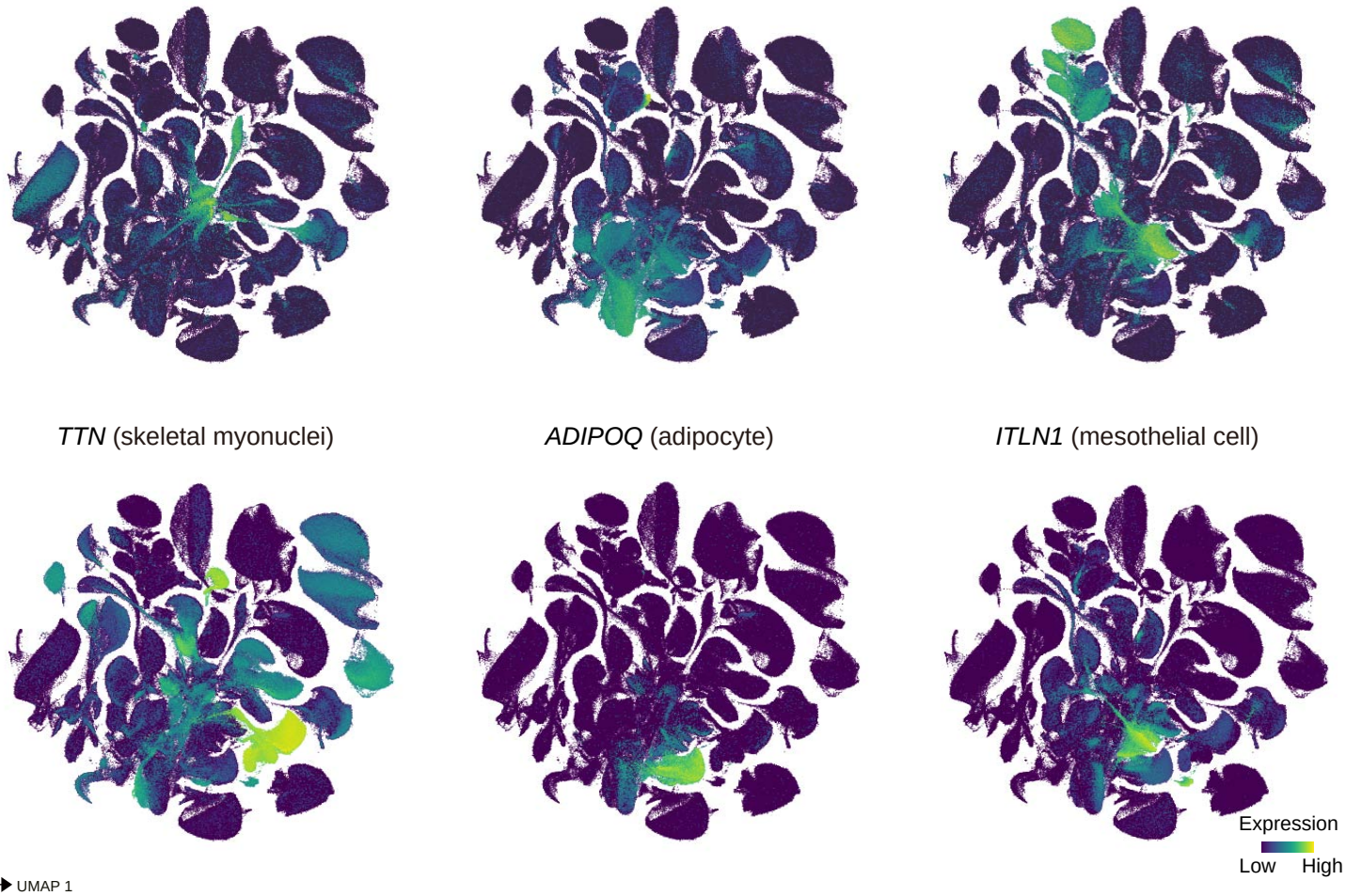


Extended Data Figure 13

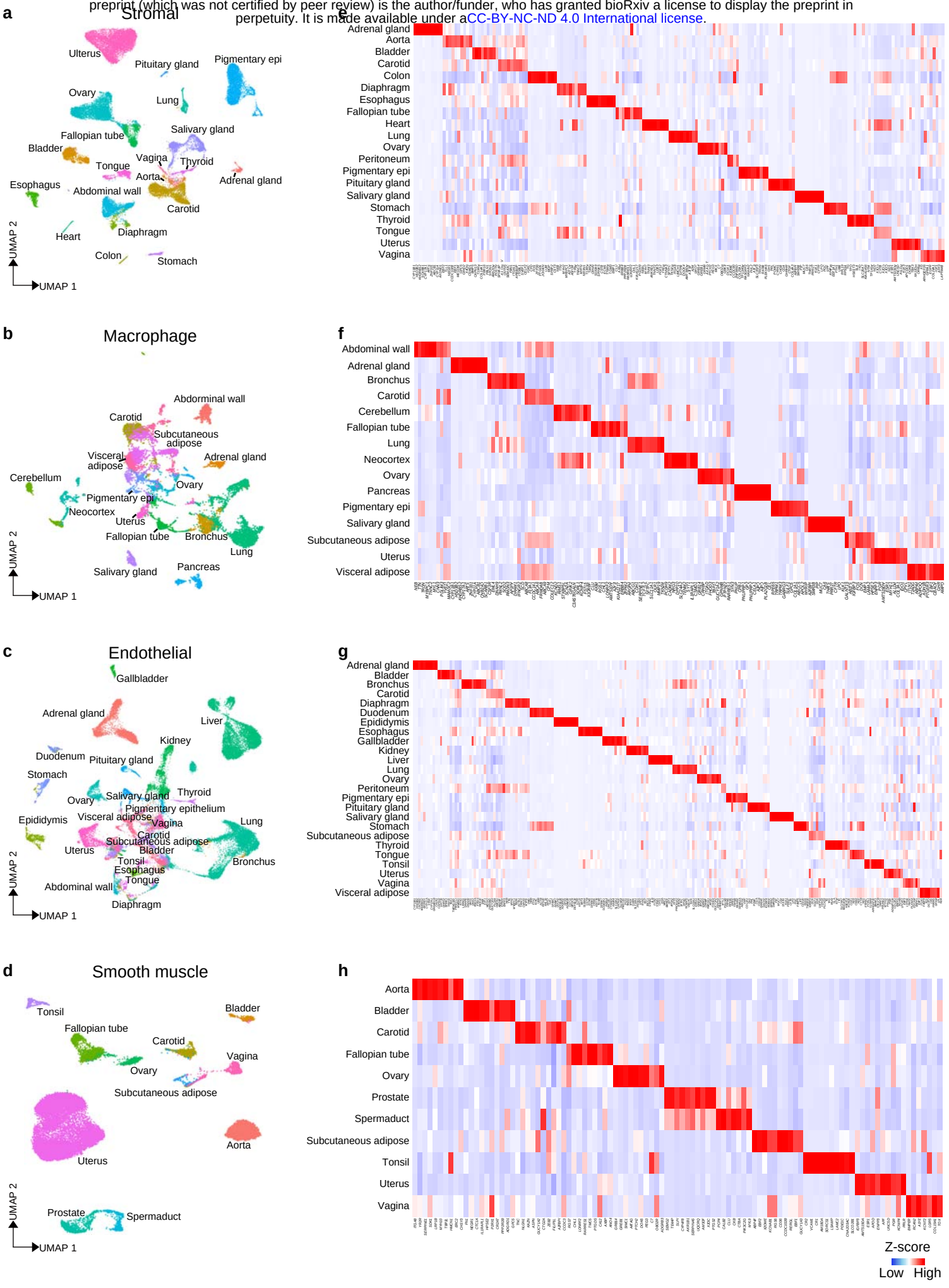
a



b

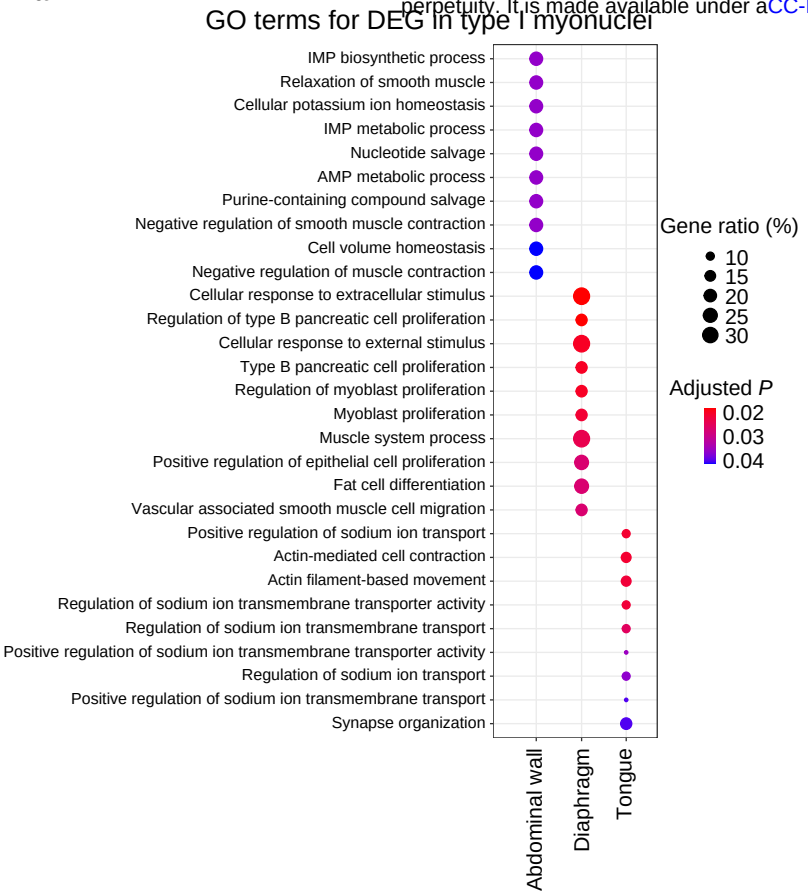


Extended Data Figure 14

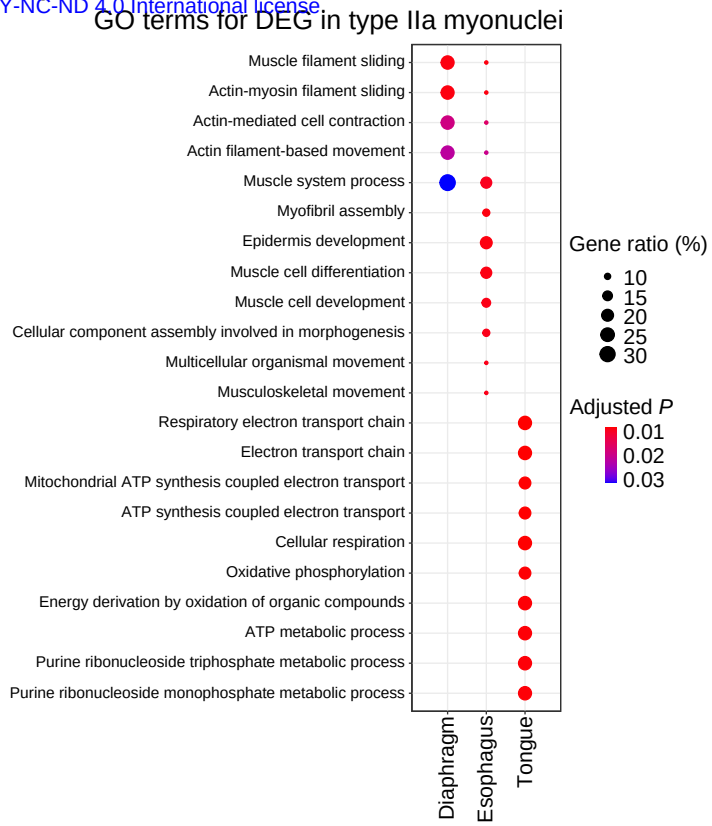


Extended Data Figure 15

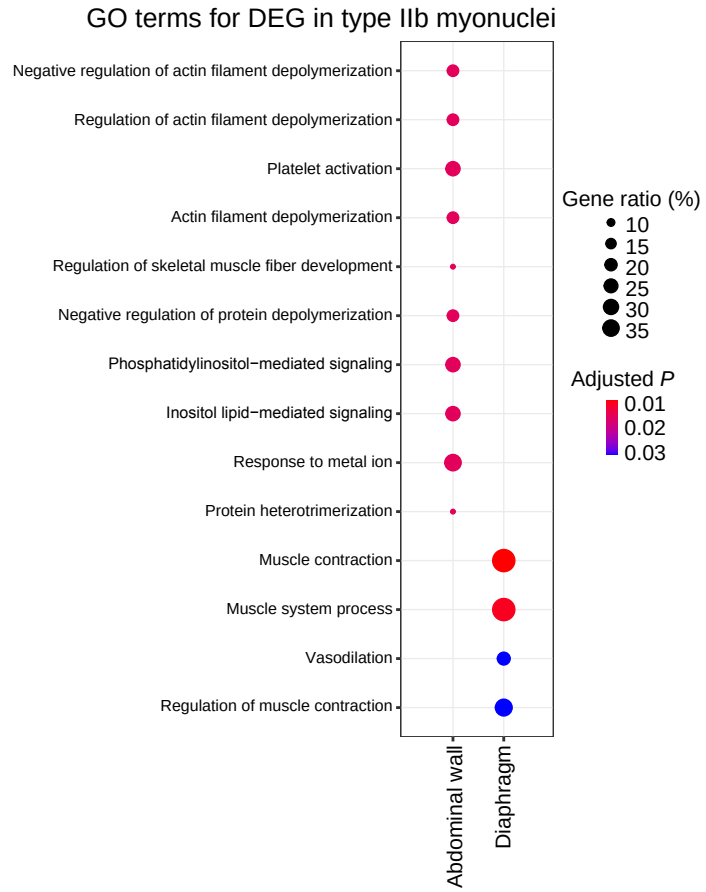
a



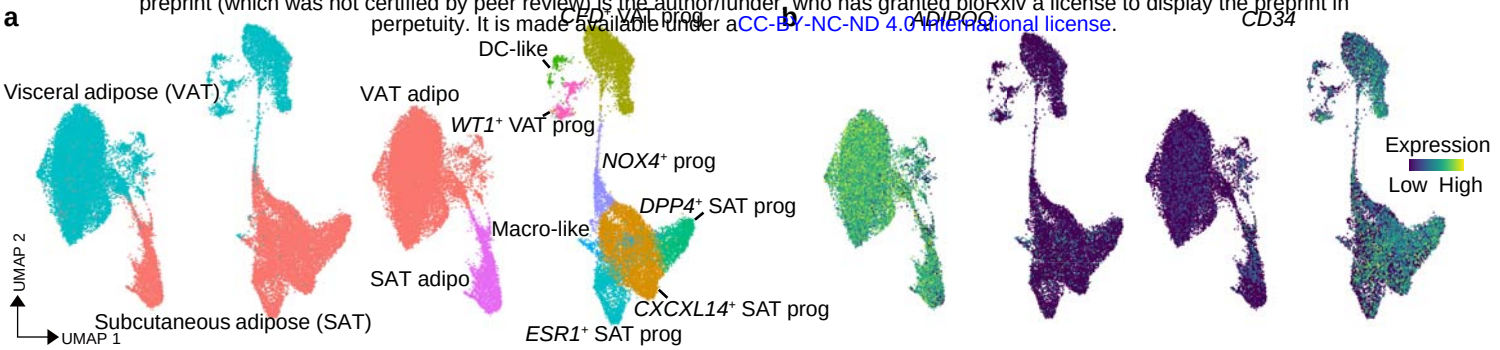
b

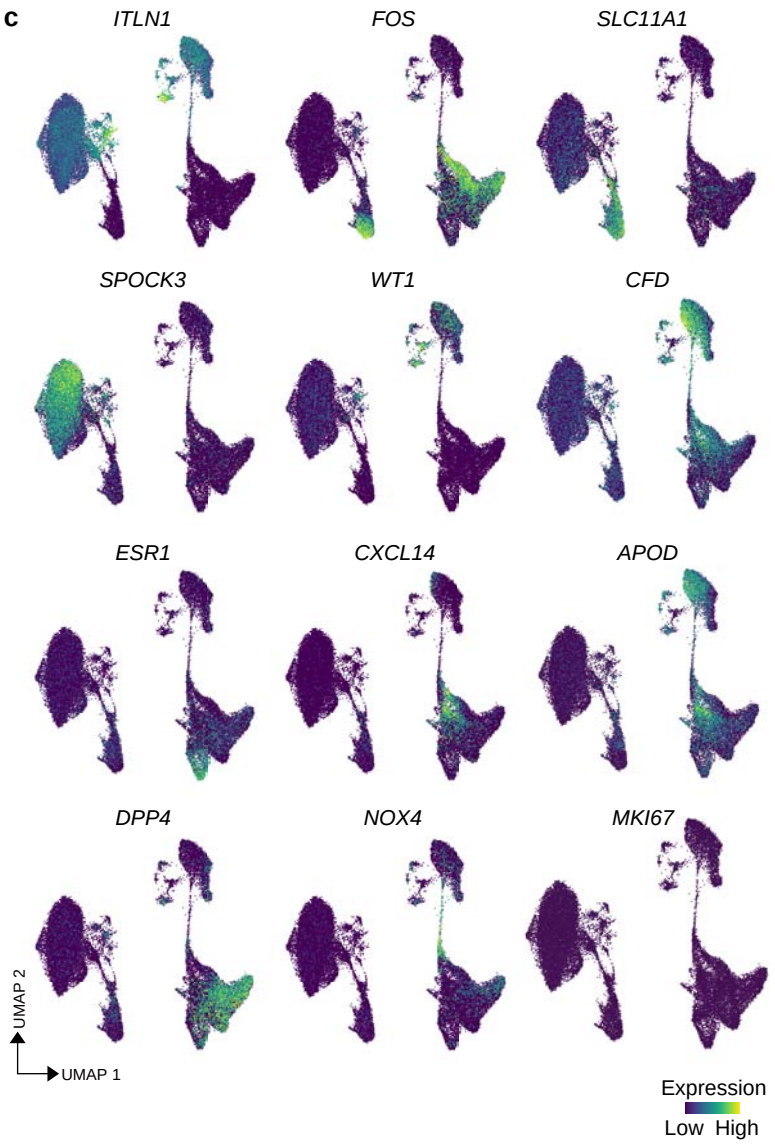


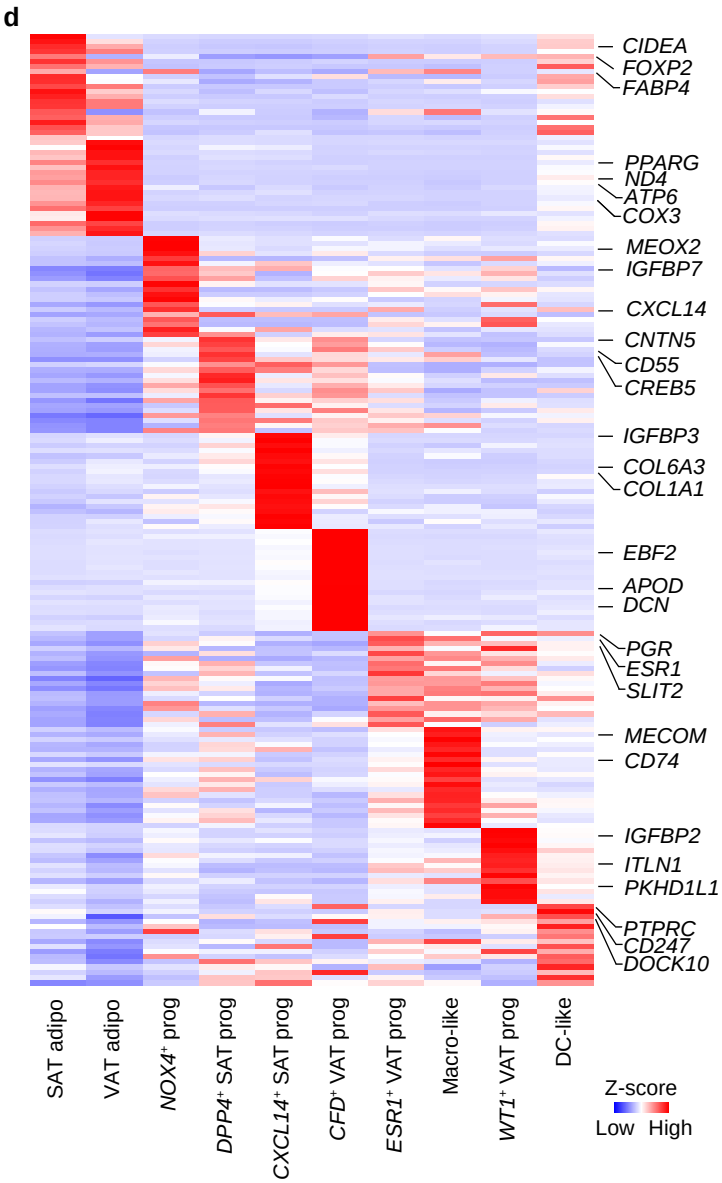
c



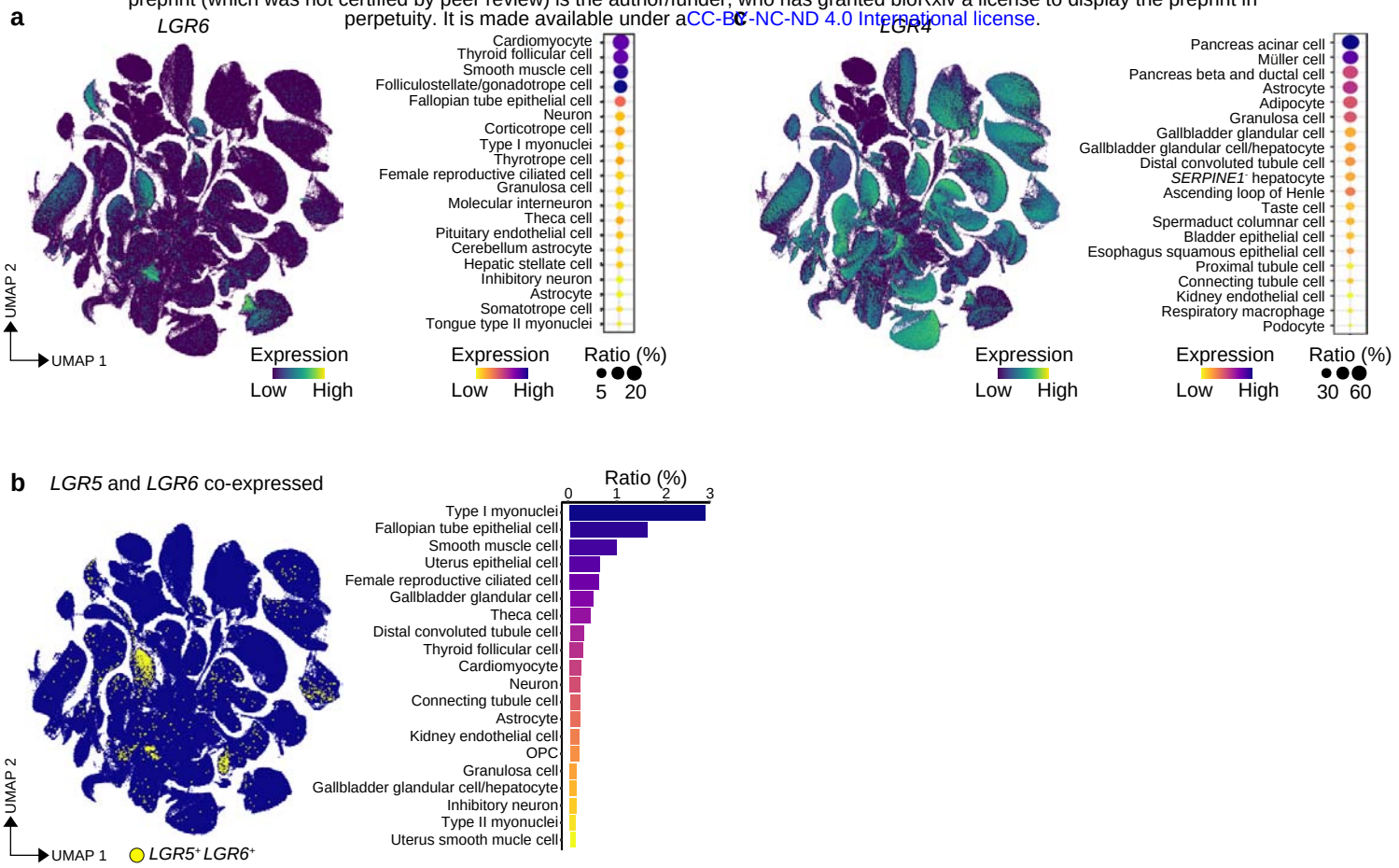
Extended Data Figure 16



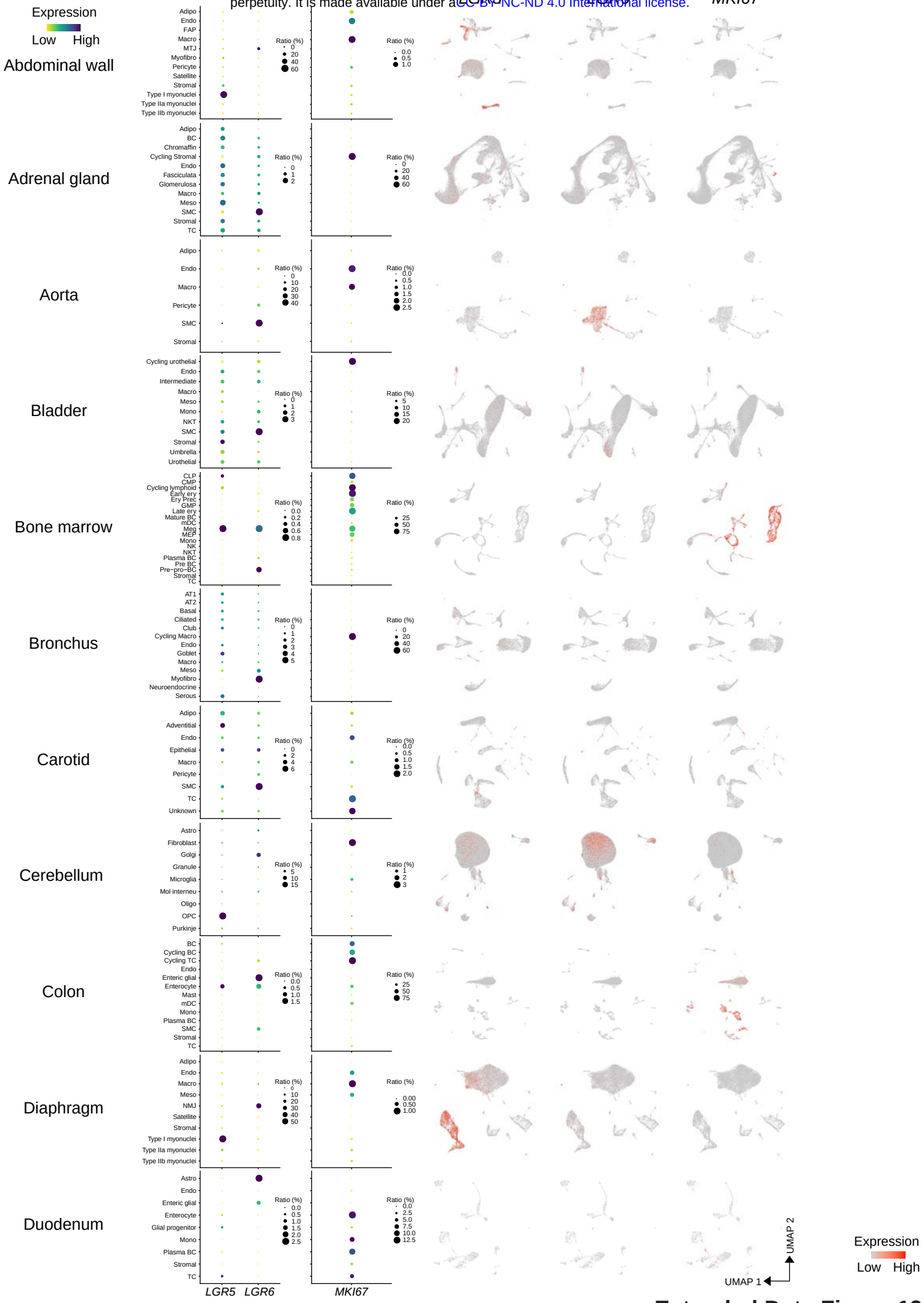




Extended Data Figure 17

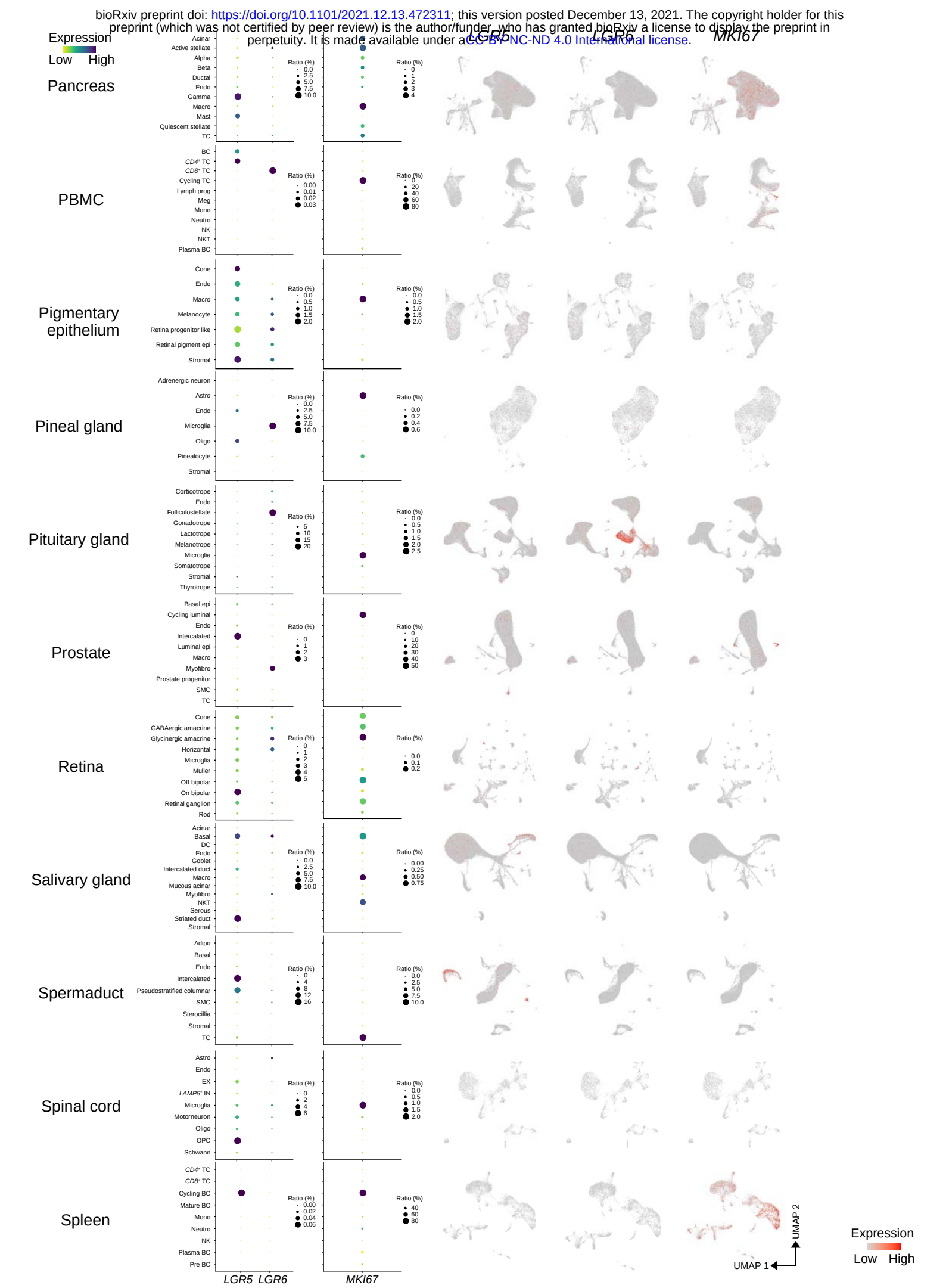


Extended Data Figure 18

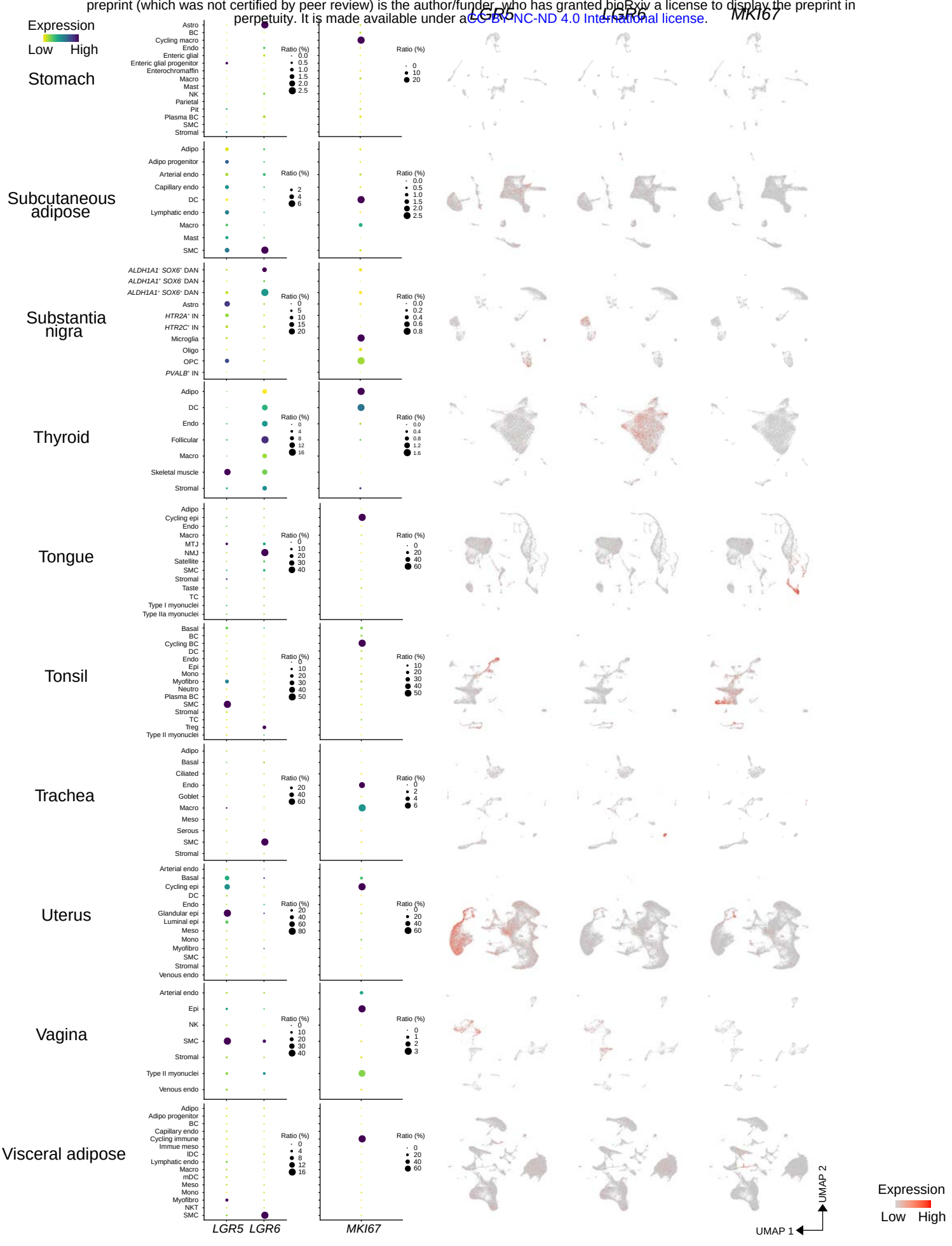


Extended Data Figure 19

Extended Data Figure 20

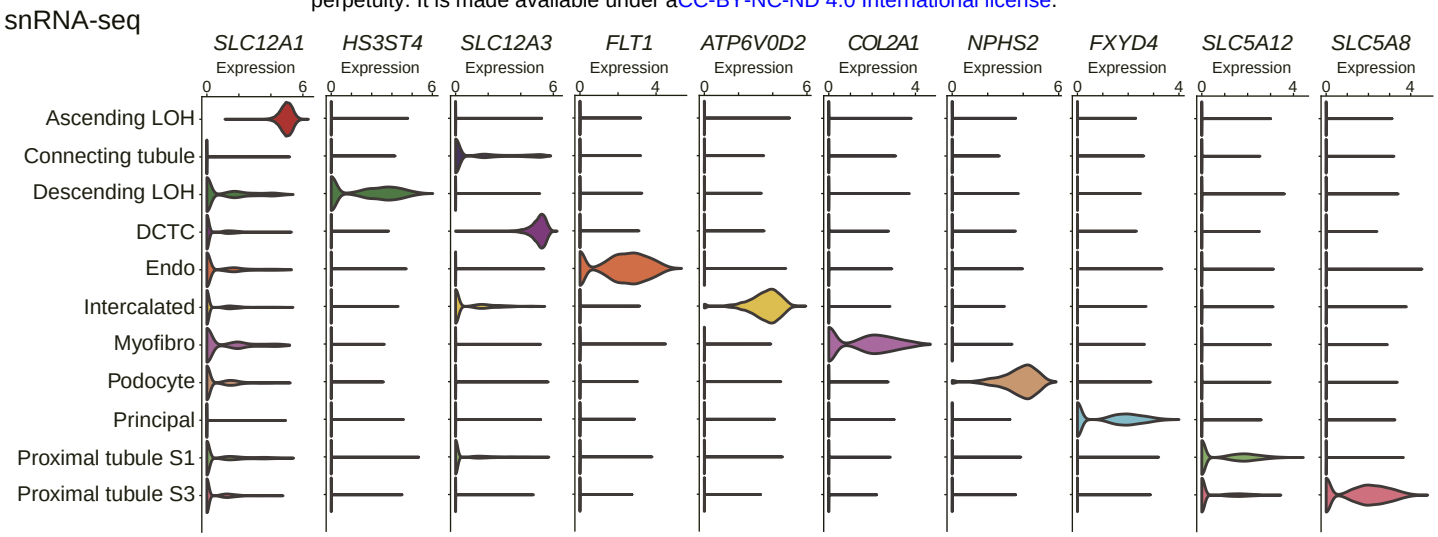


Extended Data Figure 21

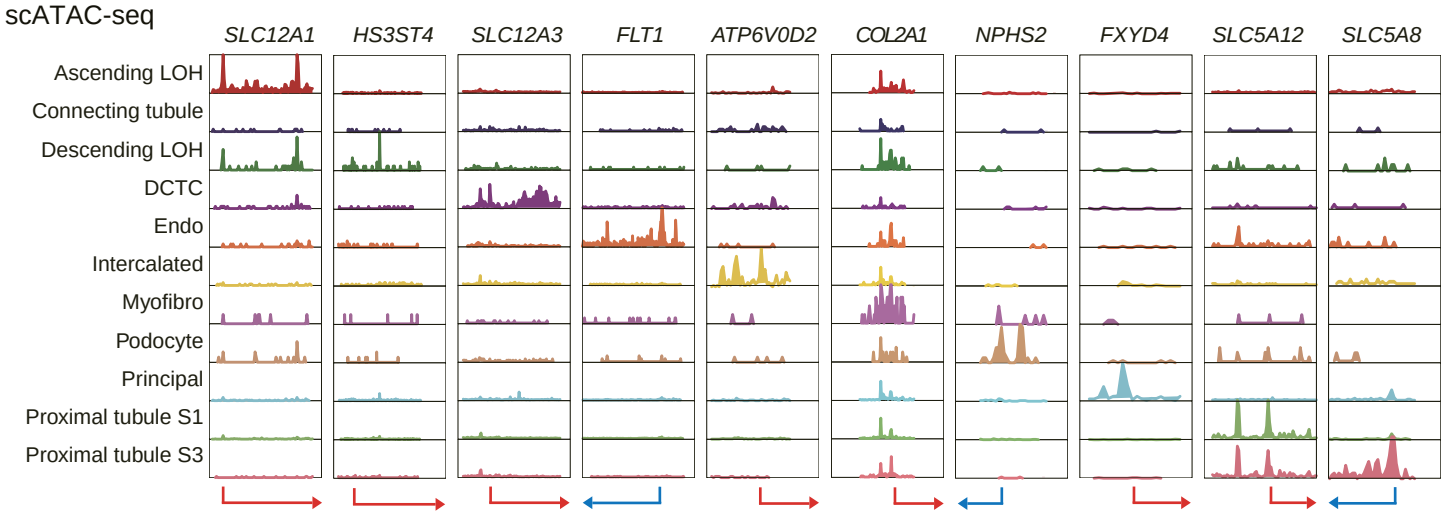


Extended Data Figure 22

a

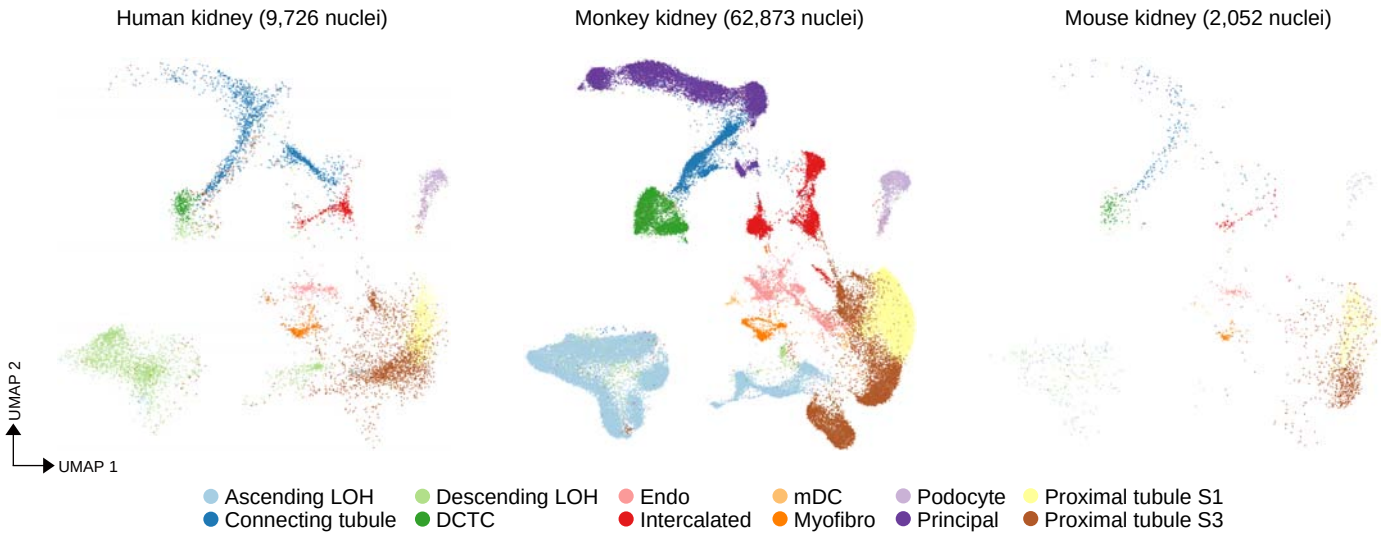


b

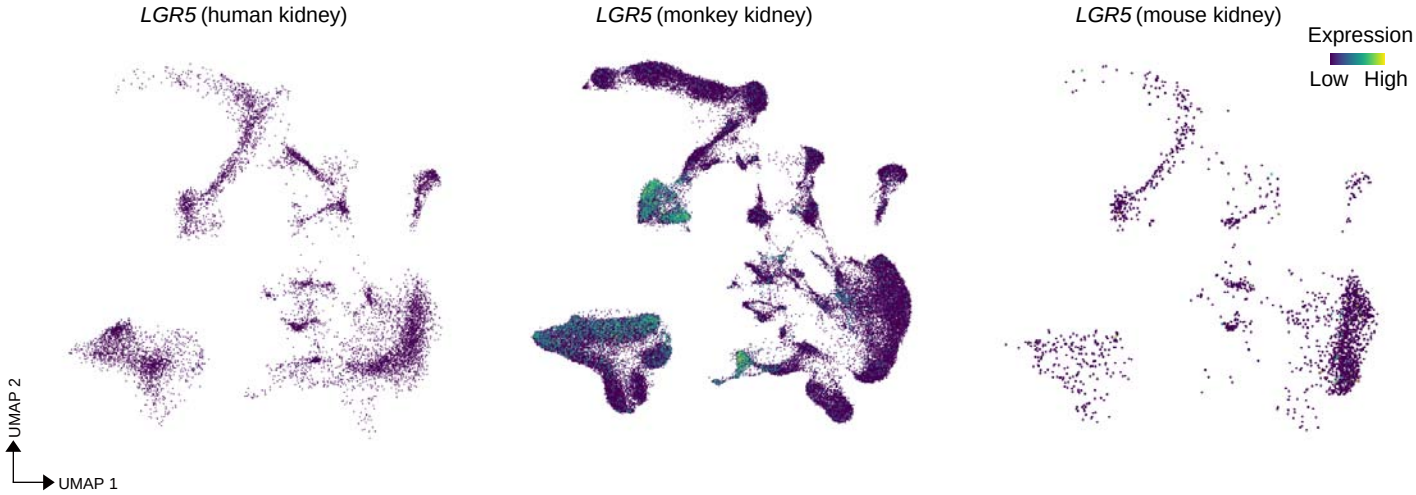


Extended Data Figure 23

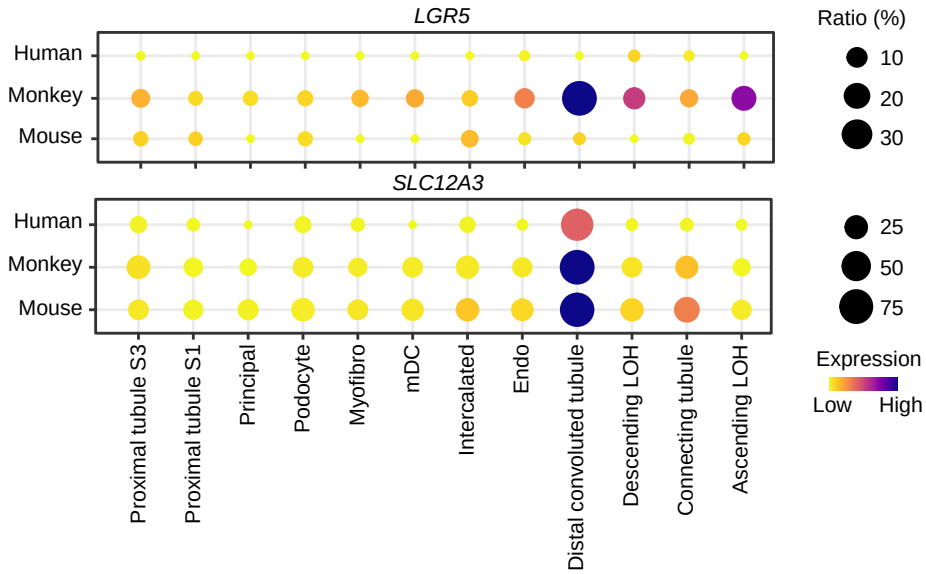
a



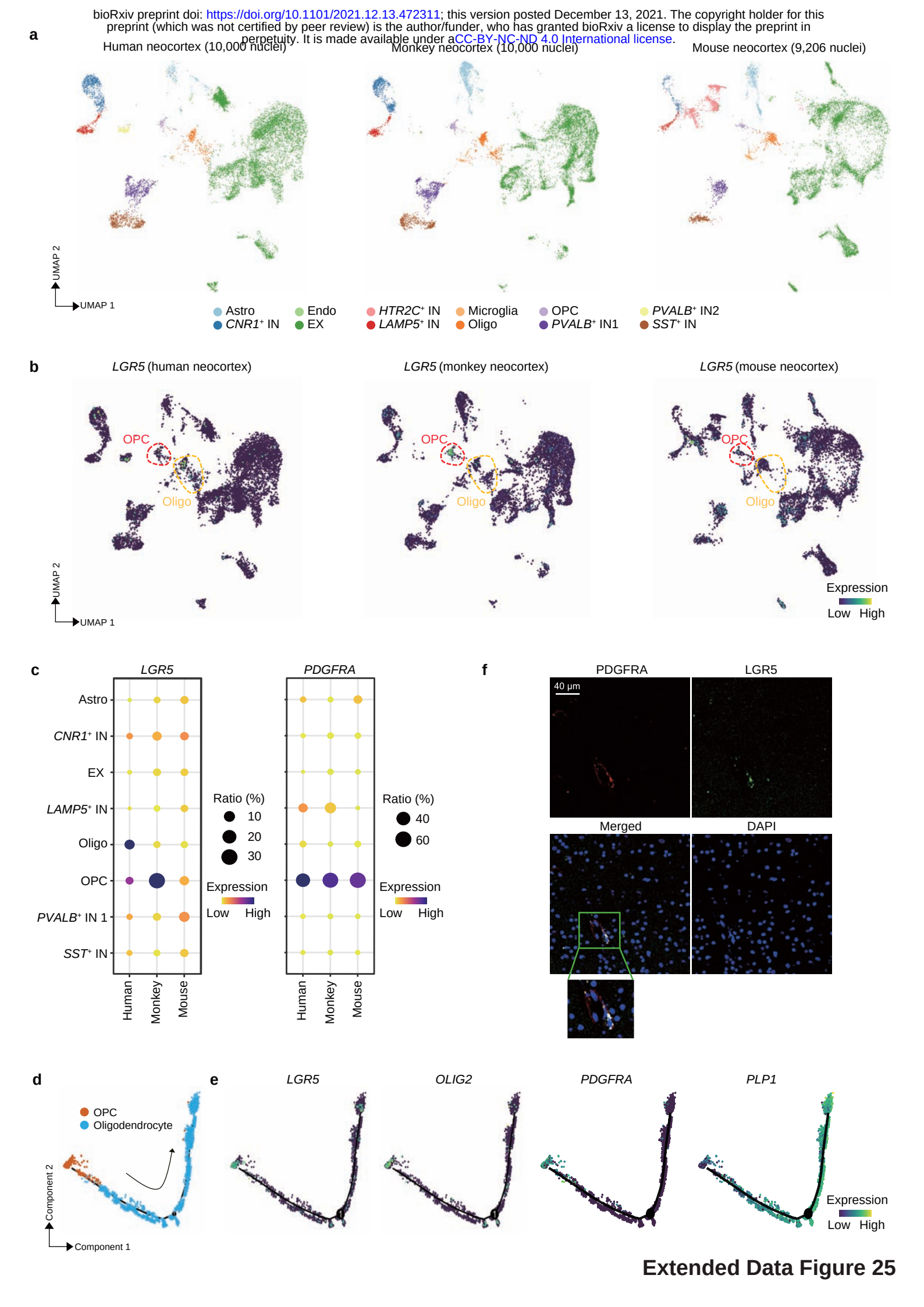
b



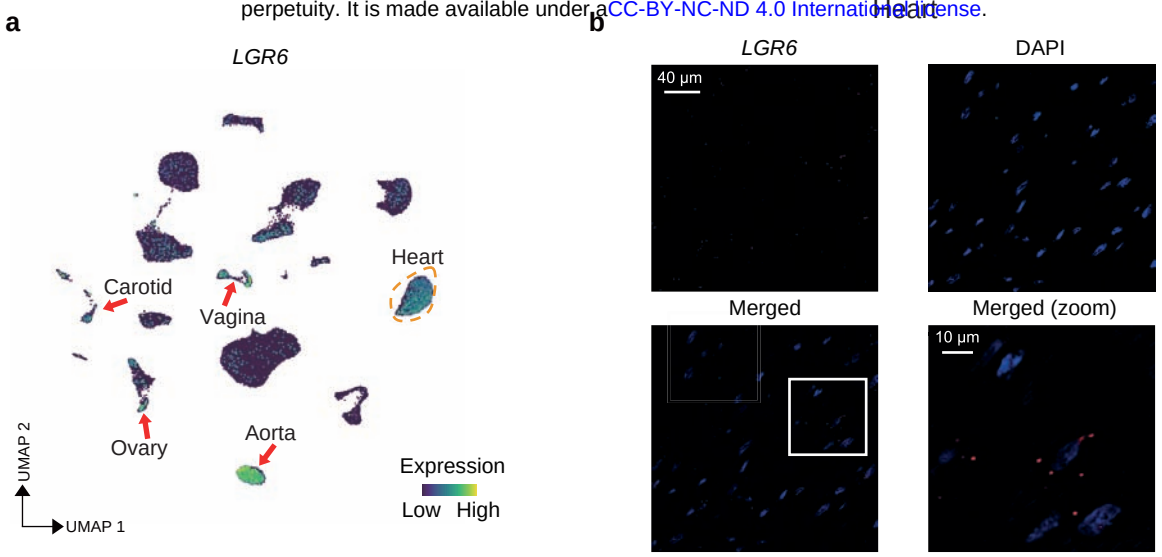
c

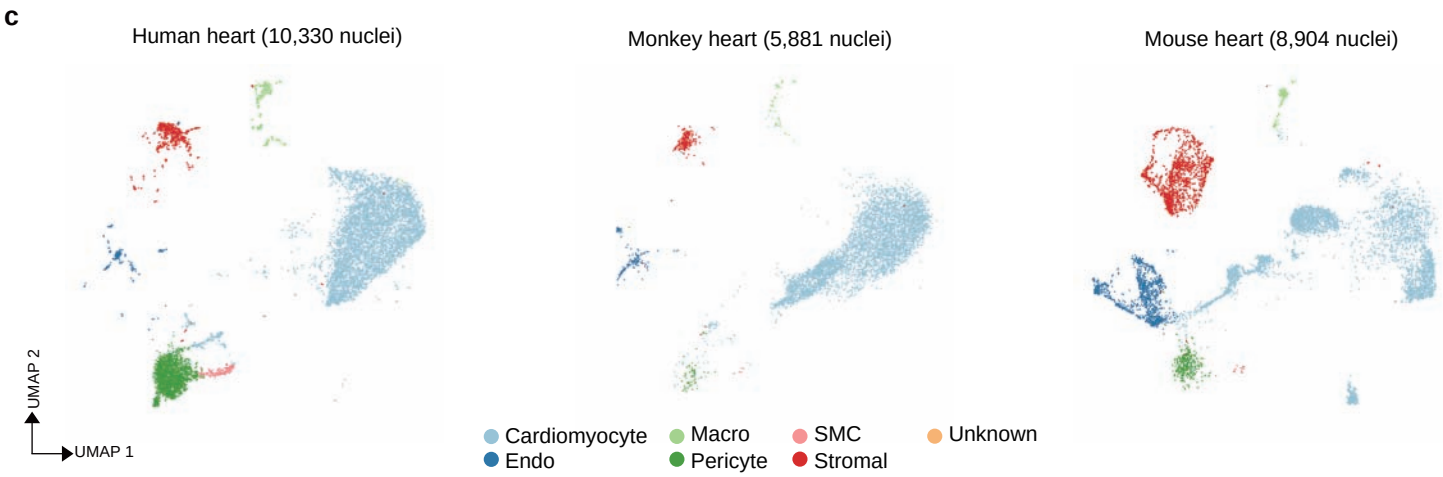


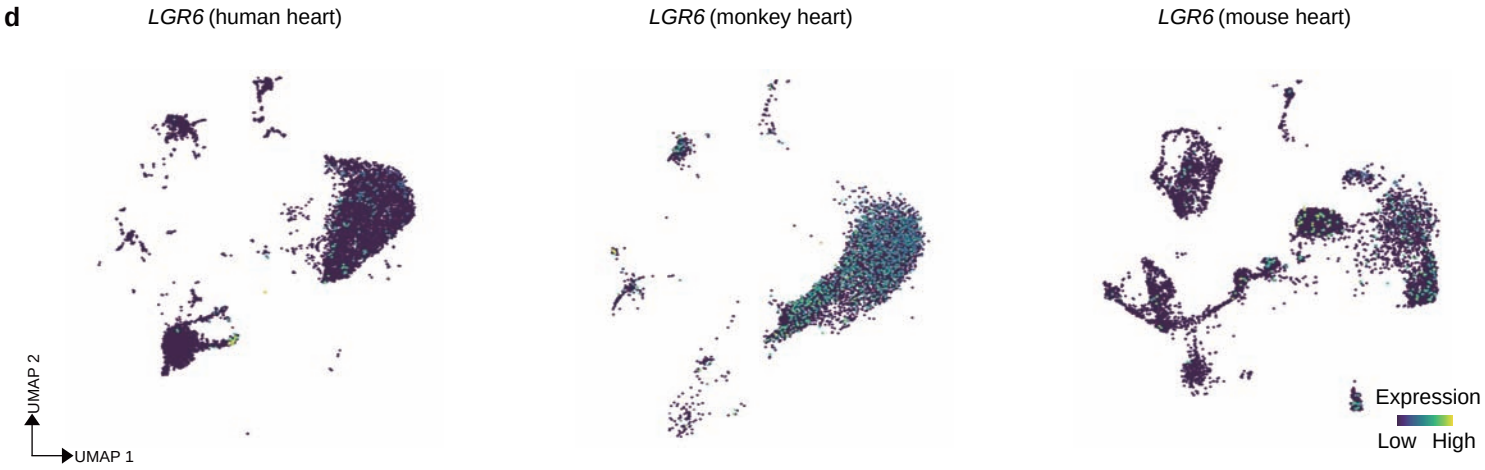
Extended Data Figure 24

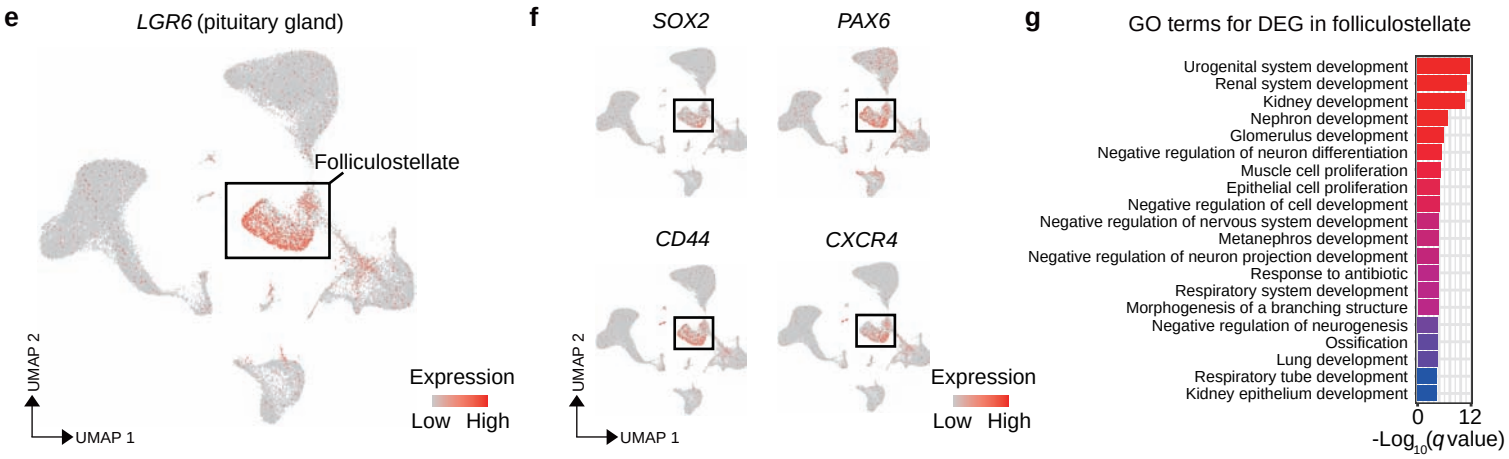


Extended Data Figure 25

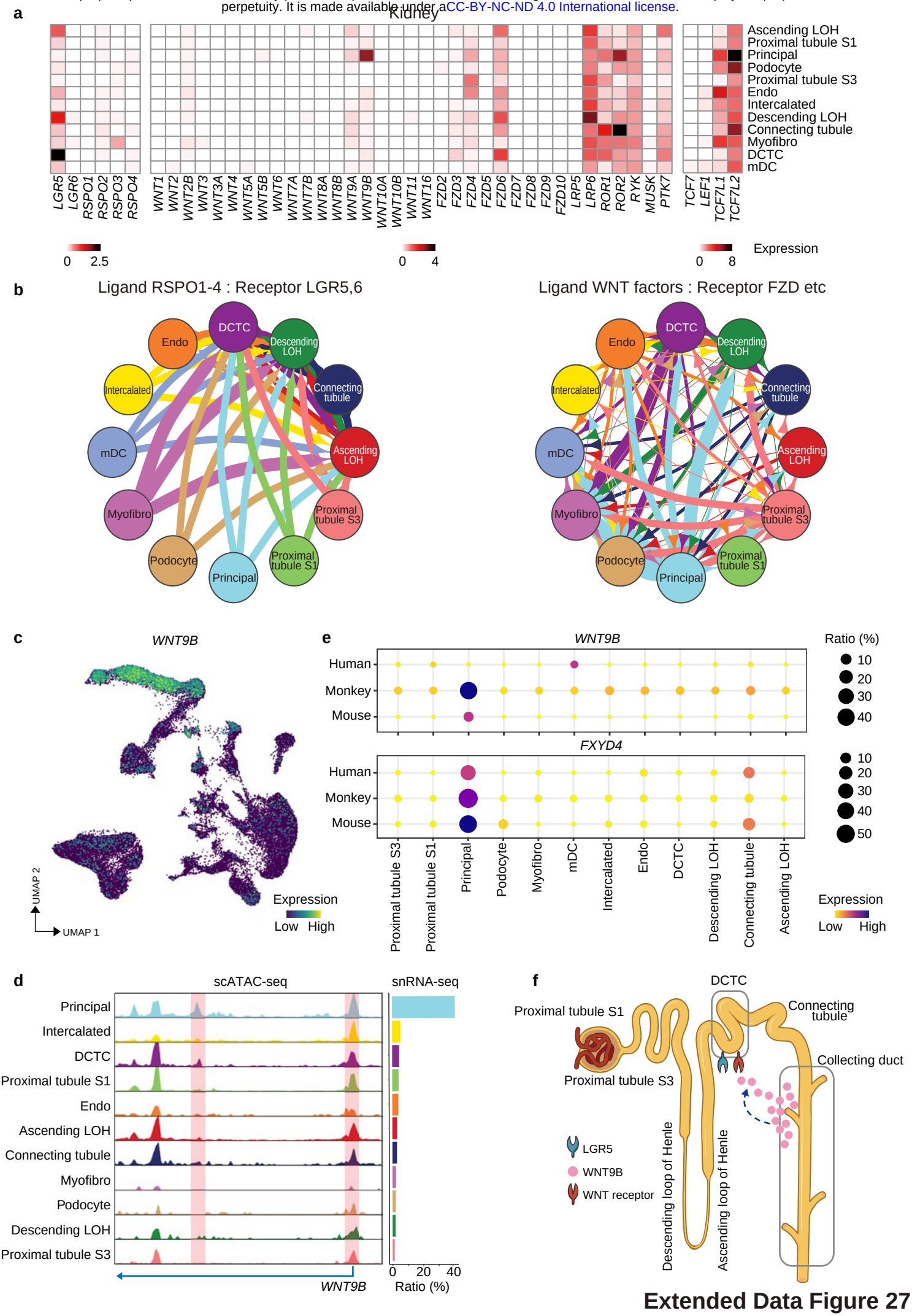






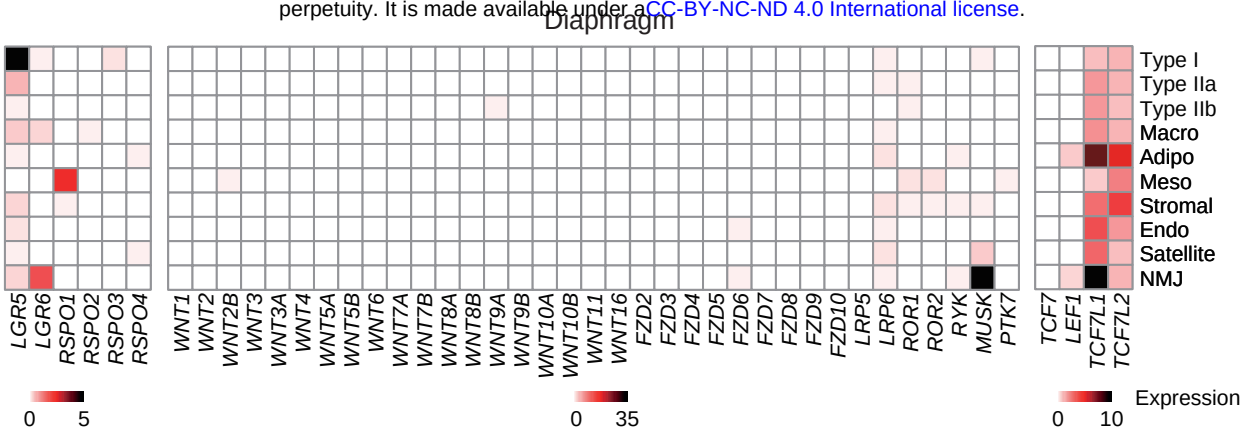


Extended Data Figure 26

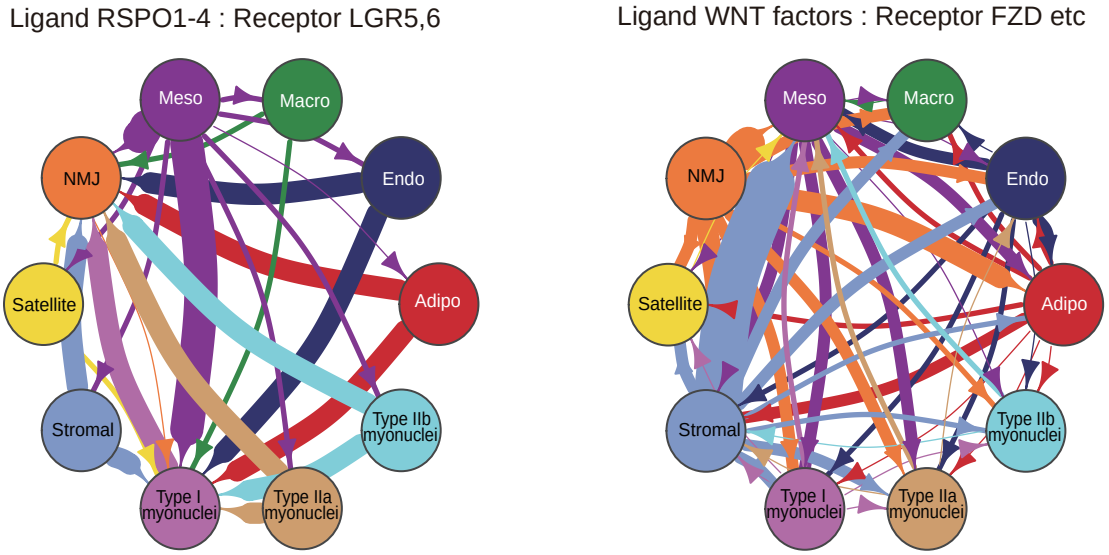


Extended Data Figure 27

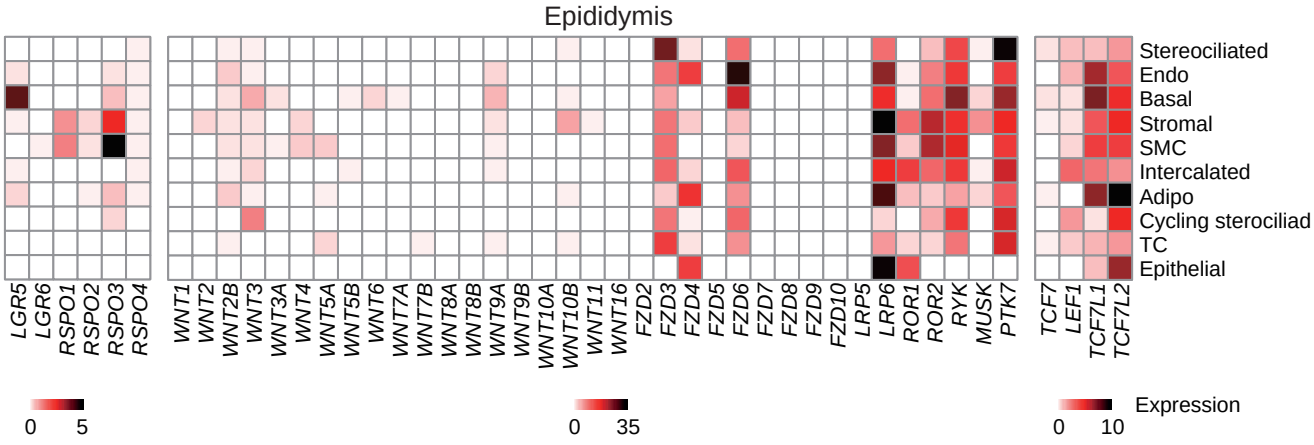
a



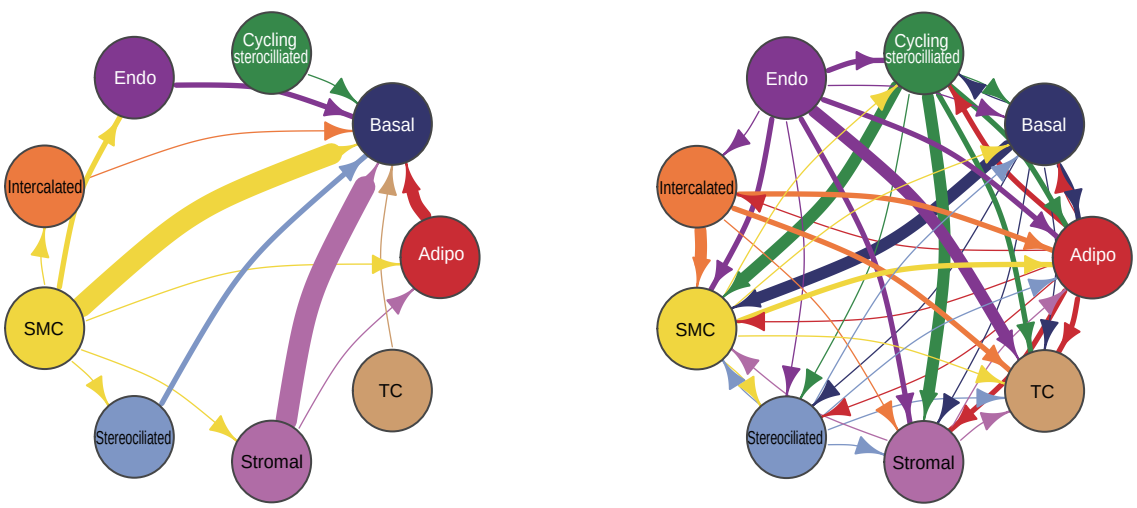
b



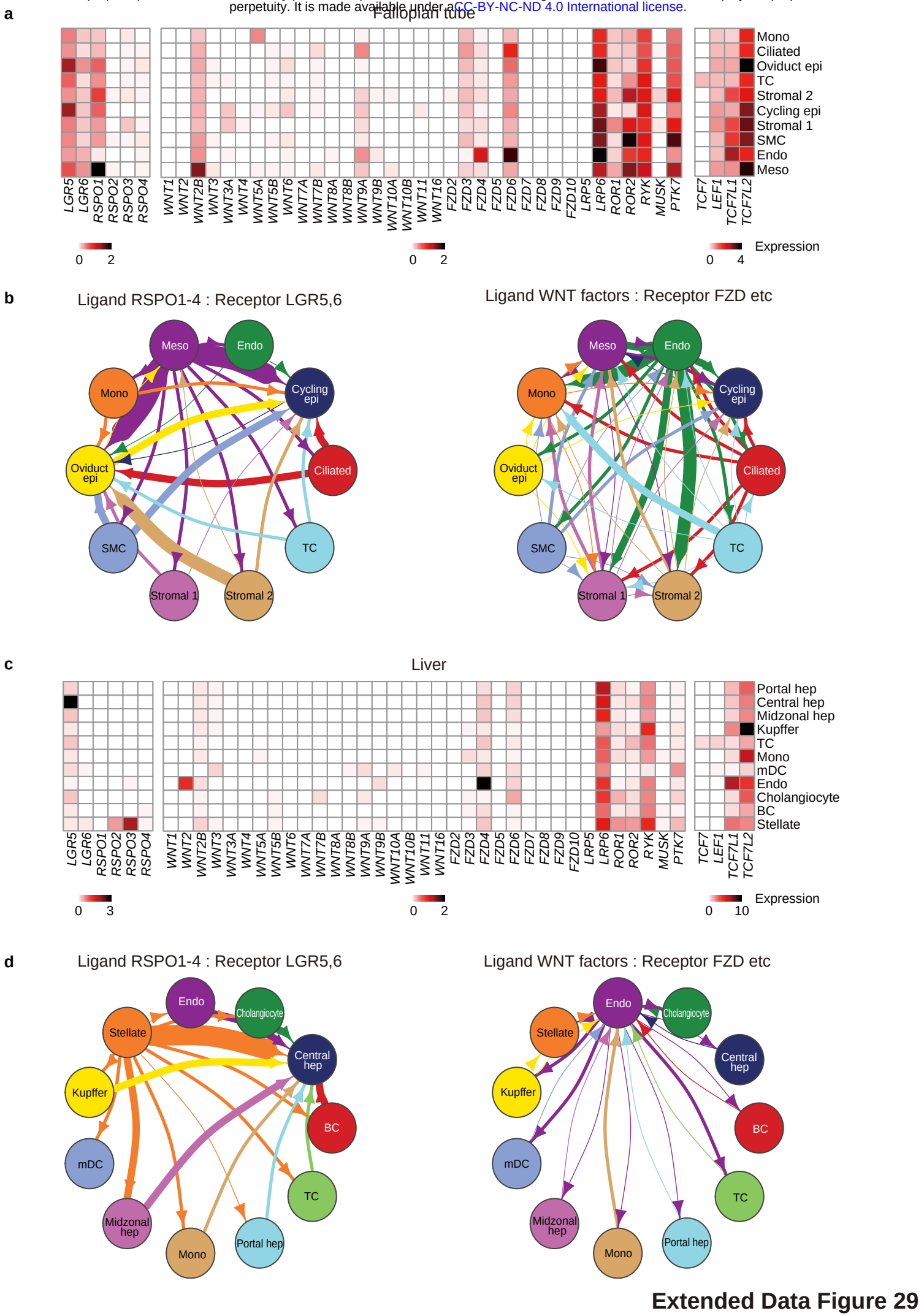
c



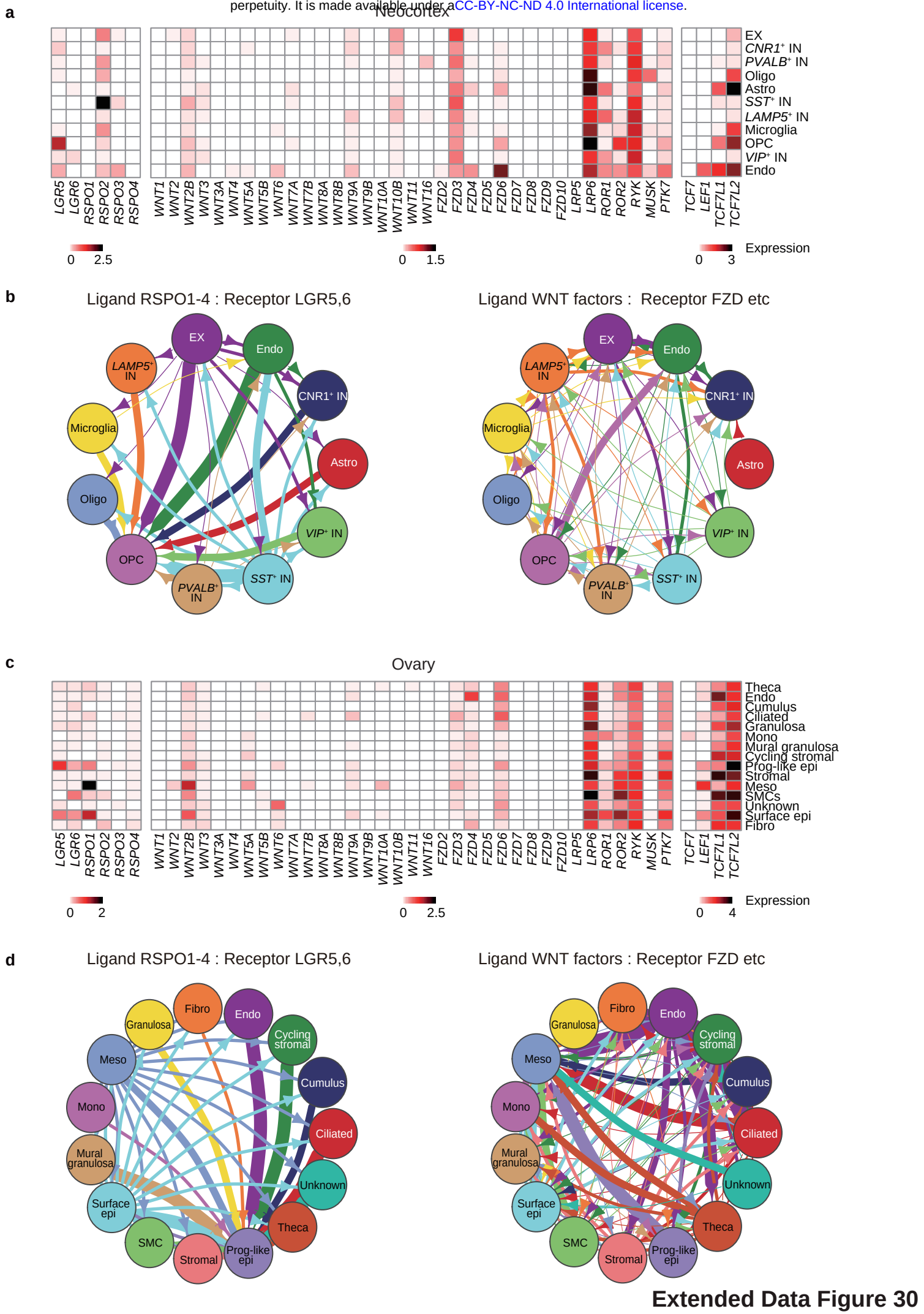
d



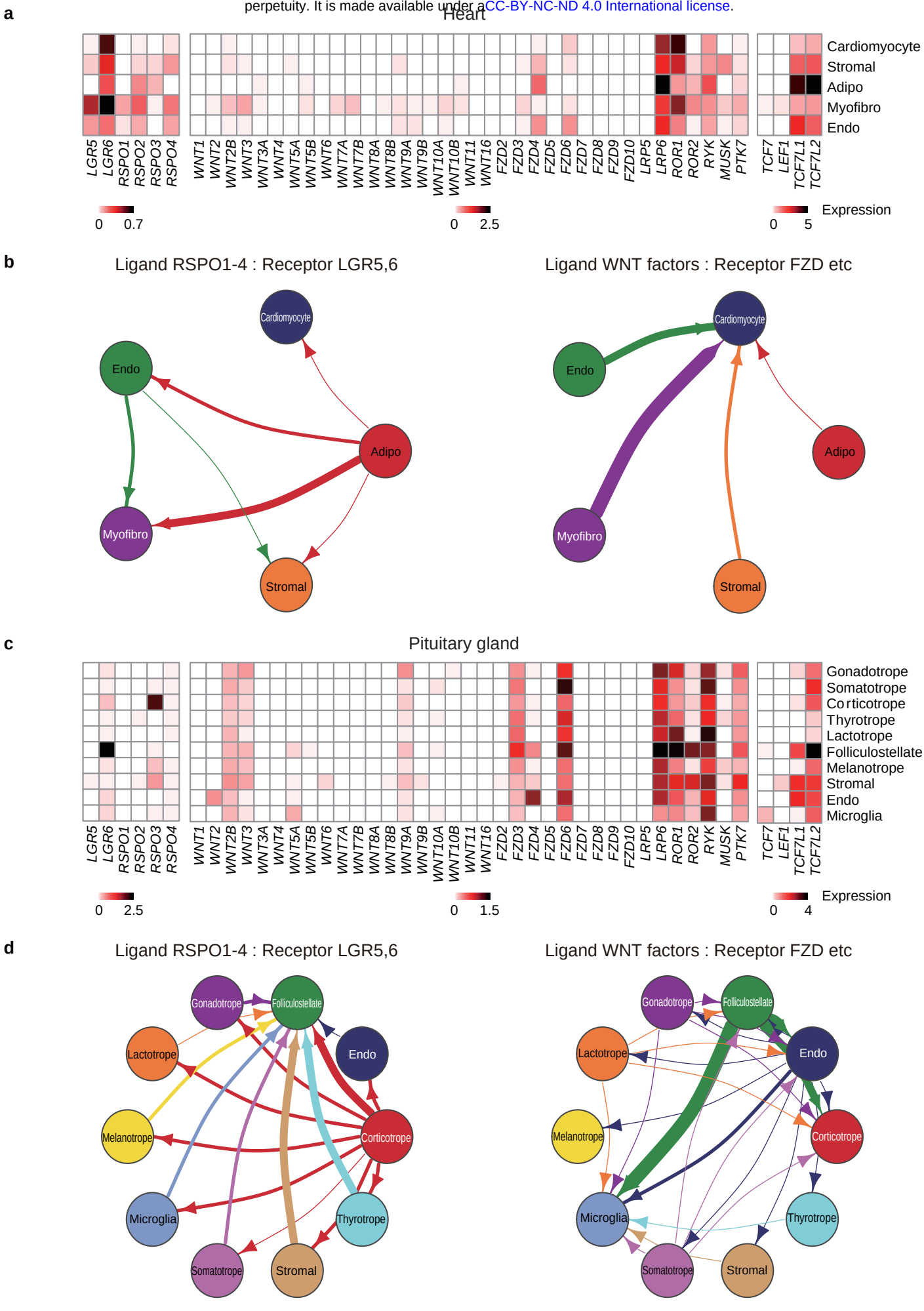
Extended Data Figure 28



Extended Data Figure 29

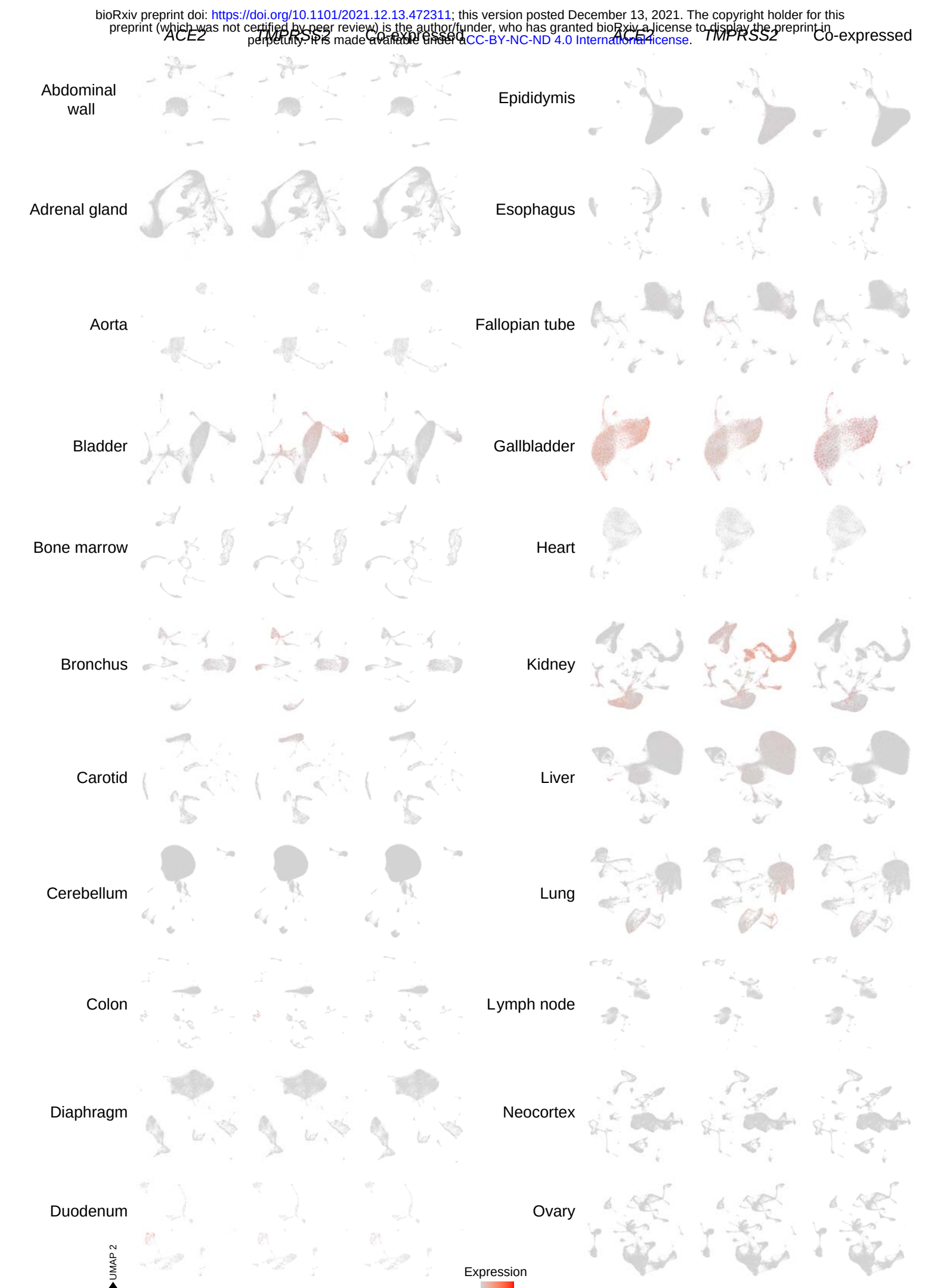


Extended Data Figure 30

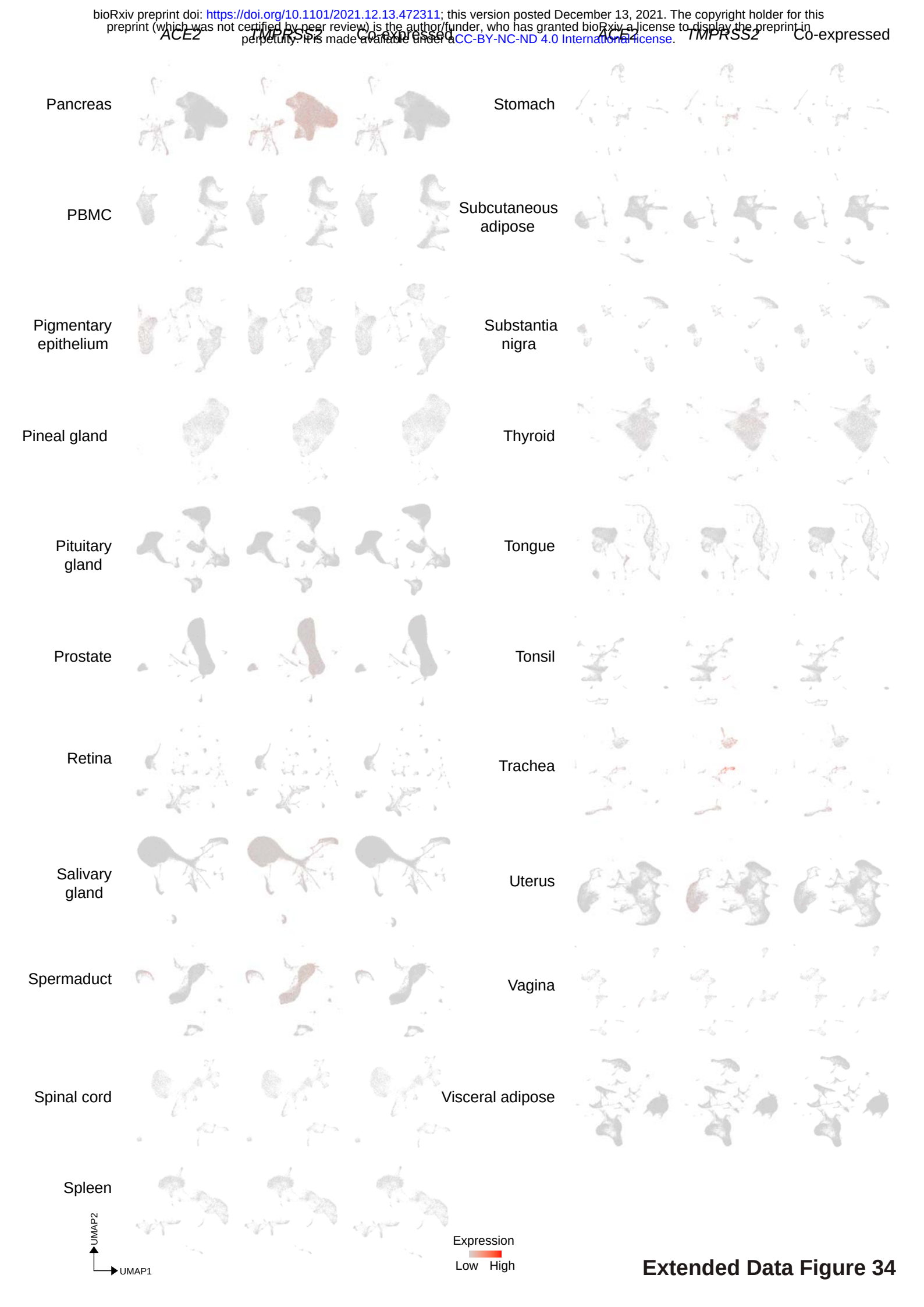


Extended Data Figure 31

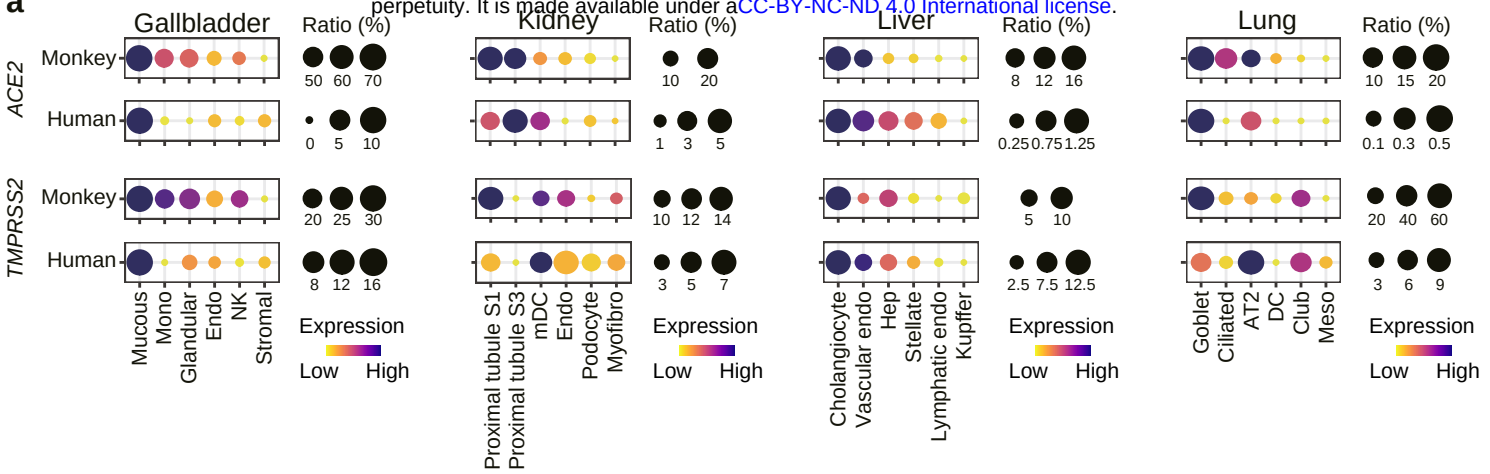
Extended Data Figure 32

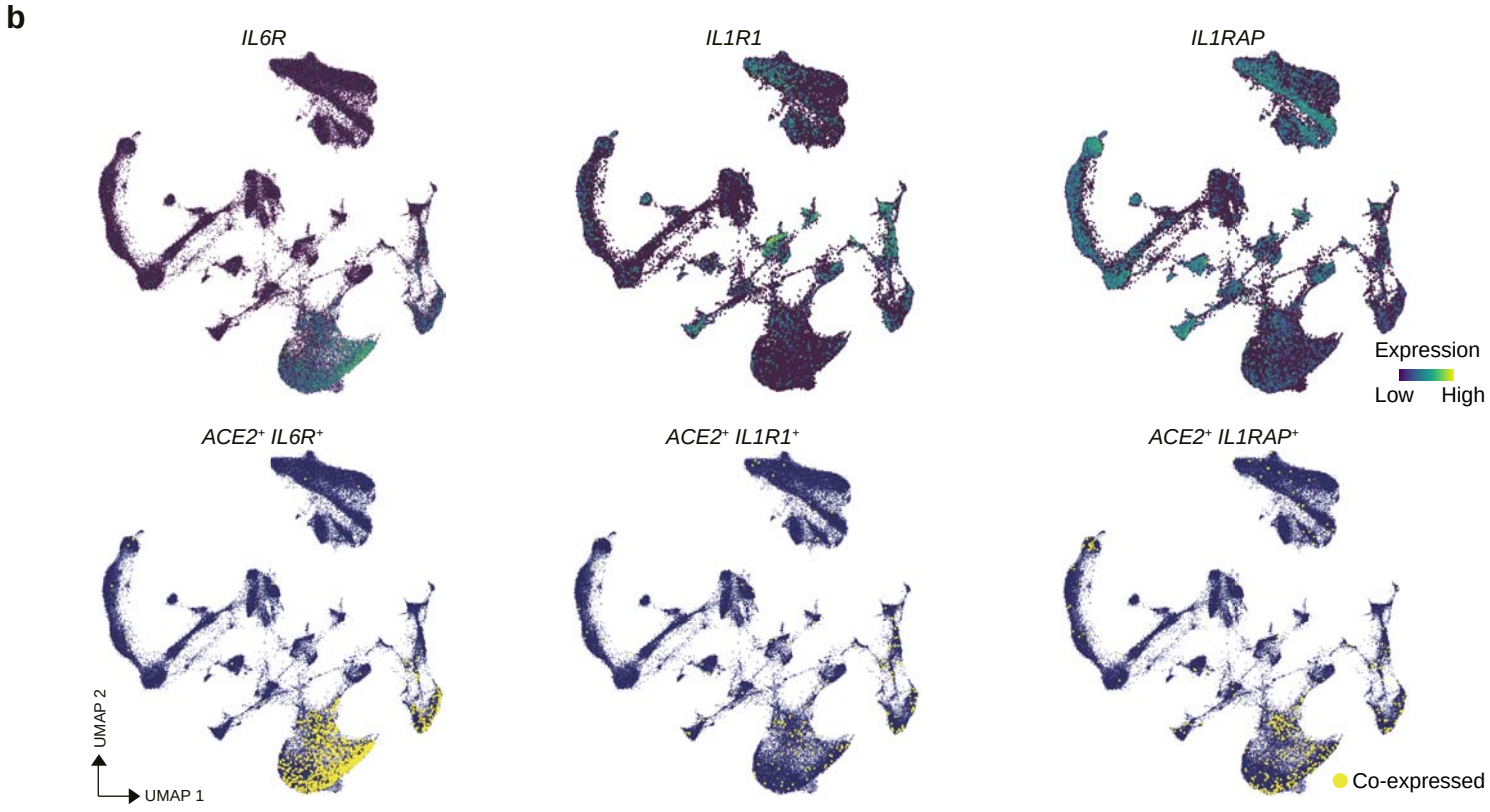


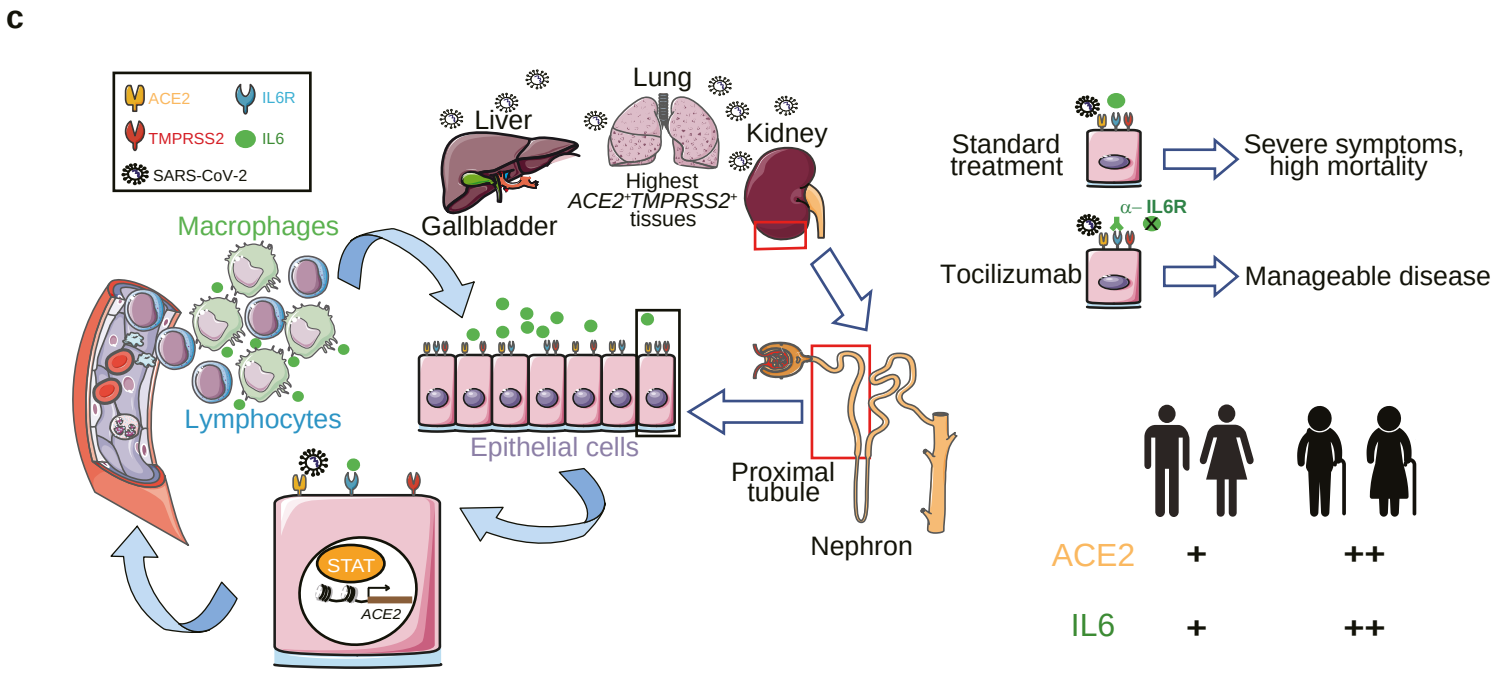
Extended Data Figure 33



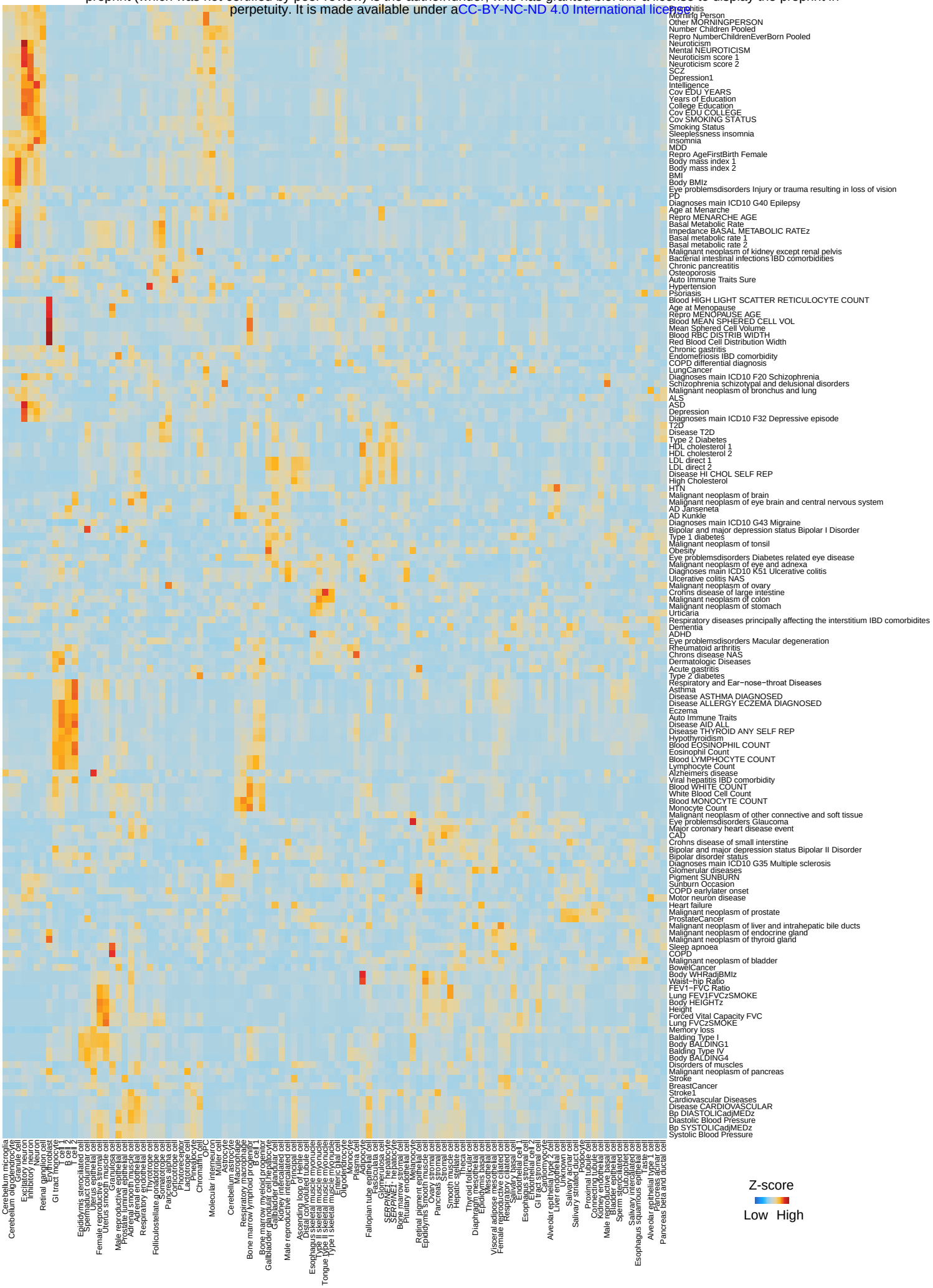
Extended Data Figure 34



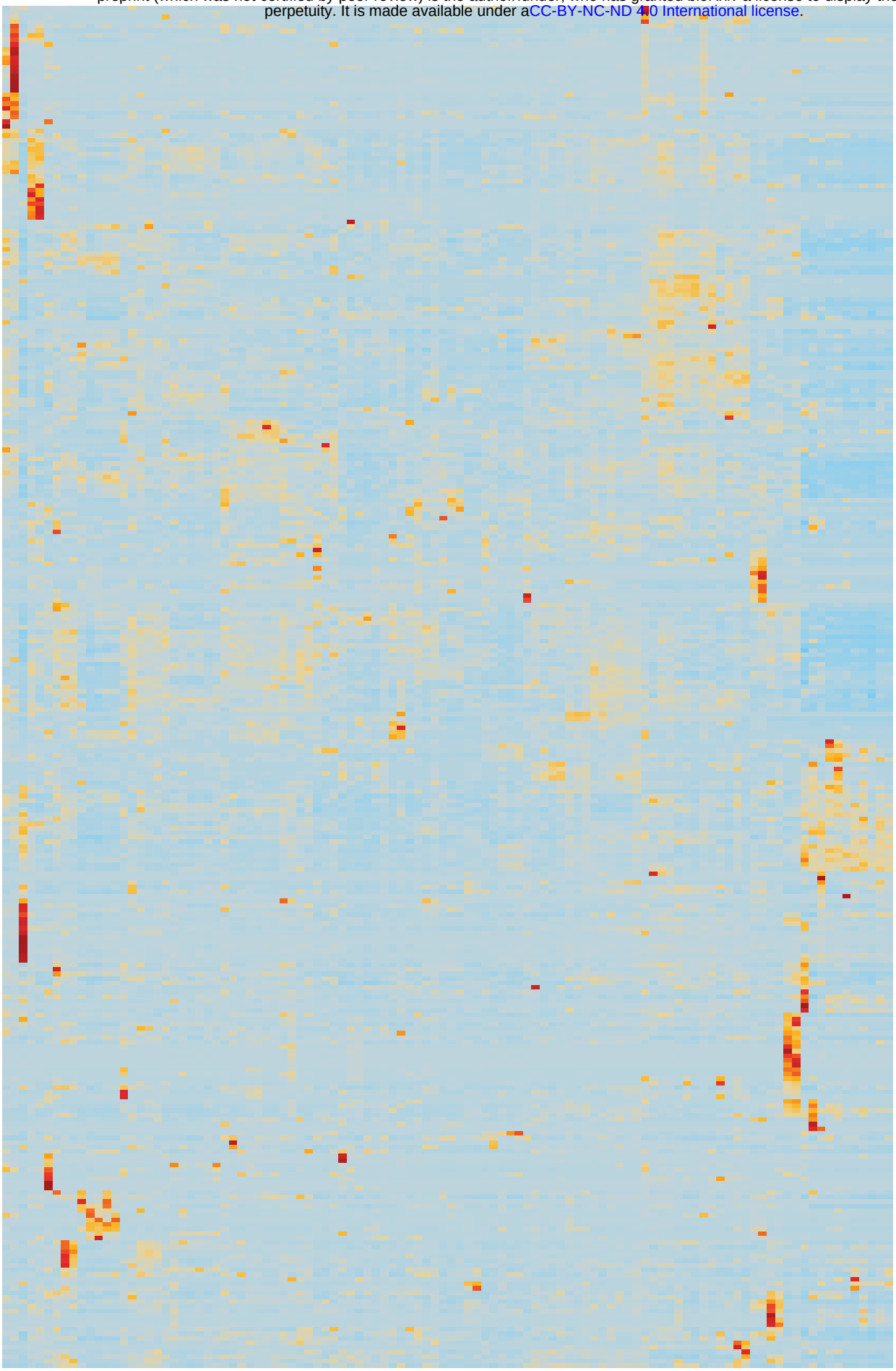




Extended Data Figure 35



Extended Data Figure 36



Extended Data Figure 37