Title (79 characters): The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans

Authors: The Tabula Sapiens Consortium

One Sentence Summary: We used single cell transcriptomics to create a molecularly defined phenotypic reference of human cell types which spans 24 human tissues and organs.

Abstract: Molecular characterization of cell types using single cell transcriptome sequencing is revolutionizing cell biology and enabling new insights into the physiology of human organs. We created a human reference atlas comprising nearly 500,000 cells from 24 different tissues and organs, many from the same donor. This atlas enabled molecular characterization of more than 400 cell types, their distribution across tissues and tissue specific variation in gene expression. Using multiple tissues from a single donor enabled identification of the clonal distribution of T cells between tissues, the tissue specific mutation rate in B cells, and analysis of the cell cycle state and proliferative potential of shared cell types across tissues. Cell type specific RNA splicing was discovered and analyzed across tissues within an individual.

Main Text

Introduction

Although the genome is often called the blueprint of an organism, it is perhaps more accurate to describe it as a parts list composed of the various genes which may or may not be used in the different cell types of a multicellular organism. While nearly every cell in the body has essentially the same genome, each cell type makes different use of that genome and expresses a subset of all possible genes (*1*). Therefore, the genome in and of itself does not provide an understanding of the molecular complexity of the various cell types of that organism. This has motivated efforts to characterize the molecular composition of various cell types within humans and multiple model organisms, both by transcriptional (*2*) and proteomic (*3*, *4*) approaches.

While such efforts are yielding insights (*5*–*7*), one caveat to current approaches is that individual organs are often collected at different locations, from different donors (*8*) and processed using different protocols, or lack replicate data (*9*). Controlled comparisons of cell types between different tissues and organs are especially difficult when donors differ in genetic background, age, environmental exposure, and epigenetic effects. To address this, we developed an approach to analyzing large numbers of organs from the same individual (*10*), which we originally used to characterize age-related changes in gene expression in various cell types in the mouse (*11*).

Data Collection and Cell Type Representation

We collected multiple tissues from individual human donors (designated TSP 1-15) and performed coordinated single cell transcriptome analysis on live cells (*12*)**.** We collected 17 tissues from one donor, 14 tissues from a second donor, and 5 tissues from two

other donors (**Fig. 1**). We also collected smaller numbers of tissues from a further 11 donors, creating biological replicates for nearly all tissues. The donors comprise a range of ethnicities, are balanced by gender, have a mean age of 51 years and a variety of medical backgrounds (**table S1**). Single cell transcriptome sequencing was performed with both FACS sorted cells in well plates with smartseq2 amplification as well as 10x microfluidic droplet capture and amplification for each tissue (**fig. S1**). Tissue experts used a defined cell ontology terminology to annotate cell types consistently across the different tissues (*13*), leading to a total of 475 distinct cell types with reference transcriptome profiles (**tables S2, S3**). The full dataset can be explored online with the cellxgene tool via the Tabula Sapiens data portal (*14*)**.**

Data was collected for bladder, blood, bone marrow, eye, fat, heart, kidney, large intestine, liver, lung, lymph node, mammary, muscle, pancreas, prostate, salivary gland, skin, small intestine, spleen, thymus, tongue, trachea, uterus and vasculature. Fifty-nine separate specimens in total were collected, processed, and analyzed, and 481,120 cells passed QC filtering (**figs. S2-S7** and **table S2**). On a per compartment basis, the dataset includes 264,009 immune cells, 102,580 epithelial cells, 32,701 endothelial cells and 81,529 stromal cells. Working with live cells as opposed to isolated nuclei ensured that the dataset includes all mRNA transcripts within the cell, including transcripts that have been processed by the cell's splicing machinery, thereby enabling insight into variation in alternative splicing.

To characterize the relationship between transcriptome data and conventional histologic analysis, a team of pathologists analyzed H&E stained sections prepared from 9 tissues from donor TSP2 and 13 tissues from donor TSP14 (*14*). Cells were identified by morphology and classified broadly into epithelial, endothelial, immune and stromal compartments, as well as rarely detected peripheral nervous system (PNS) cell types. (**Fig. 2A**). These classifications were used to estimate the relative abundances of cell types across four compartments, as well as to the uncertainties in these abundances due to spatial heterogeneity of each tissue type (**Fig. 2B, fig. S8**)**.** We compared the histologically determined abundances with those obtained by single cell sequencing (**fig. S9**). Although, as expected, there can be substantial variation between the abundances determined by these methods, in aggregate we observe broad concordance over a large range of tissues and relative abundances. This approach enables an estimate of true cell type proportions since not every cell type survives dissociation with equal efficiency (*15*). For several of the tissues we also performed literature searches and collected tables of prior knowledge of cell type identity and abundance within those tissues (**table S4**). We compared literature values with our experimentally observed frequencies for three well annotated tissues: lung, muscle and bladder (**fig. S10**).

Immune Cells: Variation in Gene Expression Across Tissues and a Shared Lineage History

The Tabula Sapiens can be used to study differences in the gene expression programs and lineage histories of cell types that are shared across tissues. We analyzed tissue

differences in the 36,475 macrophages distributed amongst 20 tissues, as tissue-resident macrophages are known to carry out specialized functions (*16*). These shared and orthogonal signatures are summarized in a correlation map (**fig. S11A**). For example, macrophages in the spleen were different from most other macrophages, and this was driven largely by higher expression of CD5L, a regulator of lipid synthesis (**fig. S11B**). We also observed a shared signature of elevated EREG (epiregulin) expression in solid tissues such as the skin, uterus and mammary compared to circulatory tissues (**fig. S11B**).

We characterized lineage relationships between T cells by assembling the T cell receptor sequences from from donor TSP2. Multiple T cell lineages were distributed across various tissues in the body, and we mapped their relationships (**Fig. 3A**). Large clones often reside in multiple organs, and several clones of mucosal associated invariant T cells are shared across donors (**fig. S11C**); these cells had characteristic expression of *TRAV1-2* as they are thought to be innate-like effector cells (*17*). Lineage information can also reveal tissue-specific somatic hyper-mutation rates in B cells. We assembled the B Cell Receptor sequences from donor TSP2 and inferred the germline ancestor of each cell. The mutational load varies dramatically by tissue of residence, with blood having the lowest mutational load compared to solid tissues (**fig. S11D**); solid tissues have an order of magnitude more mutations per nucleotide (mean=0.076, s.d.=0.026) compared to the blood (0.0069), suggesting that the immune infiltrates of solid tissues are dominated by mature B cells.

B cells also undergo class-switch recombination which diversifies the humoral immune response by using constant region genes with distinct roles in immunity. We classified every B cell in the dataset as IgA, IgG, or IgM expressing and then calculated the relative amounts of each cellular isotype in each tissue (**Fig. 3B, table S5**). Secretory IgA is known to interact with pathogens and commensals at the mucosae, IgG is often involved in direct neutralization of pathogens, and IgM is typically expressed in naive B cells or secreted in first response to pathogens. Consistent with this, our analysis revealed opposing gradients of prevalence of IgA and IgM expressing B cells across the tissues with blood having the lowest relative abundance of IgA producing cells and the large intestine having the highest relative abundance, and the converse for IgM expressing B cells (**Fig. 3B**).

Endothelial Cells Subtypes with Tissue-Specific Gene Expression Programs

As another example of analyzing shared cell types across organs, we focused on endothelial cells (ECs). While ECs are often categorized as a single cell type, they exhibit differences in morphology, structure, immunomodulatory and metabolic phenotypes depending on their tissue of origin. Here, we discovered that tissue-specificity is also reflected in their transcriptomes, as ECs mainly cluster by tissue-of-origin (**table S6**). UMAP analysis (**fig. S12A**) revealed that lung, heart, uterus, liver, pancreas, fat and muscle ECs exhibited the most distinct transcriptional signatures, reflecting their highly specialized roles. These distributions were conserved across donors (**fig. S12B**).

Interestingly, ECs from the thymus, vasculature, prostate, and eye were similarly distributed across several clusters, suggesting not only similarity in transcriptional profiles but in their sources of heterogeneity. Differential gene expression analysis between ECs of these 16 tissues revealed several canonical and previously undescribed tissue-specific vascular markers (**Fig. 3C**). These data recapitulate tissue-specific vascular markers such as LCN1 (tear lipocalin) in the eye, ABCG2 (transporter at the blood-testis barrier) in the prostate, and OIT3 (oncoprotein induced transcript 3) in the liver. Of the potential novel markers determined by this analysis, SLC14A1 (solute carrier family 14 member 1) appears to be a new specific marker for endothelial cells in the heart, whose expression was independently validated with data from the Human Protein Atlas (*18*) (**fig. S13**).

Notably, lung ECs formed two distinct populations, which is in line with the aerocyte (aCap- EDNRB+) and general capillary (gCap - PLVAP+) cells described in the mouse and human lung (*19*) (**fig. S12 C,D**). The transcriptional profile of gCaps were also more similar to ECs from other tissues, indicative of their general vascular functions in contrast to the more specialized aCap populations. Lastly, we detected two distinct populations of ECs in the muscle, including a MSX1+ population with strong angiogenic and endothelial cell proliferation signatures, and a CYP1B1+ population enriched in metabolic genes, suggesting the presence of functional specialization in the muscle vasculature (**fig. S12 E,F**).

Alternative Splice Variants are Cell Type Specific

We used SICILIAN (*20*) to identify alternative splice junctions in Tabula Sapiens using both 10X and Smart-Seq2 sequencing technologies and found a total of 955,785 junctions (**fig. S14A-E, table S7**). 217,855 of these were previously annotated, and thus our data provides independent validation of 61% of the total junctions catalogued in the entire RefSeq database. Although annotated junctions made up only 22.8% of the unique junctions, they represent 93% of total reads, indicating that previously annotated junctions tend to be expressed at higher levels than novel junctions. We additionally found 34,624 novel junctions between previously annotated 3' and 5' splice sites (3.6%). We also identified 119,276 junctions between a previously annotated site and a novel site in the gene (12.4%). This leaves 584,030 putative junctions for which both splice sites were previously unannotated, i.e. about 61% of the total detected junctions. Most of these have at least one end in a known gene (94.7%), while the remainder represent potential new splice variants from unannotated regions (5.3%). In the absence of independent validation, we conservatively characterized all of the unannotated splices as putative novel junctions. We then used the GTEx database (*21*) to seek independent corroborating evidence of these putative novel junctions, and found that reads corresponding to nearly one third of these novel junctions can be found within GTEx data (**table S7**); this corresponds to more than 300,000 new validated splice variants revealed by the Tabula Sapiens.

Hundreds of splice variants are used in a highly cell-type specific fashion; these can be explored in the cellxgene browser (*14*) which uses a statistic called SpliZ (*22*). Here we focus on two examples of cell type specific splicing of two well studied genes: MYL6

and CD47; similar cell-type specific splice usage was also observed with TPM1, TPM2, and ATP5F1C, three other genes with well-characterized splice variants (**fig. S15**).

MYL6 is an “essential light chain” (ELC) for myosin and is highly expressed in all tissues and compartments. Yet, splicing of MYL6, in particular involving the inclusion/exclusion of exon 6 (**Fig. 4A**) varies in a cell-type and compartment-specific manner (**Fig. 4B**). While the -exon6 isoform has previously been mainly described in phasic smooth muscle (*23*), we discovered it can also be the predominant isoform in non-smooth-muscle cell types. Our analysis establishes pervasive regulation of MYL6 splicing in many cell types, such as endothelial and immune cells. These previously unknown compartment-specific expression patterns of the two MYL6 isoforms are reproduced in multiple individuals from the Tabula Sapiens dataset (**Fig. 4A,B**).

CD47 is a multi-spanning membrane protein involved in many cellular processes, including angiogenesis, cell migration, and as a “don’t eat me” signal to macrophages (*24*). Differential use of exons 7-10 (**Fig. 4C and fig. S14F**) compose a variably long cytoplasmic tail (*25*). Immune cells – but also stromal and endothelial cells – have a distinct, consistent splicing pattern in CD47 that dominantly excludes two proximal exons and splicing directly to exon 8. In contrast to other compartments, epithelial cells exhibit a different splicing pattern that increases the length of the cytoplasmic tail by splicing more commonly to exon 9 and exon 10 (**Fig. 4D**). Characterization of the splicing programs of CD47 in single cells may have implications for understanding the

differential signaling activities of CD47 and for therapeutic manipulation of CD47 function.

Cell State Dynamics Can Be Inferred From A Single Time Point

Although the Tabula Sapiens was created from a single moment in time for each donor, it is possible to infer dynamic information from the data. Cell division is an important transient change of internal cell state, and we computed a cycling index for each cell type to identify actively proliferating versus quiescent or post-mitotic cell states. Rapidly dividing progenitor cells had among the highest cycling indices, while cell types from the endothelial and stromal compartments, which are known to be largely quiescent, had low cycling indices (**Fig. 5A**). In intestinal tissue, transient amplifying cells and the crypt stem cells divide rapidly in the intestinal crypts to give rise to terminally differentiated cell types of the villi (*26*). These cells were ranked with the highest cycling indices whereas terminally differentiated cell types such as the goblet cells had the lowest ranks (**fig. S16A**). To complement the computational analysis of cell cycling, we performed immunostaining of intestinal tissue for MKI67 protein (commonly referred to as Ki-67) and confirmed that transient amplifying cells abundantly express this proliferation marker (**fig. S16B,C**), supporting that this marker is differentially expressed in the G2/M cluster.

We observed several interesting tissue-specific differences in cell cycling. To illustrate one example, UMAP clustering of macrophages showed tissue-specific clustering of this cell type, and that blood, bone marrow, and lung macrophages have the highest cycling

indices compared to macrophages found in the bladder, skin, and muscle (**fig. S16D-G**). Consistent with this finding, the expression values of CDK-inhibitors (in particular the gene CDKN1A), which block the cell cycle, have the lowest overall expression in macrophages from tissues with high cycling indices (**fig. S16F**).

We used RNA velocity (*27*) as a further dynamic approach to study trans-differentiation of bladder mesenchymal cells to myofibroblasts (**Fig. 5B**). Latent time analysis, which provides an estimate of each cell's internal clock using RNA velocity trajectories (*28*), correctly identified the direction of differentiation (**Fig. 5C**) across multiple donors. Ordering cells as a function of latent time shows clustering of the mesenchymal and myofibroblast gene expression programs for the most dynamically expressed genes (**Fig. 5D**). Among these genes, ACTN1 (Alpha Actinin 1) – a key actin crosslinking protein that stabilizes cytoskeleton-membrane interactions (*29*) – increases across the mesenchymal to myofibroblast trans-differentiation trajectory (**fig. S16H**). Another gene with a similar trajectory is MYLK (myosin light-chain kinase), which also rises as myofibroblasts attain more muscle-like properties (*30*). Finally, a random sampling of the most dynamic genes shared across TSP1 and TSP2 demonstrated that they share concordant trajectories and revealed some of the core genes in the transcriptional program underlying this trans-differentiation event within the bladder (**fig. S16I**).

Unexpected Spatial Variation in the Microbiome

The Tabula Sapiens provided an opportunity to densely and directly sample the human microbiome throughout the gastrointestinal tract. The intestines from donors TSP2 and

TSP14 were sectioned into five regions: the duodenum, jejunum, ileum, and ascending and sigmoid colon (**Fig. 6A**). Each section was transected, and three to nine samples of were collected from each location, followed by amplification and sequencing of the 16S rRNA gene. Uniformly there was a high (~10-30%) relative abundance of Proteobacteria, particularly Enterobacteriaceae (**Fig. 6B**), even in the colon. Samples from each of the duodenum, jejunum, and ileum were largely distinct from one another, with samples exhibiting individual patterns of blooming or absence of certain families (**Fig. 6B**). These data reveal that the microbiota is patchy even at a 3-inch length scale. We observed similar heterogeneity in both donors (**fig. S17A-C**). In the small intestine, richness (number of observed species) was also variable, and was negatively correlated with the relative abundance of Burkholderiaceae (**Fig. 6B**); in TSP2, the Proteobacteria phylum was dominated by Enterobacteriaceae, which was present at >30% in all samples at a level negatively correlated with richness (**fig. S17A-C**). In a comparison of species from adjacent regions across the gut, a large fraction of species was unique to each region (**Fig. 6C**), reflecting the patchiness. These data are derived from only two donor samples and further conclusions about the statistics and extent of microbial patchiness will require larger studies.

We analyzed host immune cells in conjunction with the spatial microbiome data; UMAP clustering analysis revealed that the small intestine T cell pool from TSP14 contained a population with distinct transcriptomes (**Fig. 6D**). The most significant transcriptional differences in T cells between the small and large intestine were genes associated with trafficking, survival, and activation (**Fig. 6E**, **table S8**). For example, expression of the

long non-coding RNA MALAT1, which impacts the regulatory function of T cells, and CCR9, which mediates T lymphocyte development and migration to the intestine (*31*), were high only in the small intestine, while GPR15 (colonic T cell trafficking), SELENBP1 (selenium transporter), ANXA1 (repressor of inflammation in T cells), KLRC2 (T cell lectin), CD24 (T cell survival), GDF15 (T cell inhibitor), and RARRES2 (T cell chemokine) exhibited much higher expression in the large intestine. Within the epithelial cells, we observed distinct transcriptomes between small and large intestine Paneth cells and between small and large intestine enterocytes, while there was some degree of overlap for each of the two cell types for either location (**fig. S17E,F**). The site-specific composition of the microbiome in the intestine, paired with distinct T cell populations at each site helps define local host-microbe interactions that occur in the GI tract and is likely reflective of a gradient of physiological conditions that influence host-microbe dynamics.

Conclusion

The Tabula Sapiens is part of a growing set of data which when analyzed together will enable many interesting comparisons of both a biological and a technical nature. Studying particular cell types across organs, datasets, and species will yield new biological insights – as shown with fibroblasts (*32*). Similarly, comparing fetal human cell types (*33*) to those determined here in adults may give insight into the loss of plasticity from early development to maturity. Having multi-organ data from individual donors may facilitate development of methods to compare diverse datasets and yield understanding of technical artifacts from various approaches (*8*, *9*, *34*, *35*). The Tabula Sapiens has

enabled discoveries relating to shared behavior and organ specific differences across cell types. For example, we found T cell clones shared between organs, and characterized organ dependent hypermutation rates amongst resident B cells. Endothelial cells and macrophages are cell types which are shared across tissues, but often show subtle tissue-specific differences in gene expression. We found an unexpectedly large and diverse amount of cell-type specific RNA splice variant usage, and discovered and validated many new splices. These are but a few examples of how the Tabula Sapiens represents a broadly useful reference to understand and explore human biology deeply at cellular resolution.

Brief synopsis of methods

Fresh whole non-transplantable organs, or 1-2cm^2 organ samples, were obtained from surgery and then transported on ice by courier to tissue expert labs where they were immediately prepared for transcriptome sequencing. Single-cell suspensions were prepared for 10x Genomics 3' V3.1 droplet-based sequencing and for FACS sorted 384-well plate Smart-seq2. Preparation began with dissection, digestion with enzymes and physical manipulation; tissue specific details are in the methods supplement (*12*). Cell suspensions from some organs were normalized by major cell compartment (epithelial, endothelial, immune, and stromal) using antibody-labelled magnetic microbeads to enrich rare cell types. cDNA and sequencing libraries were prepared and run on the Illumina NovaSeq 6000 with the goal to obtain 10,000 droplet-based cells and 1000 plate-based cells for each organ. Sequences were de-multiplexed and aligned to the

GRCh38 reference genome. Gene count tables were generated with CellRanger (droplet samples), or STAR and HTSEQ (plate samples). Cells with low UMI counts and low gene counts were removed. Droplet cells were filtered to remove barcode-hopping events and filtered for ambient RNA using DecontX. Sequencing batches were harmonized using scVI and projected to 2-D space with UMAP for analysis by the tissue experts. Expert annotation was made through the cellxgene browser and regularized with a public cell ontology. Annotation was manually QC checked and cross-validated with PopV, an annotation tool, which employs seven different automated annotation methods. For complete methods, see supplementary materials (*12*).

References:

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Molecular Biology of the Cell* (Garland Science, ed. 4th, 2002).

2. A. Regev, S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, H. Clevers, B. Deplancke, I. Dunham, J. Eberwine, R. Eils, W. Enard, A. Farmer, L. Fugger, B. Göttgens, N. Hacohen, M. Haniffa, M. Hemberg, S. Kim, P. Klenerman, A. Kriegstein, E. Lein, S. Linnarsson, E. Lundberg, J. Lundeberg, P. Majumder, J. C. Marioni, M. Merad, M. Mhlanga, M. Nawijn, M. Netea, G. Nolan, D. Pe'er, A. Phillipakis, C. P. Ponting, S. Quake, W. Reik, O. Rozenblatt-Rosen, J. Sanes, R. Satija, T. N. Schumacher, A. Shalek, E. Shapiro, P. Sharma, J. W. Shin, O. Stegle, M. Stratton, M. J. T. Stubbington, F. J. Theis, M. Uhlen, A. Van Oudenaarden, A. Wagner, F. Watt, J. Weissman, B. Wold, R. Xavier, N. Yosef, The human cell atlas. *eLife*. **6** (2017), doi:10.7554/eLife.27041.

3. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Tissue-based map of the human proteome. *Science*. **347** (2015), doi:10.1126/SCIENCE.1260419.

4. P. J. Thul, L. Åkesson, M. Wiking, D. Mahdessian, A. Geladaki, H. A. Blal, T. Alm, A. Asplund, L. Björk, L. M. Breckels, A. Bäckström, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Hober, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C. M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J. M. Schwenk, Å. Sivertsson, E. Sjöstedt, M. Skogs, C. Stadler, D. P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoglu, F. Pontén, K. von Feilitzen, K. S. Lilley, M. Uhlén, E. Lundberg, A subcellular map of the human proteome. *Science*. **356** (2017), doi:10.1126/SCIENCE.AAL3321.

5. A. O. Pisco, B. Tojo, A. McGeever, Single-Cell Analysis for Whole-Organism Datasets. *Annu. Rev. Biomed. Data Sci.* **4** (2021), doi:10.1146/ANNUREV-BIODATASCI-092820-031008.

6. S. Aldridge, S. A. Teichmann, Single cell transcriptomics comes of age. *Nat. Commun.* **11** (2020), pp. 1–4.

7. A. R. Wu, J. Wang, A. M. Streets, Y. Huang, Single-Cell Transcriptional Analysis. *Annu. Rev. Anal. Chem.* **10**, 439–462 (2017).

8. X. Han, Z. Zhou, L. Fei, H. Sun, R. Wang, Y. Chen, H. Chen, J. Wang, H. Tang, W. Ge, Y. Zhou, F. Ye, M. Jiang, J. Wu, Y. Xiao, X. Jia, T. Zhang, X. Ma, Q. Zhang, X.

Bai, S. Lai, C. Yu, L. Zhu, R. Lin, Y. Gao, M. Wang, Y. Wu, J. Zhang, R. Zhan, S. Zhu, H. Hu, C. Wang, M. Chen, H. Huang, T. Liang, J. Chen, W. Wang, D. Zhang, G. Guo, Construction of a human cell landscape at single-cell level. *Nature*. **581**, 303–309 (2020).

9. S. He, L.-H. Wang, Y. Liu, Y.-Q. Li, H.-T. Chen, J.-H. Xu, W. Peng, G.-W. Lin, P.-P. Wei, B. Li, X. Xia, D. Wang, J.-X. Bei, X. He, Z. Guo, Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biol. 2020 211*. **21**, 1–34 (2020).

10. Tabula Muris Consortium, Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*. **562**, 367–372 (2018).

11. The Tabula Muris Consortium, A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*. **583**, 590–595 (2020).

12. The Tabula Sapiens Consortium, Methods for “The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans,” (available at tabula-sapiens-portal.ds.czbiohub.org).

13. S. Wang, A. O. Pisco, A. McGeever, M. Brbic, M. Zitnik, S. Darmanis, J. Leskovec, J. Karkanias, R. B. Altman, Leveraging the Cell Ontology to classify unseen cell types. *Nat. Commun.* **12**, 5556 (2021).

14. The Tabula Sapiens Consortium, Tabula Sapiens data portal, (available at tabula-sapiens-portal.ds.czbiohub.org).

15. E. M. Steinert, J. M. Schenkel, K. A. Fraser, L. K. Beura, L. S. Manlove, B. Z. Igyártó, P. J. Southern, D. Masopust, Quantifying Memory CD8 T Cells Reveals Regionalization of Immunosurveillance. *Cell*. **161**, 737–749 (2015).

16. L. C. Davies, P. R. Taylor, Tissue-resident macrophages: then and now. *Immunology*. **144**, 541–548 (2015).

17. D. I. Godfrey, H.-F. Koay, J. McCluskey, N. A. Gherardin, The biology and functional importance of MAIT cells. *Nat. Immunol. 2019 209*. **20**, 1110–1128 (2019).

18. M. Karlsson, C. Zhang, L. Méar, W. Zhong, A. Digre, B. Katona, E. Sjöstedt, L. Butler, J. Odeberg, P. Dusart, F. Edfors, P. Oksvold, K. von Feilitzen, M. Zwahlen, M. Arif, O. Altay, X. Li, M. Ozcan, A. Mardinoglu, L. Fagerberg, J. Mulder, Y. Luo, F. Ponten, M. Uhlén, C. Lindskog, A single–cell type transcriptomics map of human tissues. *Sci. Adv.* **7**, eabh2169.

19. A. Gillich, F. Zhang, C. G. Farmer, K. J. Travaglini, S. Y. Tan, M. Gu, B. Zhou, J. A. Feinstein, M. A. Krasnow, R. J. Metzger, Capillary cell-type specialization in the alveolus. *Nat. 2020 5867831*. **586**, 785–789 (2020).

20. R. Dehghannasiri, J. E. Olivieri, A. Damljanovic, J. Salzman, Specific splice junction detection in single cells with SICILIAN. *Genome Biol.* **22**, 219 (2021).

21. J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, B. Foster, M. Moser, E. Karasik, B. Gillard, K. Ramsey, S. Sullivan, J. Bridge, H. Magazine, J. Syron, J. Fleming, L. Siminoff, H. Traino, M. Mosavel, L. Barker, S. Jewell, D. Rohrer, D. Maxim, D. Filkins, P. Harbach, E. Cortadillo, B. Berghuis, L. Turner, E. Hudson, K. Feenstra, L. Sobin, J. Robb, P. Branton, G. Korzeniewski, C. Shive, D. Tabor, L. Qi, K. Groch, S. Nampally, S. Buia, A. Zimmerman, A. Smith, R. Burges, K. Robinson, K. Valentino, D. Bradbury, M. Cosentino, N. Diaz-Mayoral, M. Kennedy, T. Engel, P. Williams, K. Erickson, K. Ardlie, W. Winckler, G. Getz, D. DeLuca, D. MacArthur, M. Kellis, A. Thomson, T. Young, E. Gelfand, M. Donovan, Y. Meng, G. Grant, D. Mash, Y. Marcus, M. Basile, J. Liu, J. Zhu, Z. Tu, N. J. Cox, D. L. Nicolae, E. R. Gamazon, H. K. Im, A. Konkashbaev, J. Pritchard, M. Stevens, T. Flutre, X. Wen, E. T. Dermitzakis, T. Lappalainen, R. Guigo, J. Monlong, M. Sammeth, D. Koller, A. Battle, S. Mostafavi, M. McCarthy, M. Rivas, J. Maller, I. Rusyn, A. Nobel, F. Wright, A. Shabalin, M. Feolo, N. Sharopova, A. Sturcke, J. Paschal, J. M. Anderson, E. L. Wilder, L. K. Derr, E. D. Green, J. P. Struewing, G. Temple, S. Volpi, J. T. Boyer, E. J. Thomson, M. S. Guyer, C. Ng, A. Abdallah, D. Colantuoni, T. R. Insel, S. E. Koester, A. R. Little, P. K. Bender, T. Lehner, Y. Yao, C. C. Compton, J. B. Vaught, S. Sawyer, N. C. Lockhart, J. Demchok, H. F. Moore, The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).

22. J. E. Olivieri, R. Dehghannasiri, P. L. Wang, S. Jang, A. de Morree, S. Y. Tan, J. Ming, A. Ruohao Wu, Tabula Sapiens Consortium, S. R. Quake, M. A. Krasnow, J. Salzman, RNA splicing programs define tissue compartments and cell types at single-cell resolution. *eLife*. **10**, e70692 (2021).

23. S. A. Fisher, Vascular smooth muscle phenotypic diversity and function. *Physiol. Genomics*. **42A**, 169–187 (2010).

24. A. A. Barkal, R. E. Brewer, M. Markovic, M. Kowarsky, S. A. Barkal, B. W. Zaro, V. Krishnan, J. Hatakeyama, O. Dorigo, L. J. Barkal, I. L. Weissman, CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. *Nature*. **572**, 392–396 (2019).

25. S. M. G. Hayat, V. Bianconi, M. Pirro, M. R. Jaafari, M. Hatamipour, A. Sahebkar, CD47: role in the immune system and application to cancer therapy. *Cell. Oncol. 2019 431*. **43**, 19–30 (2019).

26. H. Gehart, H. Clevers, Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16** (2019), pp. 19–34.

27. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P.

Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. *Nature*. **560**, 494 (2018).

28. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).

29. A. C. H. Murphy, P. W. Young, The actinin family of actin cross-linking proteins - A genetic perspective. *Cell Biosci.* **5** (2015), , doi:10.1186/s13578-015-0029-7.

30. R. Li, X. Li, J. Hagood, M.-S. Zhu, X. Sun, Myofibroblast contraction is essential for generating and regenerating the gas-exchange surface. *J. Clin. Invest.* **130**, 2859–2871 (2020).

31. S. Uehara, A. Grinberg, J. M. Farber, P. E. Love, A role for CCR9 in T lymphocyte development and migration. *J. Immunol. Baltim. Md 1950*. **168**, 2811–2819 (2002).

32. M. B. Buechler, R. N. Pradhan, A. T. Krishnamurty, C. Cox, A. K. Calviello, A. W. Wang, Y. A. Yang, L. Tam, R. Caothien, M. Roose-Girma, Z. Modrusan, J. R. Arron, R. Bourgon, S. Müller, S. J. Turley, Cross-tissue organization of the fibroblast lineage. *Nature*. **593**, 575–579 (2021).

33. J. Cao, D. R. O'Day, H. A. Pliner, P. D. Kingsley, M. Deng, R. M. Daza, M. A. Zager, K. A. Aldinger, R. Blecher-Gonen, F. Zhang, M. Spielmann, J. Palis, D. Doherty, F. J. Steemers, I. A. Glass, C. Trapnell, J. Shendure, A human cell atlas of fetal gene expression. *Science*. **370** (2020), doi:10.1126/science.aba7721.

34. G. Eraslan, E. Drokhlyansky, S. Anand, A. Subramanian, E. Fiskin, M. Slyper, J. Wang, N. V. Wittenberghe, J. M. Rouhana, J. Waldman, O. Ashenberg, D. Dionne, T. S. Win, M. S. Cuoco, O. Kuksenko, P. A. Branton, J. L. Marshall, A. Greka, G. Getz, A. V. Segrè, F. Aguet, O. Rozenblatt-Rosen, K. G. Ardlie, A. Regev, "Single-nucleus cross-tissue molecular reference maps to decipher disease gene function" (2021), p. 2021.07.19.452954, , doi:10.1101/2021.07.19.452954.

35. C. D. Conde, C. Xu, L. B. Jarvis, T. Gomes, S. K. Howlett, D. B. Rainbow, O. Suchanek, H. W. King, L. Mamanova, K. Polanski, N. Huang, E. S. Fasouli, K. T. Mahbubani, M. Prete, L. Tuck, N. Richoz, Z. K. Tuong, L. Campos, H. S. Mousa, E. J. Needham, S. Pritchard, T. Li, R. Elmentaite, J. Park, D. K. Menon, O. A. Bayraktar, L. K. James, K. B. Meyer, M. R. Clatworthy, K. Saeb-Parsy, J. L. Jones, S. A. Teichmann, "Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture in humans" (2021), p. 2021.04.28.441762, , doi:10.1101/2021.04.28.441762.

36. The Tabula Sapiens Consortium, *czbiohub/tabula-sapiens: v1.0* (Zenodo, 2022; https://zenodo.org/record/6069683).

37. The Tabula Sapiens Consortium, Tabula Sapiens Single-Cell Dataset (2021), p. 46920039266 Bytes, doi:10.6084/M9.FIGSHARE.14267219.

38. The Tabula Sapiens Consortium, Tabula Sapiens H&E Image Collection (2021), p. 120026032906 Bytes, doi:10.6084/M9.FIGSHARE.14962947.

39. The Tabula Sapiens Consortium, Tabula Sapiens Splicing (2021), p. 3624676150 Bytes, doi:10.6084/M9.FIGSHARE.14977281.

40. S. Picelli, Å. K. Björklund, O. R. Faridani, S. Sagasser, G. Winberg, R. Sandberg, *Nat. Methods*, in press, doi:10.1038/nmeth.2639.

41. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

42. J. A. Griffiths, A. C. Richard, K. Bach, A. T. L. Lun, J. C. Marioni, Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat. Commun.* **9**, 2667 (2018).

43. S. Yang, S. E. Corbett, Y. Koga, Z. Wang, W. E. Johnson, M. Yajima, J. D. Campbell, Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol.* **21**, 57 (2020).

44. R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, N. Yosef, Deep generative modeling for single-cell transcriptomics. *Nat. Methods*. **15**, 1053–1058 (2018).

45. A. Gayoso, R. Lopez, G. Xing, P. Boyeau, K. Wu, M. Jayasuriya, E. Melhman, M. Langevin, Y. Liu, J. Samaran, G. Misrachi, A. Nazaret, O. Clivio, C. Xu, T. Ashuach, M. Lotfollahi, V. Svensson, E. da V. Beltrame, C. Talavera-López, L. Pachter, F. J. Theis, A. Streets, M. I. Jordan, J. Regier, N. Yosef, "scvi-tools: a library for deep probabilistic analysis of single-cell omics data" (preprint, Bioinformatics, 2021), , doi:10.1101/2021.04.28.441833.

46. E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, E. W. Newell, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2019).

47. C. Megill, B. Martin, C. Weaver, S. Bell, L. Prins, S. Badajoz, B. McCandless, A. O. Pisco, M. Kinsella, F. Griffin, J. Kiggins, G. Haliburton, A. Mani, M. Weiden, M. Dunitz, M. Lombardo, T. Huang, T. Smith, S. Chambers, J. Freeman, J. Cool, A. Carr, cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices. *bioRxiv*, in press, doi:10.1101/2021.04.05.438318.

48. A. D. Diehl, T. F. Meehan, Y. M. Bradford, M. H. Brush, W. M. Dahdul, D. S. Dougall, Y. He, D. Osumi-Sutherland, A. Ruttenberg, S. Sarntivijai, C. E. V. Slyke, N. A. Vasilevsky, M. A. Haendel, J. A. Blake, C. J. Mungall, The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J. Biomed. Semant.* **7**, 44 (2016).

49. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D.

Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

50. C. Xu, R. Lopez, E. Mehlman, J. Regier, M. I. Jordan, N. Yosef, Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* **17**, e9620 (2021).

51. K. Polański, M. D. Young, Z. Miao, K. B. Meyer, S. A. Teichmann, J.-E. Park, BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* (2019), doi:10.1093/bioinformatics/btz625.

52. B. Hie, B. Bryson, B. Berger, Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat. Biotechnol.* **37**, 685–691 (2019).

53. N. Sofroniew, Talley Lambert, K. Evans, J. Nunez-Iglesias, P. Winston, G. Bokota, G. Peña-Castellanos, K. Yamauchi, M. Bussonnier, A. C. Solak, Ziyangczi, D. D. Pop, G. Buckley, Pam, Alisterburt, V. Hilsenstein, A. Sweet, Hector, J. Freeman, P. Boone, Jordão Bragantini, A. R. Lowe, C. Gohlke, Loic Royer, Kir0ul, H. Har-Gil, Lukasz Migas, M. Kittisopikul, R. Haase, S. Axelrod, *napari/napari: 0.4.10rc0* (Zenodo, 2021; https://zenodo.org/record/4968798).

54. S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, N. Rajpoot, Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* **58**, 101563 (2019).

55. R. Core Team, R: A Language and Environment for Statistical Computing (2017) (available at https://www.r-project.org/).

56. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods*. **13**, 581–583 (2016).

57. P. Yilmaz, L. W. Parfrey, P. Yarza, J. Gerken, E. Pruesse, C. Quast, T. Schweer, J. Peplies, W. Ludwig, F. O. Glöckner, The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* **42**, D643–D648 (2014).

58. K. P. Schliep, phangorn: phylogenetic analysis in R. *Bioinformatics*. **27**, 592–593 (2011).

59. P. J. McMurdie, S. Holmes, phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*. **8**, e61217 (2013).

60. M. L. Whitfield, G. Sherlock, A. J. Saldanha, J. I. Murray, C. A. Ball, K. E. Alexander, J. C. Matese, C. M. Perou, M. M. Hurt, P. O. Brown, D. Botstein, Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. *Mol. Biol. Cell*. **13**, 1977–2000 (2002).

61. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*. **161**, 1202–1214 (2015).

62. I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J.-R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. K. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A.-C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jane-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, L. A. Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. **352**, 189–196 (2016).

63. A. Santos, R. Wernersson, L. J. Jensen, Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Res.* **43**, D1140–D1144 (2015).

64. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies and species. *Nat. Biotechnol.* **36**, 411-420 (2018), doi:10.1038/nbt.4096.

65. S. C. van den Brink, F. Sage, Á. Vértesy, B. Spanjaard, J. Peterson-Maduro, C. S. Baron, C. Robin, A. van Oudenaarden, Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods*. **14**, 935–936 (2017).

66. P. G. Ferreira, M. Muñoz-Aguirre, F. Reverter, C. P. Sá Godinho, A. Sousa, A. Amadoz, R. Sodaei, M. R. Hidalgo, D. Pervouchine, J. Carbonell-Caballero, R. Nurtdinov, A. Breschi, R. Amador, P. Oliveira, C. Çubuk, J. Curado, F. Aguet, C. Oliveira, J. Dopazo, M. Sammeth, K. G. Ardlie, R. Guigó, The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nat. Commun.* **9**, 490 (2018).

67. R. Massoni-Badosa, G. Iacono, C. Moutinho, M. Kulis, N. Palau, D. Marchese, J. Rodríguez-Ubreva, E. Ballestar, G. Rodriguez-Esteban, S. Marsal, M. Aymerich, D. Colomer, E. Campo, A. Julià, J. I. Martín-Subero, H. Heyn, Sampling time-dependent artifacts in single-cell genomics studies. *Genome Biol.* **21**, 112 (2020).

68. C. T. J. van Velthoven, A. de Morree, I. M. Egner, J. O. Brett, T. A. Rando, Transcriptional Profiling of Quiescent Muscle Stem Cells In Vivo. *Cell Rep.* **21**, 1994–2004 (2017).

69. A. K. Savage, M. V. Gutschow, T. Chiang, K. Henderson, R. Green, M. Chaudhari, E. Swanson, A. T. Heubeck, N. Kondza, K. C. Burley, P. C. Genge, C. Lord, T. Smith, Z. Thomson, A. Beaubien, E. Johnson, J. Goldy, H. Bolouri, J. H. Buckner, P. Meijer, E. M. Coffey, P. J. Skene, T. R. Torgerson, X. Li, T. F. Bumol,

"Multimodal Analysis for Human ex vivo Studies Shows Extensive Molecular Changes from Delays in Blood Processing" (preprint, Immunology, 2020), , doi:10.1101/2020.10.18.344663.

70. M. J. T. Stubbington, T. Lönnberg, V. Proserpio, S. Clare, A. O. Speak, G. Dougan, S. A. Teichmann, *Nat. Methods*, in press, doi:10.1038/nmeth.3800.

71. I. Lindeman, G. Emerton, L. Mamanova, O. Snir, K. Polanski, S.-W. Qiao, L. M. Sollid, S. A. Teichmann, M. J. T. Stubbington, BraCeR: B-cell-receptor reconstruction and clonality inference from single-cell RNA-seq. *Nat. Methods*. **15**, 563–565 (2018).

72. J. Koster, S. Rahmann, Snakemake--a scalable bioinformatics workflow engine. *Bioinformatics*. **28**, 2520–2522 (2012).

73. J. Ye, N. Ma, T. L. Madden, J. M. Ostell, IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41**, W34–W40 (2013).

74. G. Sturm, T. Szabo, G. Fotakis, M. Haider, D. Rieder, Z. Trajanoski, F. Finotello, Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. *Bioinformatics*. **36**, 4817–4818 (2020).

75. T. P. Peixoto, The graph-tool python library (2017), p. 136152697 Bytes, doi:10.6084/M9.FIGSHARE.1164194.

76. M. Waskom, seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).

77. H. Gehart, H. Clevers, Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 19–34 (2019).

78. K. J. Travaglini, A. N. Nabhan, L. Penland, R. Sinha, A. Gillich, R. V. Sit, S. Chang, S. D. Conley, Y. Mori, J. Seita, G. J. Berry, J. B. Shrager, R. J. Metzger, C. S. Kuo, N. Neff, I. L. Weissman, S. R. Quake, M. A. Krasnow, A molecular cell atlas of the human lung from single cell RNA sequencing. Nature **587**, 619-625 (2020).

79. Z. Yu, J. Liao, Y. Chen, C. Zou, H. Zhang, J. Cheng, D. Liu, T. Li, Q. Zhang, J. Li, X. Yang, Y. Ye, Z. Huang, X. Long, R. Yang, Z. Mo, Single-Cell Transcriptomic Map of the Human and Mouse Bladders. *J. Am. Soc. Nephrol.* **30**, 2159–2176 (2019).

80. J.-P. Volkmer, D. Sahoo, R. K. Chin, P. L. Ho, C. Tang, A. V. Kurtova, S. B. Willingham, S. K. Pazhanisamy, H. Contreras-Trujillo, T. A. Storm, Y. Lotan, A. H. Beck, B. I. Chung, A. A. Alizadeh, G. Godoy, S. P. Lerner, M. van de Rijn, L. D. Shortliffe, I. L. Weissman, K. S. Chan, Three differentiation states risk-stratify bladder cancer into distinct subtypes. *Proc. Natl. Acad. Sci.* **109**, 2078–2083 (2012).

81. O. W. Hakenberg, C. Linne, A. Manseck, M. P. Wirth, Bladder wall thickness in normal adults and men with mild lower urinary tract symptoms and benign prostatic enlargement. *Neurourol. Urodyn.* **19**, 585–593 (2000).

82. A. Armulik, G. Genové, C. Betsholtz, Pericytes: Developmental, Physiological, and Pathological Perspectives, Problems, and Promises. *Dev. Cell*. **21**, 193–215 (2011).

83. S. R. Bolla, N. Odeluga, R. Jetti, in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2022; http://www.ncbi.nlm.nih.gov/books/NBK540963/).

84. J. S. Dixon, S.-A. Gilpin, C. J. GlLPlN, J. A. Gosling, Intramural Ganglia of the Human Urinary Bladder. *Br. J. Urol.* **55**, 195–198 (1983).

85. M. Joseph, D. Enting, Immune Responses in Bladder Cancer-Role of Immune Cell Populations, Prognostic Factors and Therapeutic Implications. *Front. Oncol.* **9**, 1270 (2019).

86. A. B. Rubenstein, G. R. Smith, U. Raue, G. Begue, K. Minchev, F. Ruf-Zamojski, V. D. Nair, X. Wang, L. Zhou, E. Zaslavsky, T. A. Trappe, S. Trappe, S. C. Sealfon, Single-cell transcriptional profiles in human skeletal muscle. *Sci. Rep.* **10**, 229 (2020).

87. A. J. De Micheli, J. A. Spector, O. Elemento, B. D. Cosgrove, A reference single-cell transcriptomic atlas of human skeletal muscle tissue reveals bifurcated muscle stem cell populations. *Skelet. Muscle*. **10**, 19 (2020).

88. J. A. Zoladz, D. Semik, B. Zawadowska, J. Majerczak, J. Karasinski, L. Kolodziejski, K. Duda, W. M. Kilarski, Capillary density and capillary-to-fibre ratio in vastus lateralis muscle of untrained and trained men. *Folia Histochem. Cytobiol.* **43**, 11–17 (2005).

89. A. A. Al-Shammari, R. W. P. Kissane, S. Holbek, A. L. Mackey, T. R. Andersen, E. A. Gaffney, M. Kjaer, S. Egginton, Integrated method for quantitative morphometry and oxygen transport modeling in striated muscle. *J. Appl. Physiol. Bethesda Md 1985*. **126**, 544–557 (2019).

90. C. Christov, F. Chrétien, R. Abou-Khalil, G. Bassez, G. Vallet, F.-J. Authier, Y. Bassaglia, V. Shinin, S. Tajbakhsh, B. Chazaud, R. K. Gherardi, Muscle satellite cells and endothelial cells: close neighbors and privileged partners. *Mol. Biol. Cell*. **18**, 1397–1409 (2007).

91. A. L. Mackey, M. Magnan, B. Chazaud, M. Kjaer, Human skeletal muscle fibroblasts stimulate in vitro myogenesis and in vivo muscle regeneration. *J. Physiol.* **595**, 5115–5127 (2017).

92. J. H. Wokke, F. G. Jennekens, C. J. van den Oord, H. Veldman, L. M. Smit, G. J. Leppink, Morphological changes in the human end plate with age. *J. Neurol. Sci.* **95**, 291–310 (1990).

93. L. F. Hesselmans, F. G. Jennekens, C. J. van den Oord, A. B. Oestreicher, H. Veldman, W. H. Gispen, A light and electron microscopical study of B-50 (GAP-43) in human intramuscular nerve and neuromuscular junctions during development. *J. Neurol. Sci.* **89**, 301–311 (1989).

94. R. W. Emmens, L. Woudstra, A. Papageorgiou, P. Carai, S. Smit, S. Seven-Deniz, L. Rozendaal, W. J. Paulus, D. Wouters, S. Zeerleder, J.-L. Murk, M. S. van Ham, S. Heymans, A. C. van Rossum, H. W. M. Niessen, P. A. J. Krijnen, Lymphocytes infiltrate the quadriceps muscle in lymphocytic myocarditis patients: a potential new diagnostic tool. *Can. J. Cardiol.* **30**, 1547–1554 (2014).

95. C. S. Klein, G. D. Marsh, R. J. Petrella, C. L. Rice, Muscle fiber number in the biceps brachii muscle of young and old men. *Muscle Nerve*. **28**, 62–68 (2003).

Acknowledgements: We express our gratitude and thanks to donor WEM and his family, as well as to all of the anonymous organ and tissue donors and their families for giving both the gift of life and the gift of knowledge by their generous donations. We also thank Donor Network West for their cooperation in this project, S. Schmid for a close reading of the manuscript, and B. Tojo for the original artwork in Figure 1. This project has been made possible in part by grant number 2019-203354 from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation, and by support from the Chan Zuckerberg Biohub. The data portal for this publication is available at (*14*)**.**

Data and materials availability: The entire dataset can be explored interactively at http://tabula-sapiens-portal.ds.czbiohub.org/ (*14*). The code used for the analysis is available from Zenodo (https://doi.org/10.5281/zenodo.6069683) (*36*). Gene counts and metadata are available from figshare (https://doi.org/10.6084/m9.figshare.14267219) (*37*) and have been deposited in the Gene Expression Omnibus (GSE149590); the raw data files are available from a public AWS S3 bucket (https://registry.opendata.aws/tabula-sapiens/) and instructions on how to access the data have been provided in the project GitHub. The histology images are available from figshare (https://doi.org/10.6084/m9.figshare.14962947) (*38*). SpliZ scores are available from figshare (https://doi.org/10.6084/m9.figshare.14977281) (*39*).

The Tabula Sapiens Consortium Author List

Overall Project Direction and Coordination

Robert C. Jones[1], Jim Karkanias[2], Mark Krasnow[3,4], Angela Oliveira Pisco[2], Stephen R. Quake[1,2,5], Julia Salzman[3,6], Nir Yosef[2,7,8,9]

Donor Recruitment

Bryan Bulthaup[10], Phillip Brown[10], William Harper[10], Marisa Hemenez[10], Ravikumar Ponnusamy[10], Ahmad Salehi[10], Bhavani A. Sanagavarapu[10], Eileen Spallino[10]

Surgeons

Ksenia A. Aaron[11], Waldo Concepcion[10], James M. Gardner[12,13], Burnett Kelly[10,14], Nikole Neidlinger[10], Zifa Wang[10]

Logistical coordination

Sheela Crasta[1,2], Saroja Kolluru[1,2], Maurizio Morri[2], Angela Oliveira Pisco[2], Serena Y. Tan[15], Kyle J. Travaglini[3], Chenling Xu[7]

Organ Processing

Marcela Alcántara-Hernández[16], Nicole Almanzar[17], Jane Antony[18], Benjamin Beyersdorf[19], Deviana Burhan[20], Kruti Calcuttawala[21], Matthew M. Carter[16], Charles K. F. Chan[18,22], Charles A. Chang[23], Stephen Chang[3,19], Alex Colville[21,24], Sheela Crasta[1,2], Rebecca N. Culver[25], Ivana Cvijović[1,5], Gaetano D'Amato[26], Camille Ezran[3], Francisco X. Galdos[18], Astrid Gillich[3], William R. Goodyer[27], Yan Hang[23,28], Alyssa Hayashi[1], Sahar Houshdaran[29], Xianxi Huang[19,30], Juan C. Irwin[29], SoRi Jang[3], Julia Vallve Juanico[29], Aaron M. Kershner[18], Soochi Kim[21,24], Bernhard Kiss[18], Saroja Kolluru[1,2], William Kong[18], Maya E. Kumar[17], Angera H. Kuo[18], Rebecca Leylek[16],

Baoxiang Li[31], Gabriel B. Loeb[32], Wan-Jin Lu[18], Sruthi Mantri[33], Maxim Markovic[1], Patrick L. McAlpine[11,34], Antoine de Morree[21,24], Maurizio Morri[2], Karim Mrouj[18], Shravani Mukherjee[31], Tyler Muser[17], Patrick Neuhöfer[3,35,36], Thi D. Nguyen[37], Kimberly Perez[16], Ragini Phansalkar[26], Angela Oliveira Pisco[2], Nazan Puluca[18] , Zhen Qi[18], Poorvi Rao[20], Hayley Raquer-McKay[16], Nicholas Schaum[18,21], Bronwyn Scott[31], Bobak Seddighzadeh[38], Joe Segal[20], Sushmita Sen[29], Shaheen Sikandar[18], Sean P. Spencer[16], Lea Steffes[17], Varun R. Subramaniam[31], Aditi Swarup[31], Michael Swift[1,] Kyle J. Travaglini[3], Will Van Treuren[16], Emily Trimm[26], Stefan Veizades[19,39], Sivakamasundari Vijayakumar[18], Kim Chi Vo[29], Sevahn K. Vorperian[1,40], Wanxin Wang[29], Hannah N.W. Weinstein[38], Juliane Winkler[41], Timothy T.H. Wu[3], Jamie Xie[38], Andrea R.Yung[3], Yue Zhang[3]

Sequencing

Angela M. Detweiler[2], Honey Mekonen[2], Norma F. Neff[2], Rene V. Sit[2], Michelle Tan[2], Jia Yan[2]

Histology

Gregory R. Bean[15], Vivek Charu[15], Erna Forgó[15], Brock A. Martin[15], Michael G. Ozawa[15], Oscar Silva[15], Serena Y. Tan[15], Angus Toland[15], Venkata N.P. Vemuri[2]

Data Analysis

Shaked Afik[7], Kyle Awayan[2], Olga Borisovna Botvinnik[2], Ashley Byrne[2], Michelle Chen[1], Roozbeh Dehghannasiri[3,6], Angela M. Detweiler[2], Adam Gayoso[7], Alejandro A Granados[2], Qiqing Li[2], Gita Mahmoudabadi[1], Aaron McGeever[2], Antoine de Morree[21,24], Julia Eve Olivieri[3,6,42], Madeline Park[2], Angela Oliveira Pisco[2], Neha Ravikumar[1], Julia Salzman[3,6], Geoff Stanley[1], Michael Swift[1], Michelle Tan[2], Weilun Tan[2], Alexander J Tarashansky[2], Rohan Vanheusden[2], Sevahn K. Vorperian[1,40], Peter Wang[3,6], Sheng Wang[2], Galen Xing[2], Chenling Xu[6], Nir Yosef[2,6,7,8]

Expert Cell Type Annotation

Marcela Alcántara-Hernández[16], Jane Antony[18], Charles K. F. Chan[18,22], Charles A. Chang[23], Alex Colville[21,24], Sheela Crasta[1,2], Rebecca Culver[25], Les Dethlefsen[43], Camille Ezran[3], Astrid Gillich[3], Yan Hang[23,28], Po-Yi Ho[16], Juan C. Irwin[29], SoRi Jang[3], Aaron M. Kershner[18], William Kong[18], Maya E Kumar[17], Angera H. Kuo[18], Rebecca Leylek[16], Shixuan Liu[3,44], Gabriel B. Loeb[32], Wan-Jin Lu[18], Jonathan S Maltzman[45,46], Ross J. Metzger[27,47], Antoine de Morree[21,24], Patrick Neuhöfer[3,35,36], Kimberly Perez[16], Ragini Phansalkar[26], Zhen Qi[18], Poorvi Rao[20], Hayley Raquer-McKay[16,] Koki Sasagawa[19,] Bronwyn Scott[31], Rahul Sinha[15,18,35], Hanbing Song[38], Sean P. Spencer[16], Aditi Swarup[31], Michael Swift[1], Kyle J. Travaglini[3], Emily Trimm[26], Stefan Veizades[19,39], Sivakamasundari Vijayakumar[18], Bruce Wang[20], Wanxin Wang[29], Juliane Winkler[41], Jamie Xie[38], Andrea R.Yung[3]

Tissue Expert Principal Investigators

Steven E. Artandi[3,35,36], Philip A. Beachy[18,23,48], Michael F. Clarke[18], Linda C. Giudice[29], Franklin W. Huang[38,49], Kerwyn Casey Huang[1,16], Juliana Idoyaga[16], Seung K Kim[23,28], Mark Krasnow[3,4], Christin S. Kuo[17], Patricia Nguyen[19,39,46], Stephen R. Quake[1,2,5], Thomas A. Rando[21,24], Kristy Red-Horse[26], Jeremy Reiter[50], David A. Relman[16,43,46], Justin L. Sonnenburg[16], Bruce Wang[20], Albert Wu[31], Sean M. Wu[19,39], Tony Wyss-Coray[21,24]

Affiliations

[1]Department of Bioengineering, Stanford University; Stanford, CA, USA.

[2]Chan Zuckerberg Biohub; San Francisco, CA, USA.

[3]Department of Biochemistry, Stanford University School of Medicine; Stanford, CA, USA.

[4]Howard Hughes Medical Institute; USA.

[5]Department of Applied Physics, Stanford University; Stanford, CA, USA.

[6]Department of Biomedical Data Science, Stanford University; Stanford, CA, USA.

[7]Center for Computational Biology, University of California Berkeley; Berkeley, CA, USA.

[8]Department of Electrical Engineering and Computer Sciences, University of California Berkeley; Berkeley, CA, USA.

[9]Ragon Institute of MGH, MIT and Harvard; Cambridge, MA, USA.

[10]Donor Network West; San Ramon, CA, USA.

[11]Department of Otolaryngology-Head and Neck Surgery, Stanford University School of Medicine; Stanford, California, USA.

[12]Department of Surgery, University of California San Francisco; San Francisco, CA, USA.

[13]Diabetes Center, University of California San Francisco; San Francisco, CA, USA.

[14]DCI Donor Services; Sacramento, CA, USA.

[15]Department of Pathology, Stanford University School of Medicine; Stanford, CA, USA.

[16]Department of Microbiology and Immunology, Stanford University School of Medicine; Stanford, CA, USA.

[17]Department of Pediatrics, Division of Pulmonary Medicine, Stanford University; Stanford, CA, USA.

[18]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine; Stanford, CA, USA.

[19]Department of Medicine, Division of Cardiovascular Medicine, Stanford University; Stanford, CA, USA.

[20]Department of Medicine and Liver Center, University of California San Francisco; San Francisco, CA, USA.

[21]Department of Neurology and Neurological Sciences, Stanford University School of Medicine; Stanford, CA, USA.

[22]Department of Surgery - Plastic and Reconstructive Surgery, Stanford University School of Medicine; Stanford, CA, USA.

[23]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA

[24]Paul F. Glenn Center for the Biology of Aging, Stanford University School of Medicine; Stanford, CA, USA.

[25]Department of Genetics, Stanford University School of Medicine; Stanford, CA, USA.

[26]Department of Biology, Stanford University; Stanford, CA, USA.

[27]Department of Pediatrics, Division of Cardiology, Stanford University School of Medicine; Stanford, CA, USA.

[28]Stanford Diabetes Research Center, Stanford University School of Medicine, Stanford, California

[29]Center for Gynecology and Reproductive Sciences, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco; San Francisco, CA, USA.

[30]Department of Critical Care Medicine, The First Affiliated Hospital of Shantou University Medical College; Shantou, China.

[31]Department of Ophthalmology, Stanford University School of Medicine; Stanford, CA, USA.

[32]Division of Nephrology, Department of Medicine, University of California San Francisco; San Francisco, CA, USA.

[33]Stanford University School of Medicine; Stanford, CA, USA.

[34]Mass Spectrometry Platform, Chan Zuckerberg Biohub; Stanford, CA, USA.

[35]Stanford Cancer Institute, Stanford University School of Medicine; Stanford, CA, USA.

[36]Department of Medicine, Division of Hematology, Stanford University School of Medicine, Stanford, CA, USA

[37]Department of Biochemistry and Biophysics, Cardiovascular Research Institute, University of California San Francisco; San Francisco, CA, USA.

[38]Division of Hematology and Oncology, Department of Medicine, Bakar Computational Health Sciences Institute, Institute for Human Genetics, University of California San Francisco; San Francisco, CA, USA.

[39]Stanford Cardiovascular Institute; Stanford CA, USA.

[40]Department of Chemical Engineering, Stanford University; Stanford, CA, USA.

[41]Department of Cell & Tissue Biology, University of California San Francisco; San Francisco, CA, USA.

[42]Institute for Computational and Mathematical Engineering, Stanford University; Stanford, CA, USA.

[43]Division of Infectious Diseases & Geographic Medicine, Department of Medicine, Stanford University School of Medicine; Stanford, CA, USA.

[44]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA

[45]Division of Nephrology, Stanford University School of Medicine; Stanford, CA, USA.

[46]Veterans Affairs Palo Alto Health Care System; Palo Alto, CA, USA.

[47]Vera Moulton Wall Center for Pulmonary and Vascular Disease, Stanford University School of Medicine; Stanford, CA, USA.

[48]Department of Urology, Stanford University School of Medicine, Stanford, CA, USA

[49]Division of Hematology/Oncology, Department of Medicine, San Francisco Veterans Affairs Health Care System, San Francisco, CA, USA.

[50]Department of Biochemistry, University of California San Francisco; San Francisco, CA, USA.

Figure Legends:

Figure 1. Overview of Tabula Sapiens. The Tabula Sapiens was constructed with data from 15 human donors; for detailed information on which tissues were examined for each donor please refer to **table S2**. Demographic and clinical information about each donor is listed in the supplement and in **table S1**. Donors 1, 2, 7 and 14 contributed the largest number of tissues each, and the number of cells from each tissue is indicated by the size of each circle. Tissue contributions from additional donors who contributed single or small numbers of tissues are shown in the "Additional donors" column, and the total number of cells for each organ are shown in the final column.

Figure 2. Comparison of single cell transcriptomics with conventional histology. Clinical pathology was performed on nine tissues from donors TSP2 and TSP13. **A**. Hematoxylin and eosin (H&E) stained image used for histology of the colon from TSP2, with compartments (solid, colored lines) and individual cell types (dashed black ellipses) identified by the pathologists. **B**. Coarse cell type representation of TSP2 as morphologically estimated by pathologists across several tissues, ordered by increasing heterogeneity of the tissue. Compartment colors are consistent between panels A and B.

Figure 3. Analysis of immune and endothelial cell types shared across tissues. A. Illustration of clonal distribution of T cells across multiple tissues. The majority of T cell clones are found in multiple tissues and represent a variety of T cell subtypes. **B.**

Prevalence of B cell isotypes across tissues, ordered by decreasing abundance of IgA. **C.** Expression level of tissue specific endothelial markers, shown as violin plots, in the entire dataset. Many of the markers are highly tissue specific, and typically derived from multiple donors as follows: bladder (3 donors), eye (2), fat (2), heart (1), liver (2), lung (3), mammary (1), muscle (4), pancreas (2), prostate (2), salivary gland (2), skin (2), thymus (2), tongue (2), uterus (1) and vasculature (2). A detailed donor-tissue breakdown is available in **table S2**.

Figure 4. Alternative splicing analysis. A,B. The sixth exon in MYL6 is skipped at different proportions in different compartments. Cells in the immune and epithelial compartments tend to skip the exon, whereas cells in the endothelial and stromal compartments tend to include the exon. Boxes are grouped by compartment and colored by tissue. The fraction of junctional reads that include exon 6 was calculated for each cell with more than 10 reads mapping to the exon skipping event. Horizontal box plots in **B** show the distribution of exon inclusion in each cell type. **C,D**. Alternative splicing in CD47 involves one 5' splice site (exon 11, 108,047,292) and four 3' splice sites. Horizontal box plots in **D** show the distribution of weighted averages of alternative 3' splice sites in each cell type. Epithelial cells tend to use closer exons to the 5' splice site compared to immune and stromal cells. Boxes are grouped by compartment and colored by tissue.

Figure 5. Dynamic changes in cell state. A. Cell types ordered by magnitude of cell cycling index, per donor (each a separate color) with the most highly proliferative at the

top and quiescent cells at the bottom of the list. **B**. RNA velocity analysis demonstrating mesenchymal to myofibroblast transition in the bladder. The arrows represent a flow derived from the ratio of unspliced to spliced transcripts which in turn predicts dynamic changes in cell identity. **C**,**D.** Latent time analysis of the mesenchymal to myofibroblast transition in the bladder demonstrating stereotyped changes in gene expression trajectory.

Figure 6. High-resolution view highlights patchiness of the gut microbiome. A. Schematic (left) and photo of the colon from donor TSP2 (right), with numbers 1-5 representing microbiota sampling locations. **B**. Relative abundances and richness (number of observed species) at the family level in each sampling location, as determined by 16S rRNA sequencing. The Shannon diversity, a metric of evenness, mimics richness. Variability in relative abundance and/or richness/Shannon diversity was higher in the duodenum, jejunum and ileum as compared with the ascending and sigmoid colon. **C**. A Sankey diagram showing the inflow and outflow of microbial species from each section of the gastrointestinal tract. The stacked bar for each gastrointestinal section represents the number of observed species in each family as the union of all sampling locations for that section. The stacked bar flowing out represents gastrointestinal species not found in the subsequent section and the stacked bar flowing into each gastrointestinal section represents the species not found in the previous section. **D**. UMAP clustering of T cells reveals distinct transcriptome profiles in the distal and proximal small and large intestines. **E**. Dots in volcano plot highlight

genes up-regulated in the large (left) and small (right) intestines**.** Labeled dots include genes with known roles in trafficking, survival, and activation.

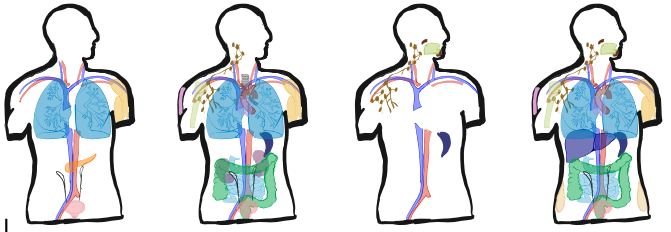
Supplementary Material

Materials and Methods
Figs. S1 to S16
Tables S1 to S9
References (40-95)

Resources

Data portal for Tabula Sapiens (*14*)
Code for the analysis (*36*)
Single cell gene counts and metadata (*37*)
Histology images (*38*)
SpliZ scores (*39*)

Fig. 1



	Cell count
Bladder	21,517
Blood	50,115
Bone Marrow	12,297
Eye	10,650
Fat	20,263
Heart	11,505
Kidney	9,641
Large Intestine	13,680
Liver	5,007
Lung	35,682
Lymph Node	53,275
Mammary Gland	11,375
Muscle	30,746
Pancreas	13,497
Prostate	16,375
Salivary Gland	27,199
Skin	9,424
Small Intestine	12,467
Spleen	34,004
Thymus	33,664
Tongue	15,020
Trachea	9,522
Uterus	7,124
Vasculature	17,071

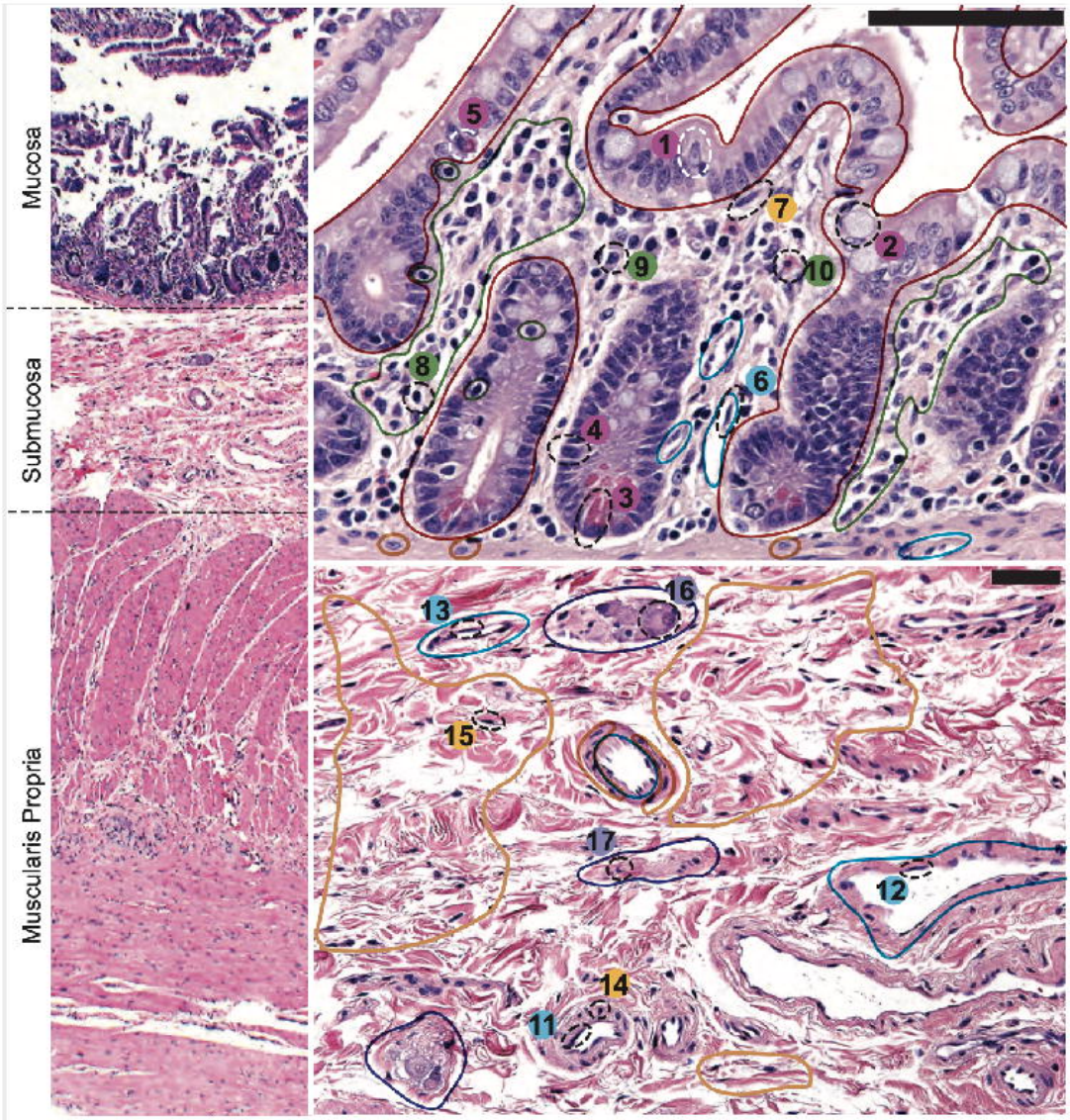
TSP1 | TSP2 | TSP7 | TSP14 | Additional 11 Donors

Number of cells
6000
12000
18000
24000
30000

Total 481,120 cells

Fig. 2

A


Mucosa
Submucosa
Muscularis Propria
Epithelial
1. Absorptive enterocyte
2. Goblet cell
3. Paneth cell
4. Crypt enterocyte
5. Enteroendocrine cell
Endothelial
6. Endothelial cell (capillary or lymphatic)
Stromal
7. Fibroblast
Immune
8. Lymphocyte
9. Plasma cell
10. Eosinophil
Endothelial
11. Endothelial cell (arteriole)
12. Endothelial cell (venule)
13. Endothelial cell (capillary)
Stromal
14. Smooth muscle cell (arteriole)
15. Fibroblast
PNS/Neuroglial
16. Ganglion cell (Meissner plexus)
17. Schwann cell

B

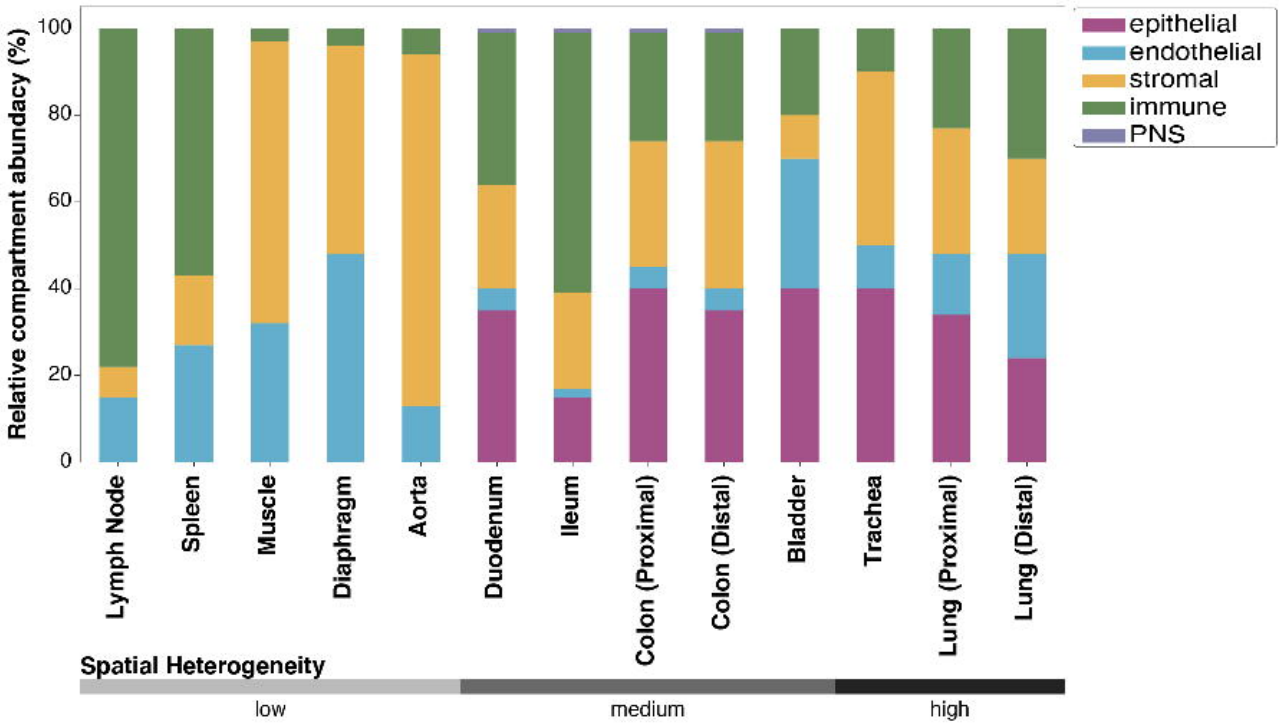

Relative compartment abundacy (%)
0
20
40
60
80
100
epithelial
endothelial
stromal
immune
PNS
Lymph Node
Spleen
Muscle
Diaphragm
Aorta
Duodenum
Ileum
Colon (Proximal)
Colon (Distal)
Bladder
Trachea
Lung (Proximal)
Lung (Distal)
Spatial Heterogeneity
low
medium
high

Fig. 3

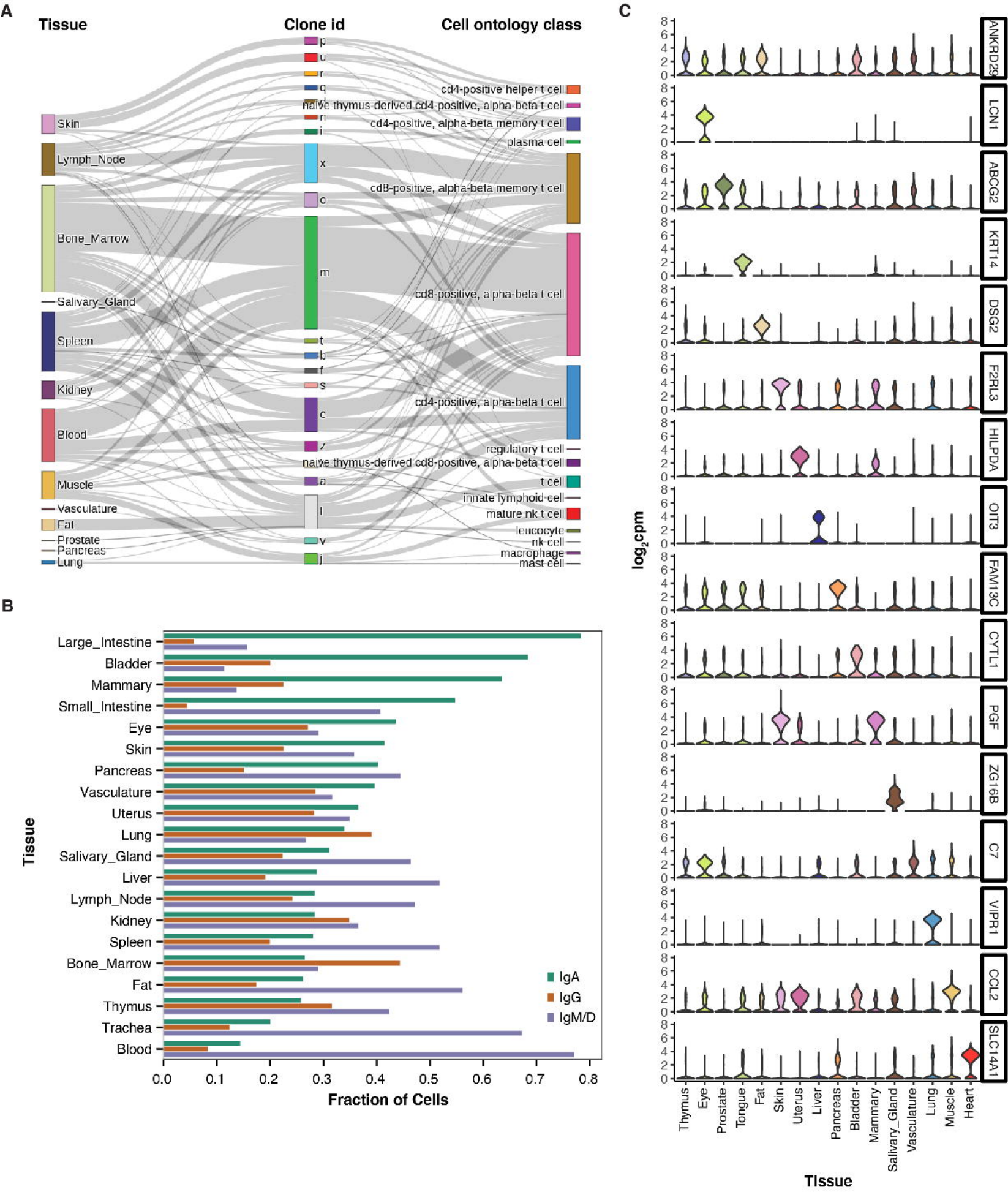
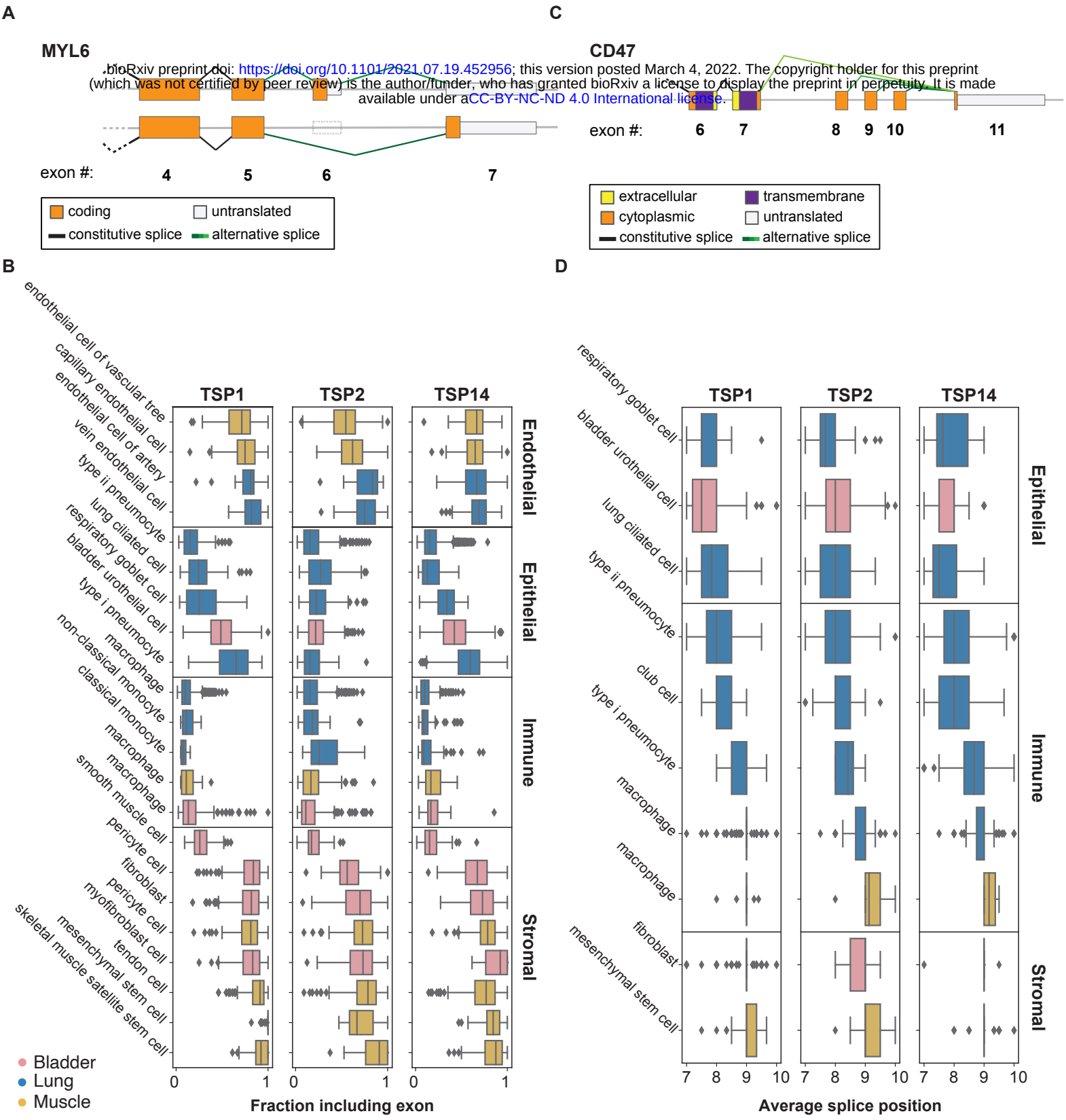

A
Tissue
Clone id
Cell ontology class
Skin
Lymph_Node
Bone_Marrow
Salivary_Gland
Spleen
Kidney
Blood
Muscle
Vasculature
Fat
Prostate
Pancreas
Lung
p
u
r
q
n
i
x
c
m
t
b
f
s
e
z
a
l
v
j
cd4-positive helper t cell
naive thymus-derived cd4-positive, alpha-beta t cell
cd4-positive, alpha-beta memory t cell
plasma cell
cd8-positive, alpha-beta memory t cell
cd8-positive, alpha-beta t cell
cd4-positive, alpha-beta t cell
regulatory t cell
naive thymus-derived cd8-positive, alpha-beta t cell
t cell
innate lymphoid cell
mature nk t cell
leucocyte
nk cell
macrophage
mast cell
B
Tissue
Large_Intestine
Bladder
Mammary
Small_Intestine
Eye
Skin
Pancreas
Vasculature
Uterus
Lung
Salivary_Gland
Liver
Lymph_Node
Kidney
Spleen
Bone_Marrow
Fat
Thymus
Trachea
Blood
IgA
IgG
IgM/D
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Fraction of Cells
C
log2cpm
ANKRD29
LCN1
ABCG2
KRT14
DSG2
F2RL3
HILPDA
OIT3
FAM13C
CYTL1
PGF
ZG16B
C7
VIPR1
CCL2
SLC14A1
Thymus
Eye
Prostate
Tongue
Fat
Skin
Uterus
Liver
Pancreas
Bladder
Mammary
Salivary_Gland
Vasculature
Lung
Muscle
Heart
TIssue

Fig. 4



Mesenchymal to myofibroblas transdifferen-
tiation (bladder)

Fig. 5

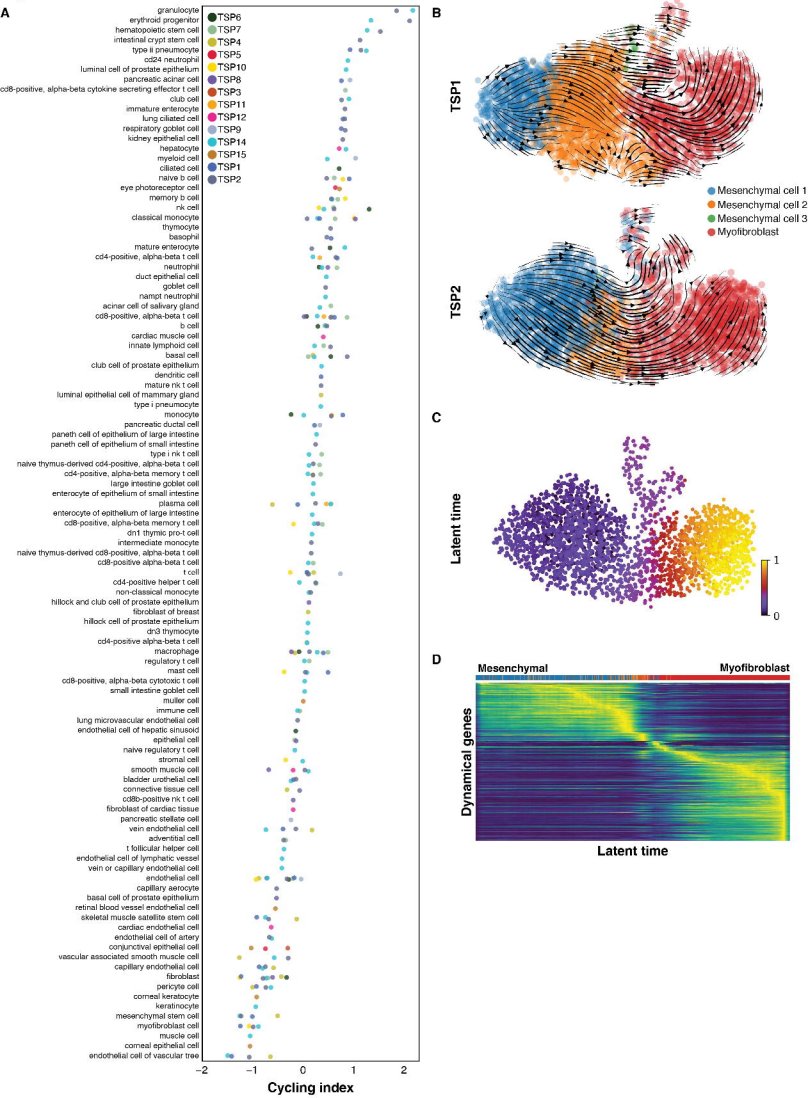


Fig. 6

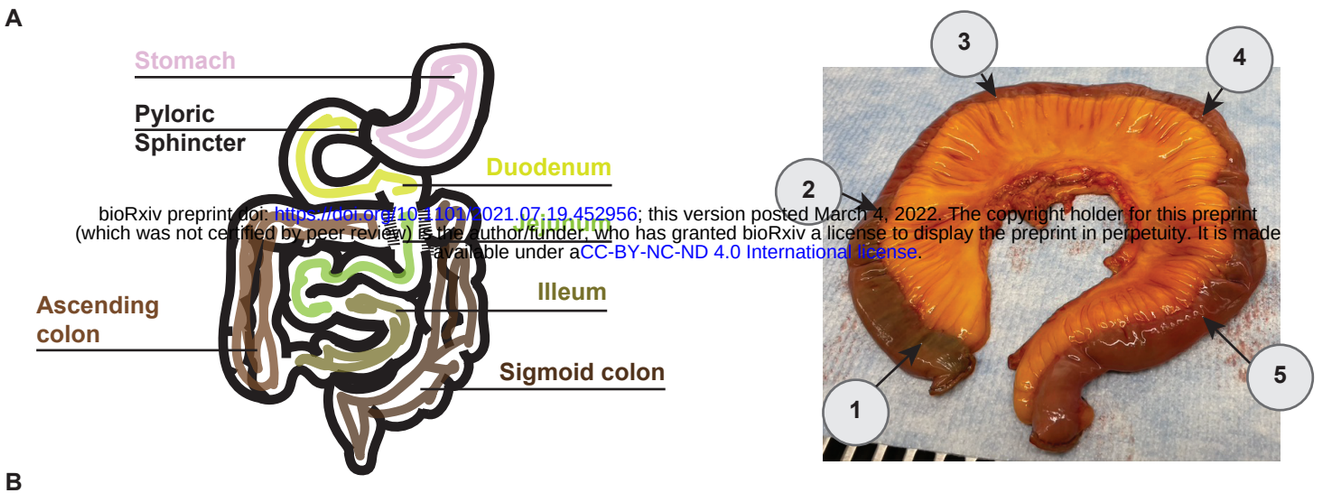


Fig. S1

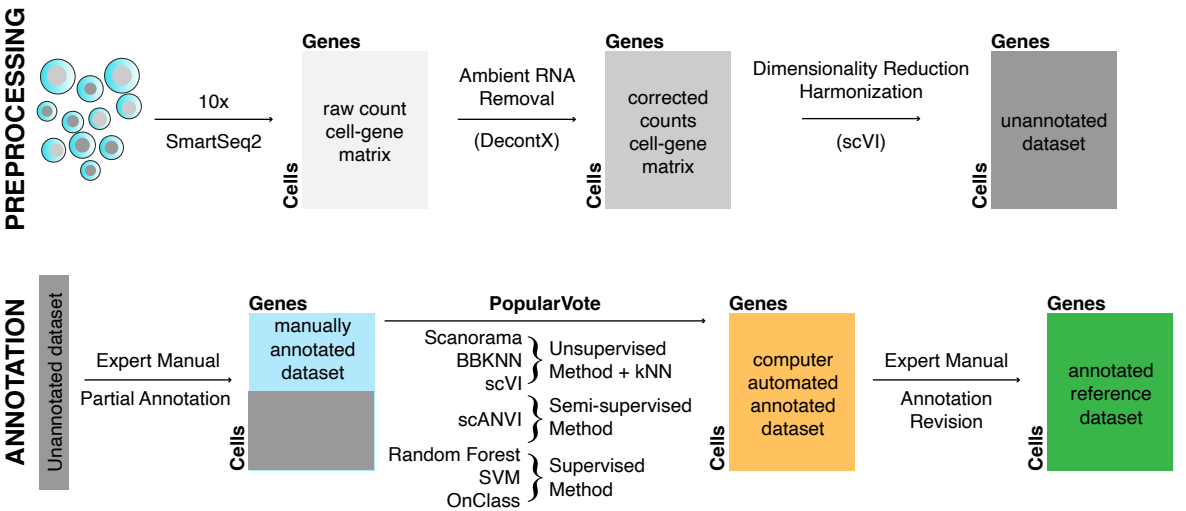
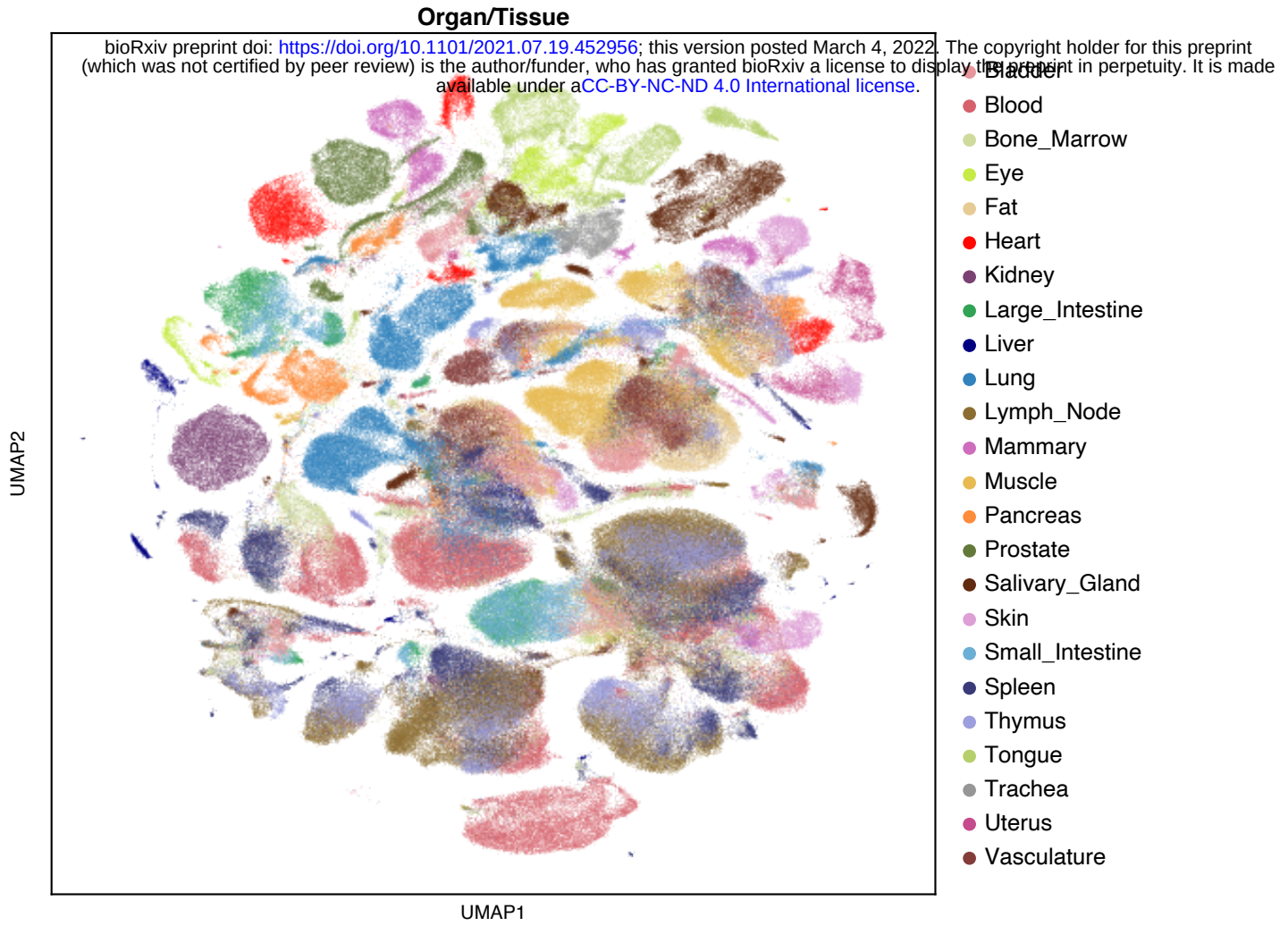
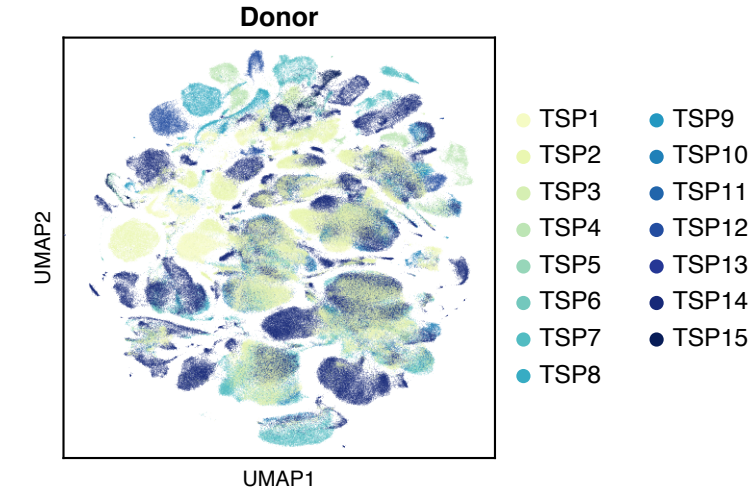

PREPROCESSING
10x
SmartSeq2
Genes
Cells
raw count
cell-gene
matrix
Ambient RNA
Removal
(DecontX)
Genes
Cells
corrected
counts
cell-gene
matrix
Dimensionality Reduction
Harmonization
(scVI)
Genes
Cells
unannotated
dataset
ANNOTATION
Unannotated dataset
Expert Manual
Partial Annotation
Genes
Cells
manually
annotated
dataset
PopularVote
Scanorama
BBKNN
scVI
Unsupervised
Method + kNN
scANVI
Semi-supervised
Method
Random Forest
SVM
OnClass
Supervised
Method
Genes
Cells
computer
automated
annotated
dataset
Expert Manual
Annotation
Revision
Genes
Cells
annotated
reference
dataset

Fig. S2

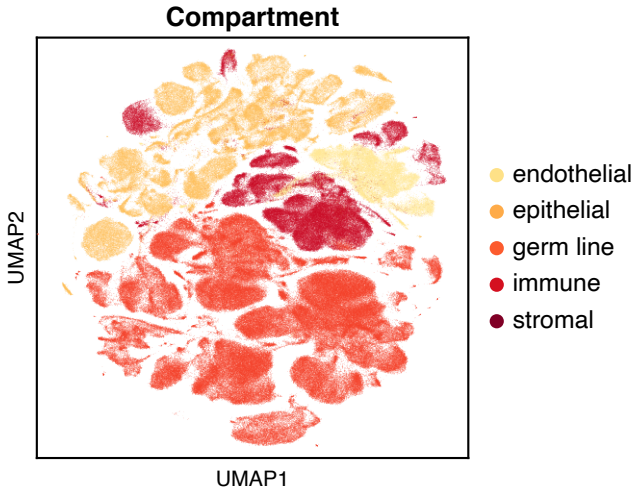
A



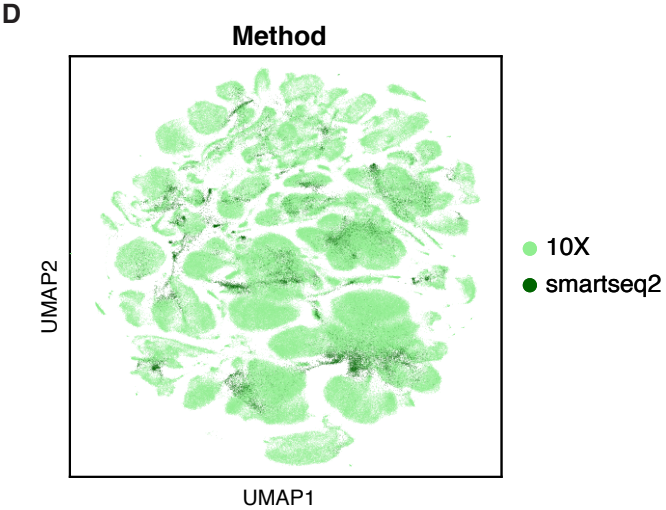
B



C



D



E

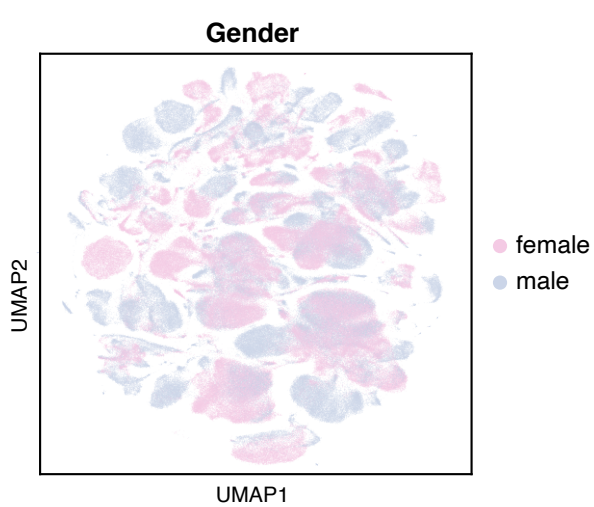


Fig. S3

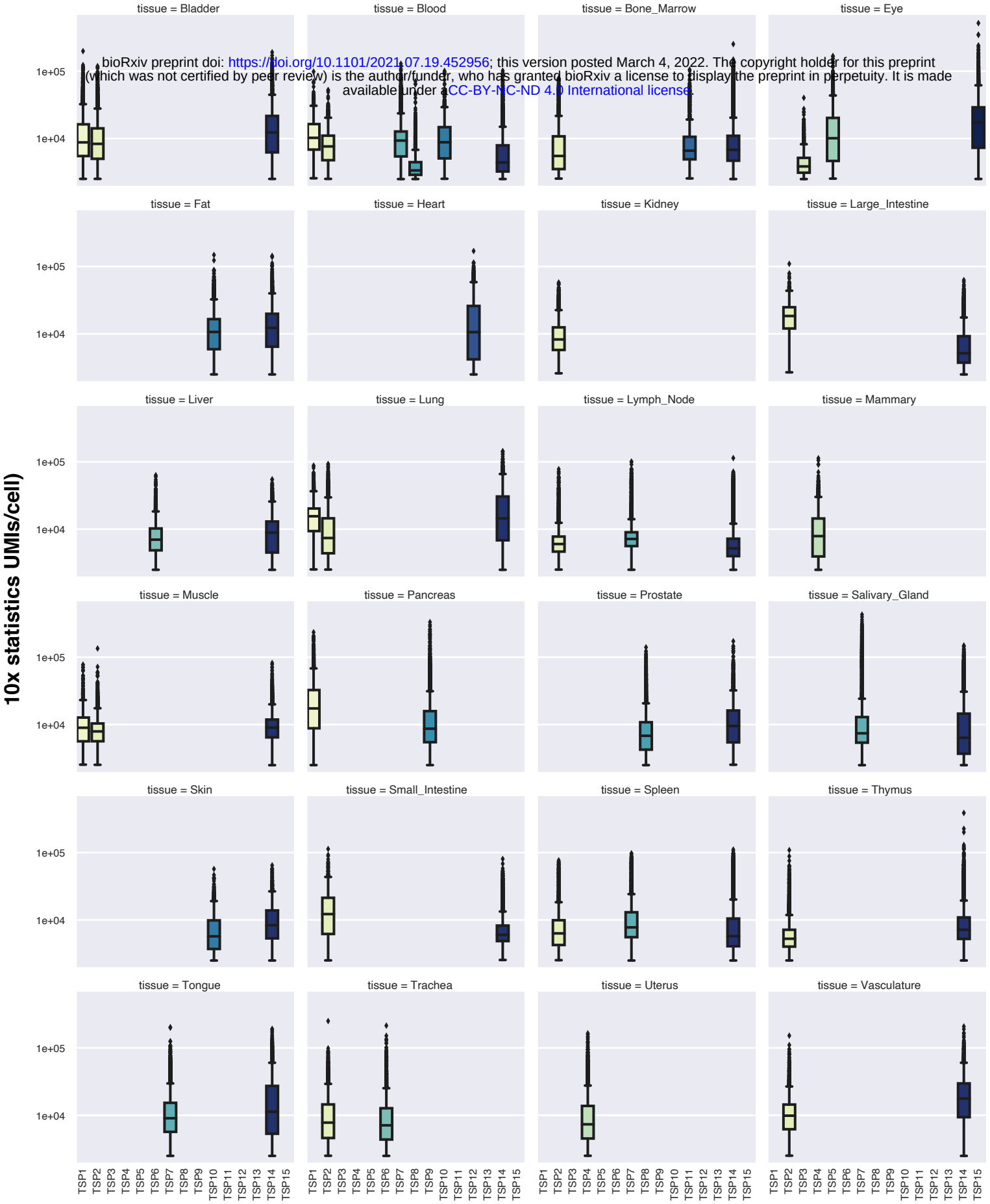


Fig. S4

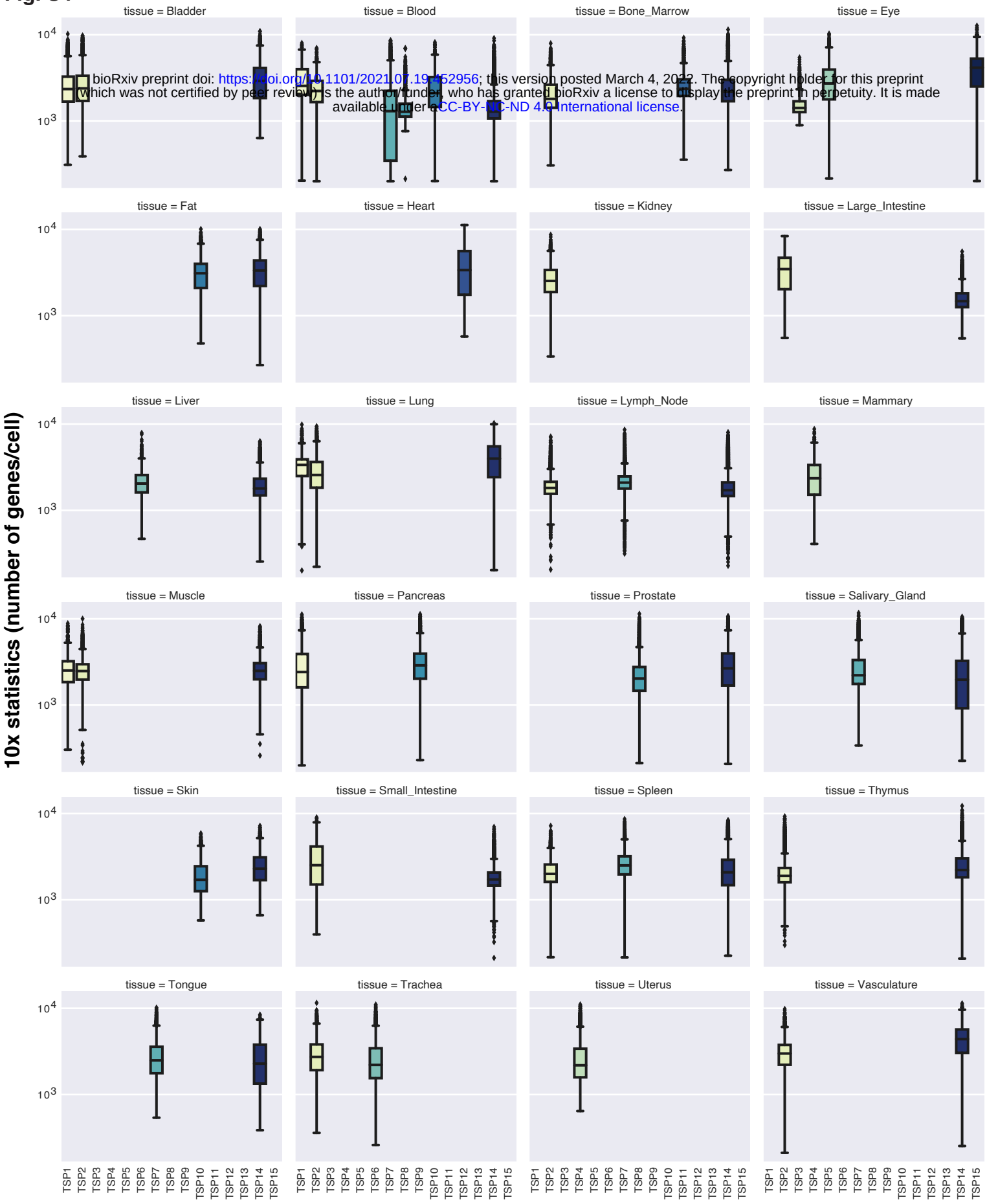


Fig. S5

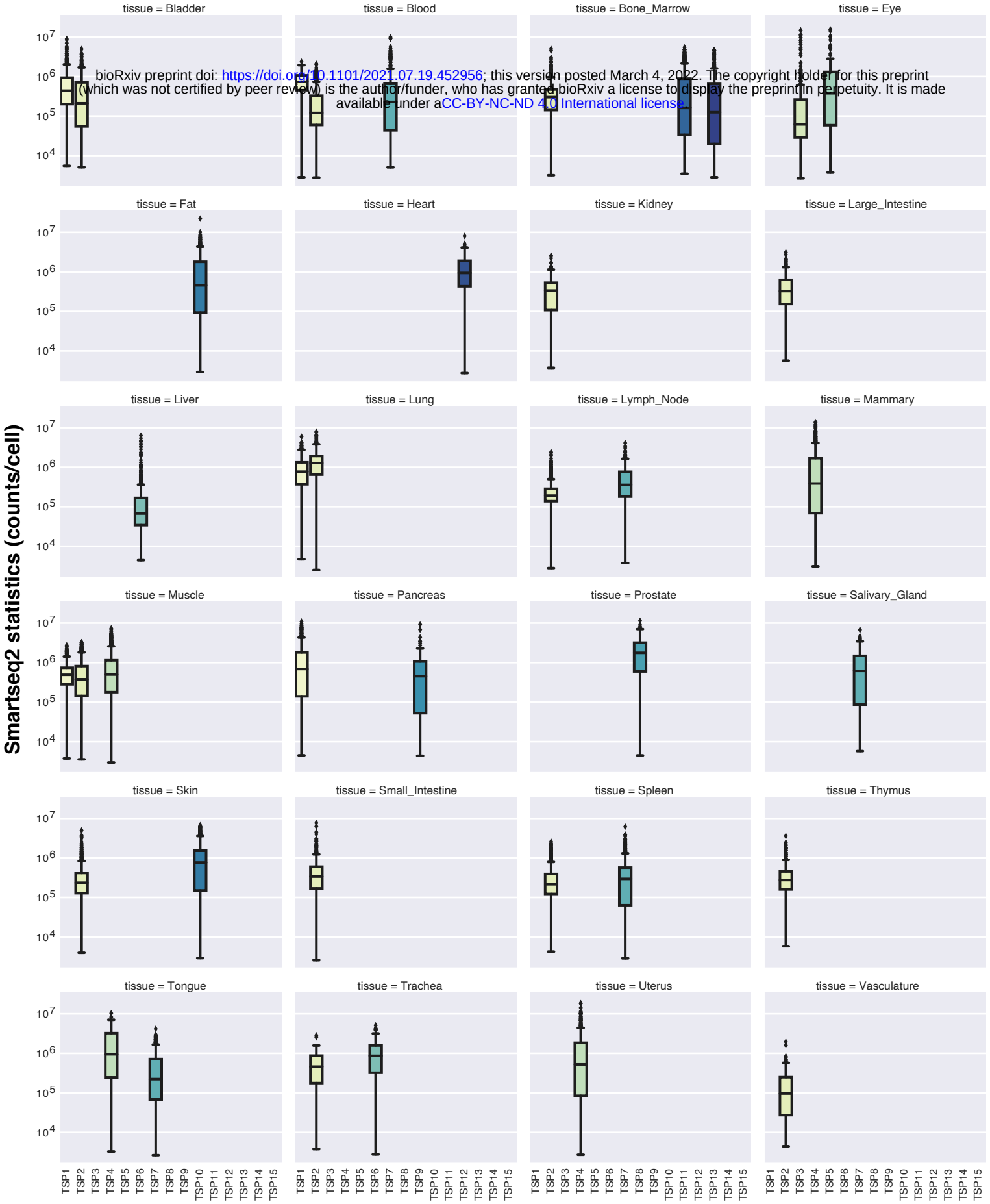


Fig. S6

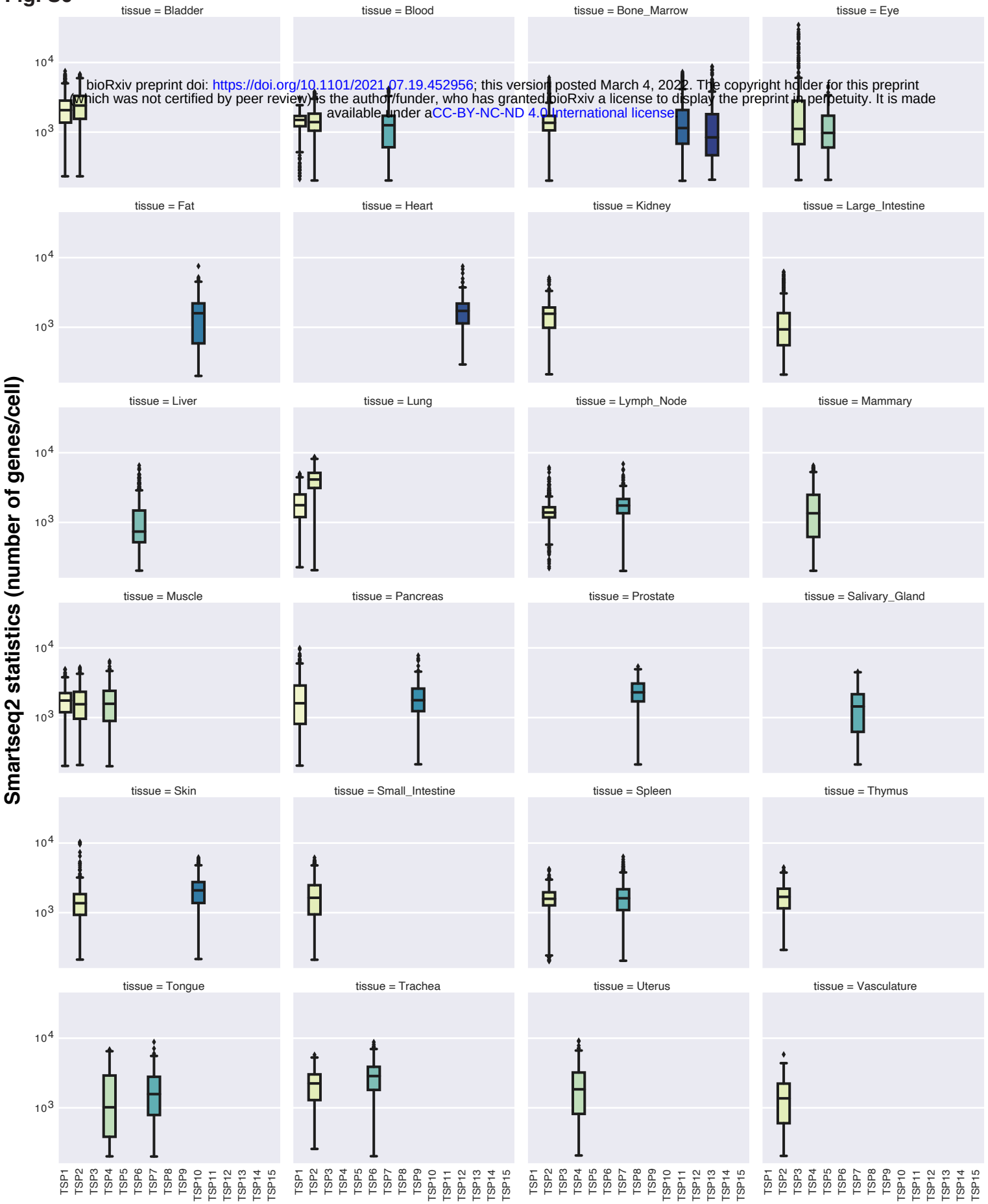


Fig. S7

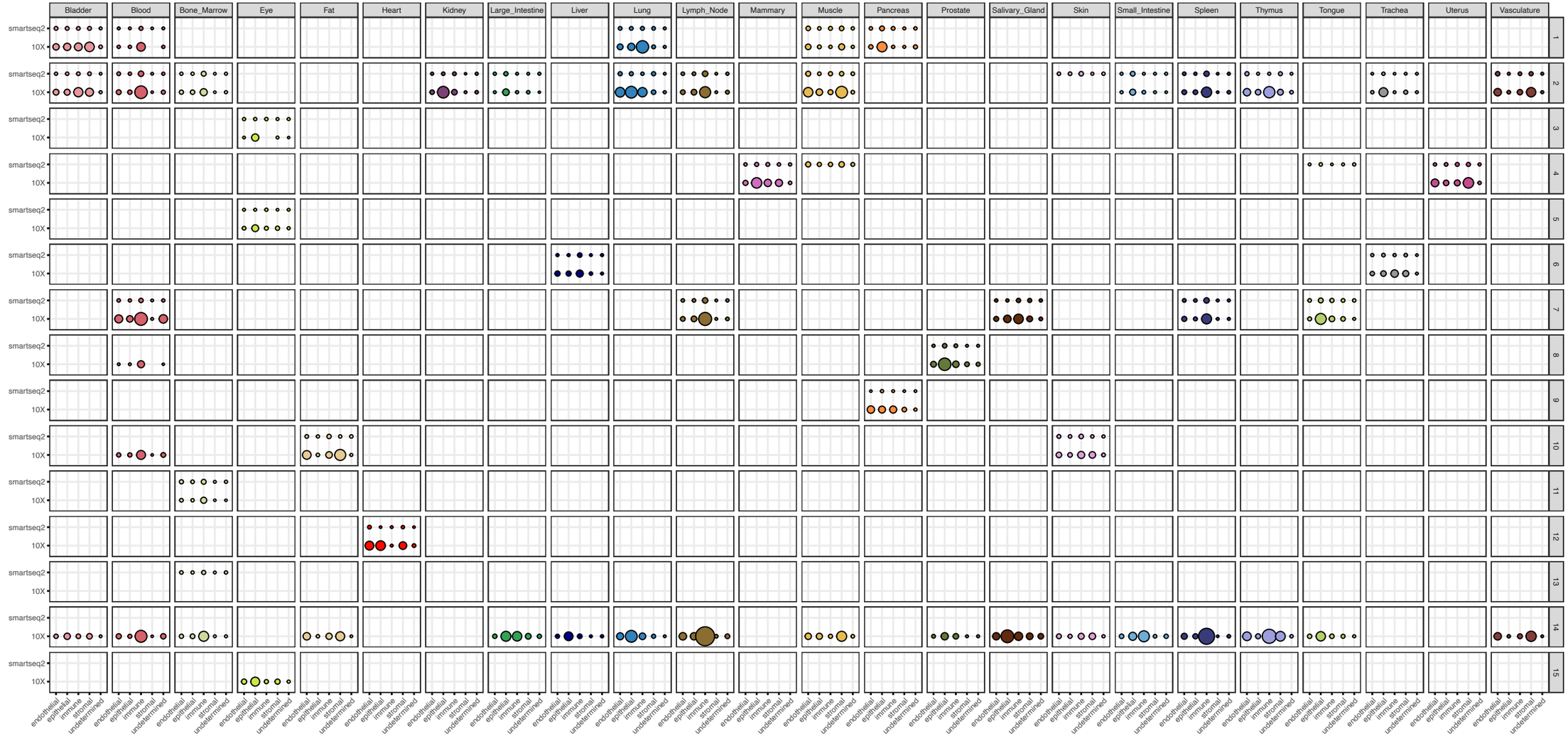

Bladder
Blood
Bone_Marrow
Eye
Fat
Heart
Kidney
Large_Intestine
Liver
Lung
Lymph_Node
Mammary
Muscle
Pancreas
Prostate
Salivary_Gland
Skin
Small_Intestine
Spleen
Thymus
Tongue
Trachea
Uterus
Vasculature
smartseq2
10X
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
endothelial
epithelial
immune
stromal
undetermined
Tissue/Organ
Bladder
Blood
Bone_Marrow
Eye
Fat
Heart
Kidney
Large_Intestine
Liver
Lung
Lymph_Node
Mammary
Muscle
Pancreas
Prostate
Salivary_Gland
Skin
Small_Intestine
Spleen
Thymus
Tongue
Trachea
Uterus
Vasculature
Number of cells
25000
20000
15000
10000
5000

Fig. S8

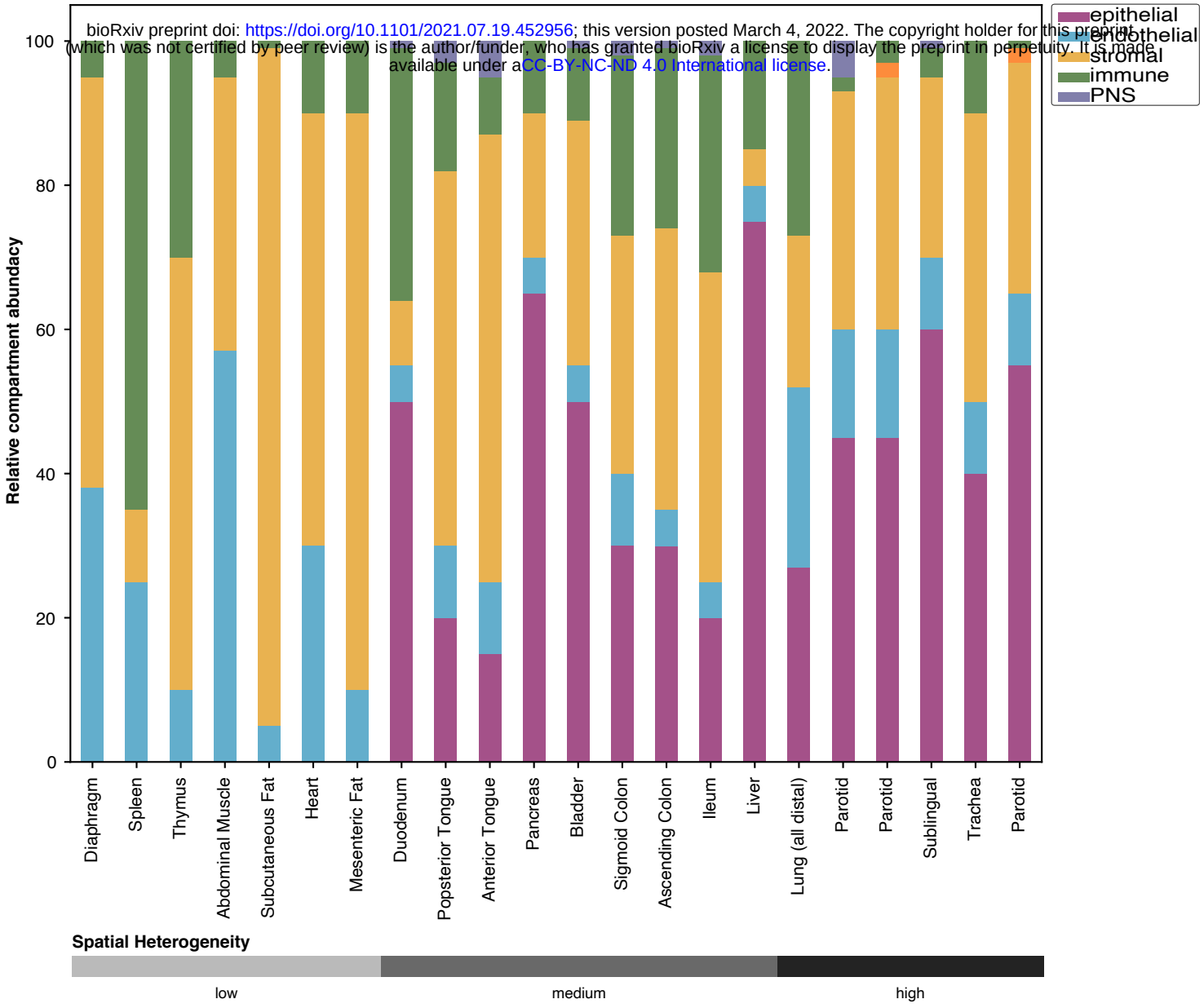


Fig. S9

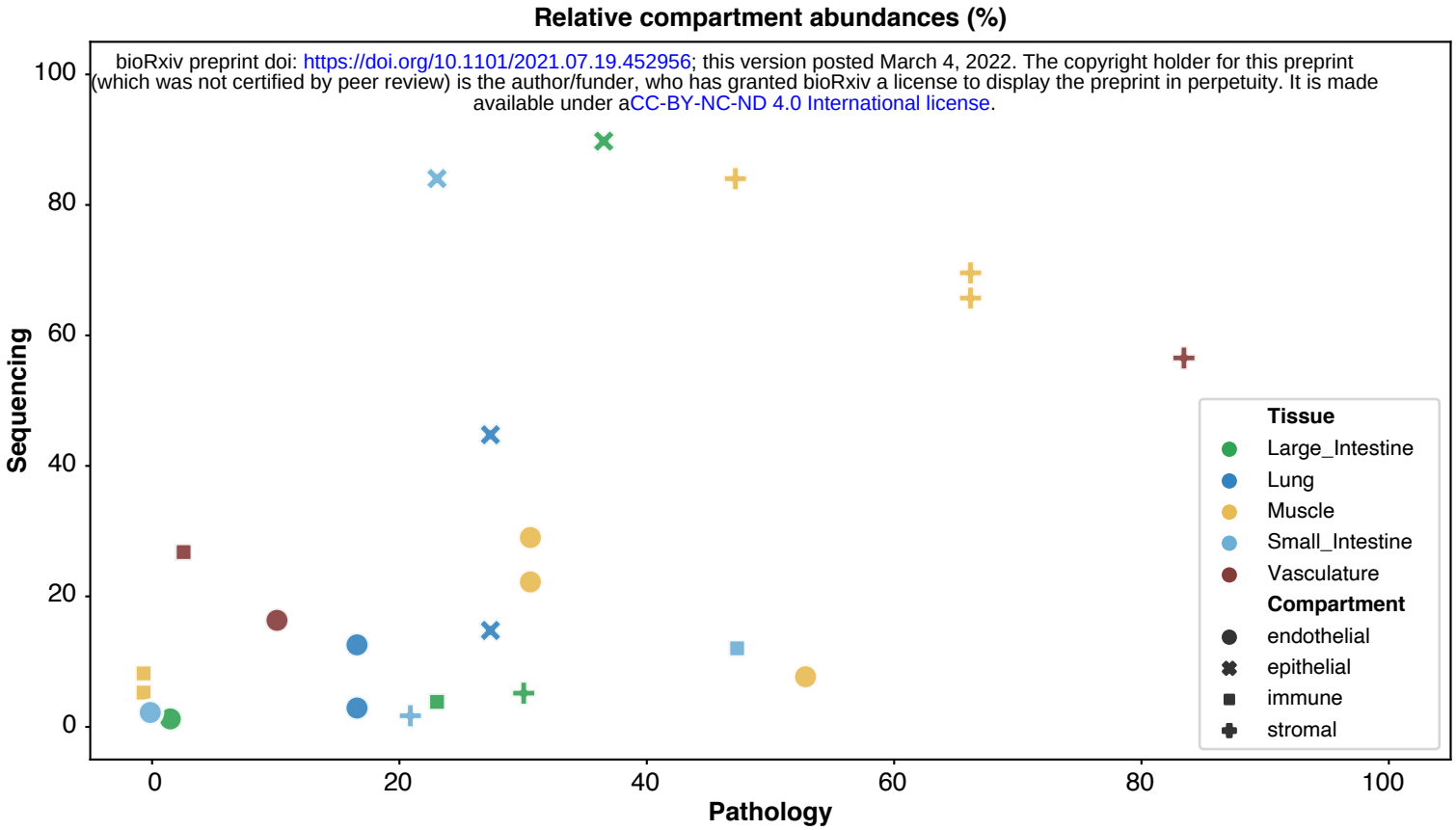


Fig. S10

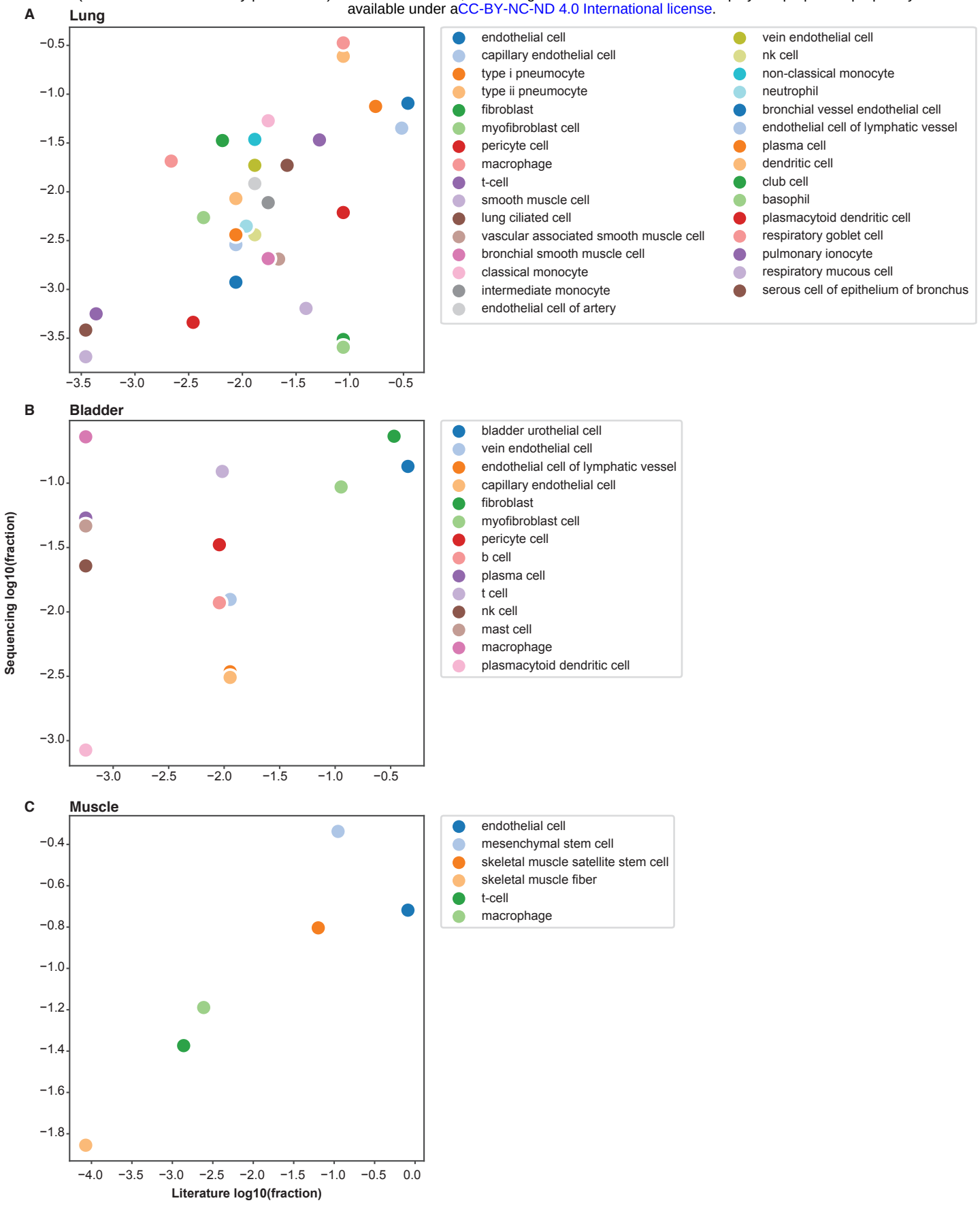


Fig. S11

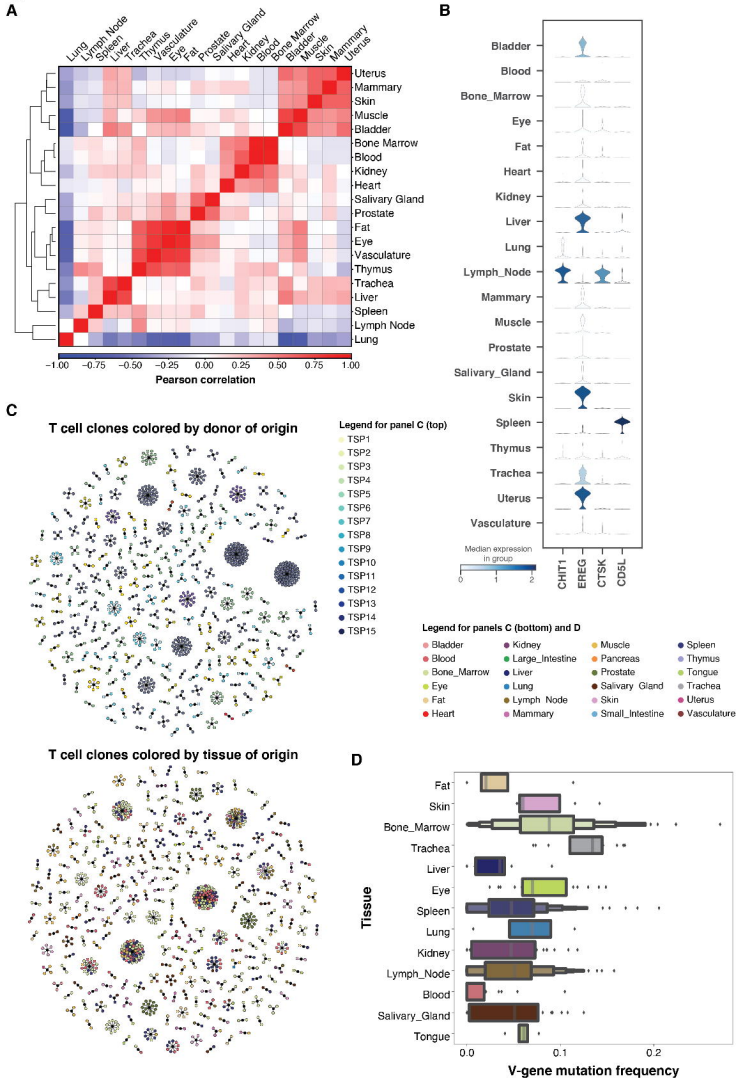

A
Lung
Lymph Node
Spleen
Liver
Trachea
Thymus
Vasculature
Eye
Fat
Prostate
Salivary Gland
Heart
Kidney
Blood
Bone Marrow
Bladder
Muscle
Skin
Mammary
Uterus
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00
Pearson correlation
B
CHIT1
EREG
CTSK
CD5L
Median expression in group
C
T cell clones colored by donor of origin
Legend for panel C (top)
TSP1
TSP2
TSP3
TSP4
TSP5
TSP6
TSP7
TSP8
TSP9
TSP10
TSP11
TSP12
TSP13
TSP14
TSP15
T cell clones colored by tissue of origin
Legend for panels C (bottom) and D
Bladder
Blood
Bone_Marrow
Eye
Fat
Heart
Kidney
Liver
Lung
Lymph Node
Mammary
Muscle
Pancreas
Prostate
Salivary Gland
Skin
Small_Intestine
Spleen
Thymus
Tongue
Trachea
Uterus
Vasculature
D
Tissue
V-gene mutation frequency

Fig. S12

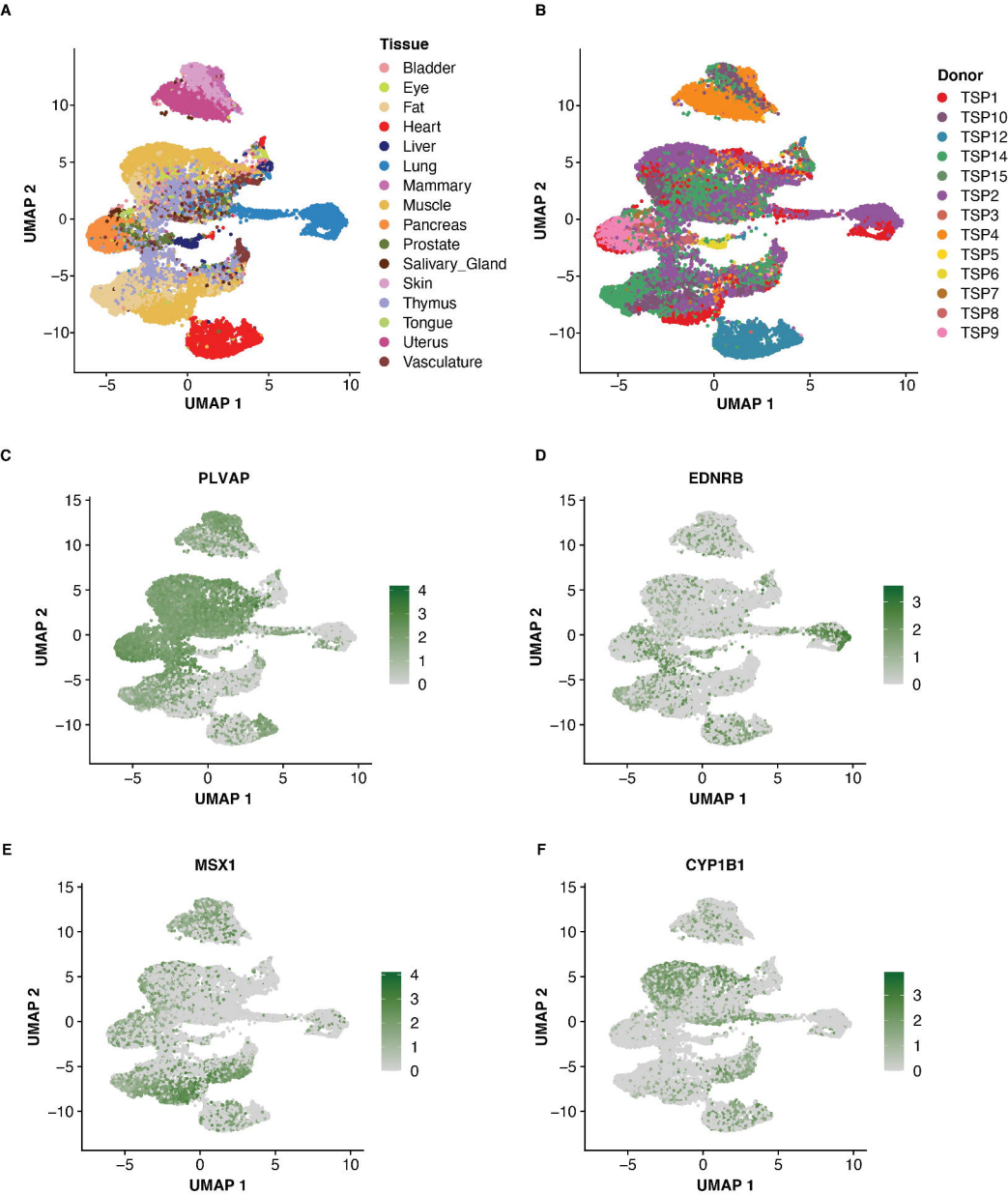
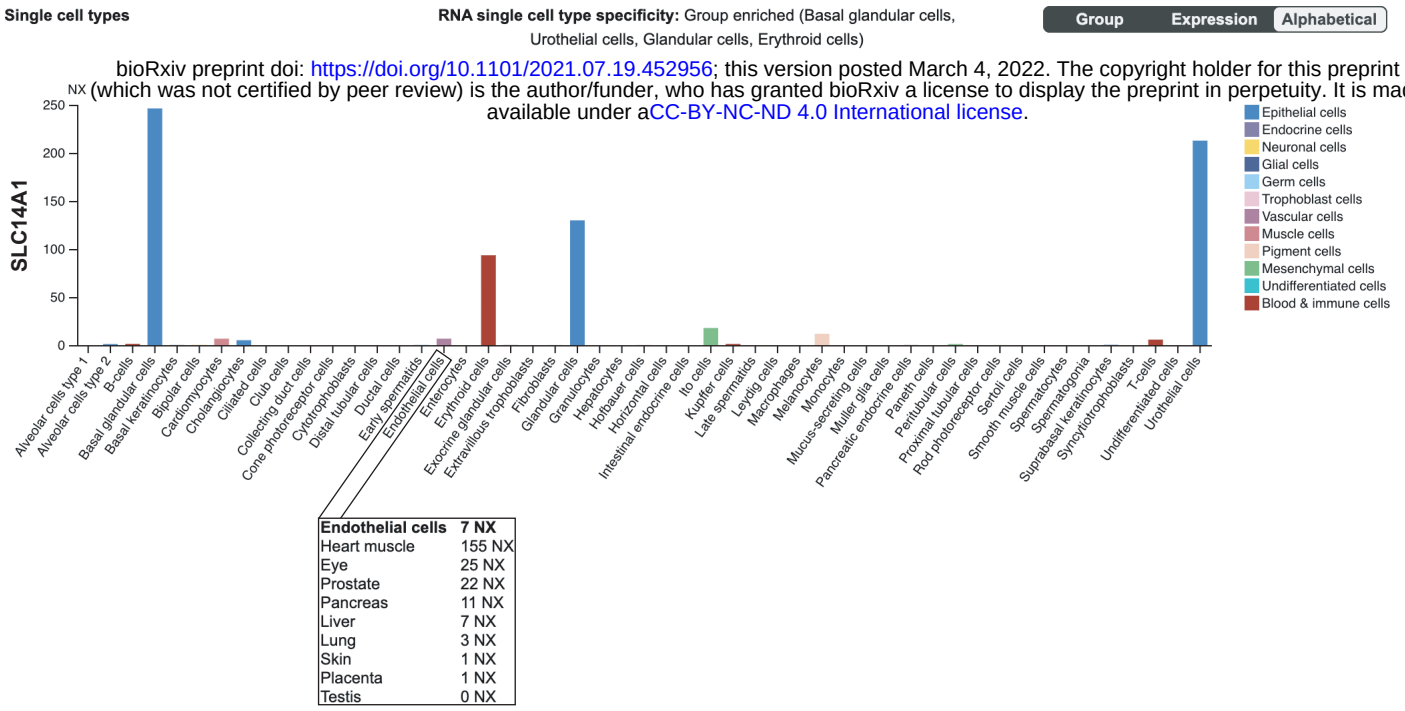

A
Tissue
Bladder
Eye
Fat
Heart
Liver
Lung
Mammary
Muscle
Pancreas
Prostate
Salivary_Gland
Skin
Thymus
Tongue
Uterus
Vasculature
UMAP 1
UMAP 2
B
Donor
TSP1
TSP10
TSP12
TSP14
TSP15
TSP2
TSP3
TSP4
TSP5
TSP6
TSP7
TSP8
TSP9
C
PLVAP
D
EDNRB
E
MSX1
F
CYP1B1

Fig. S13

A



B

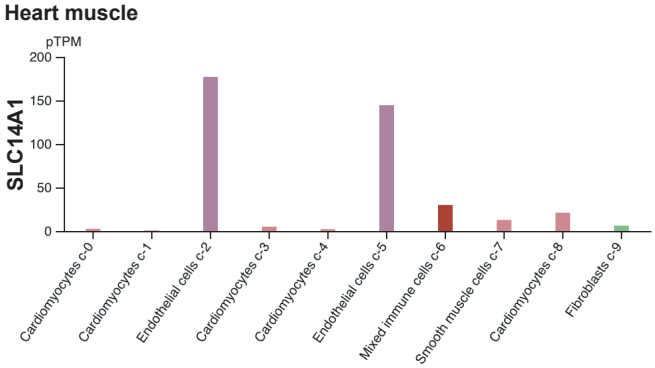
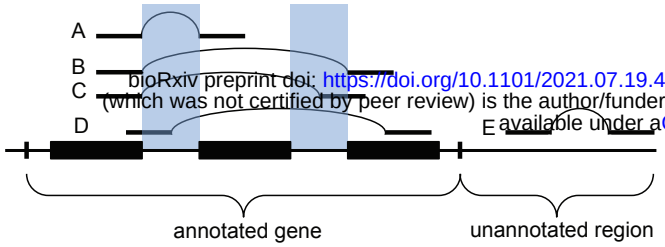


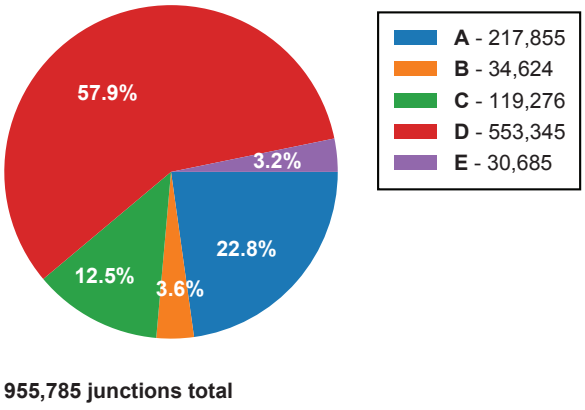
Fig. S14

A



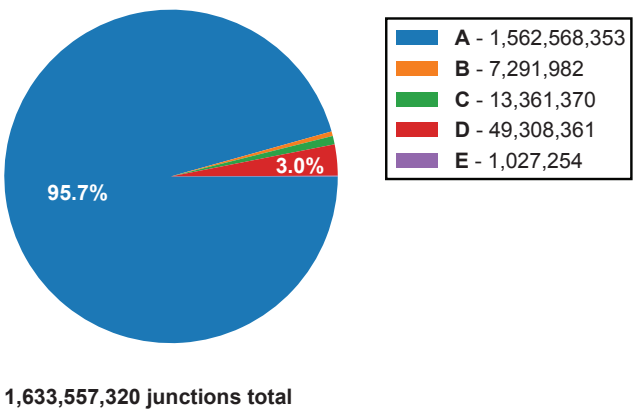
A: annotated junction
B: Both boundaries are annotated but the splice is not
C: One boundary is annotated but the other is not
D: Neither boundary is annotated but at least one end is in a known gene
E: Neither boundary is in a known gene

B



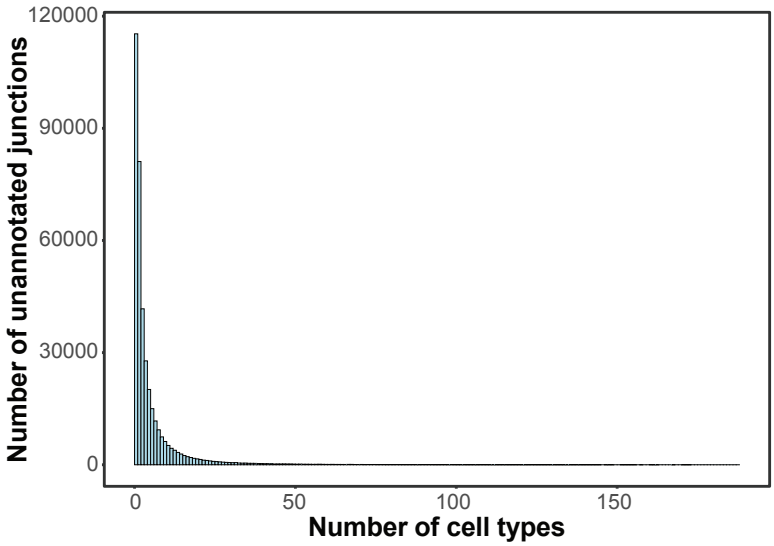
955,785 junctions total

C

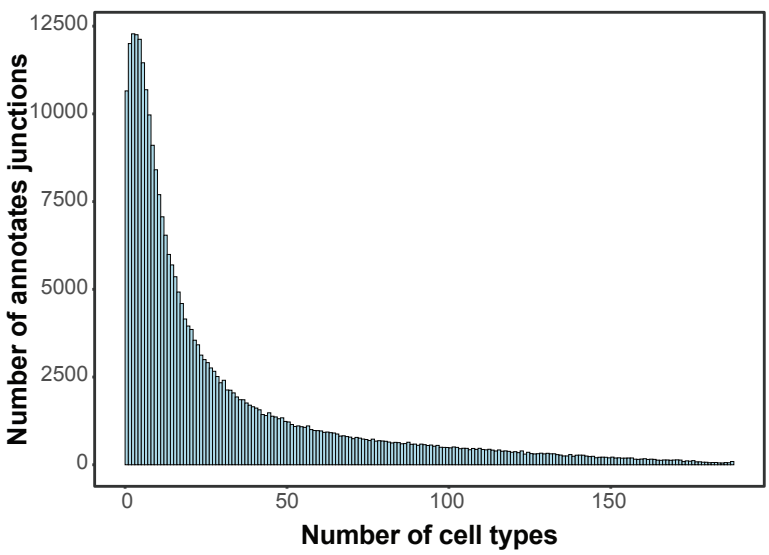


1,633,557,320 junctions total

D



E



F

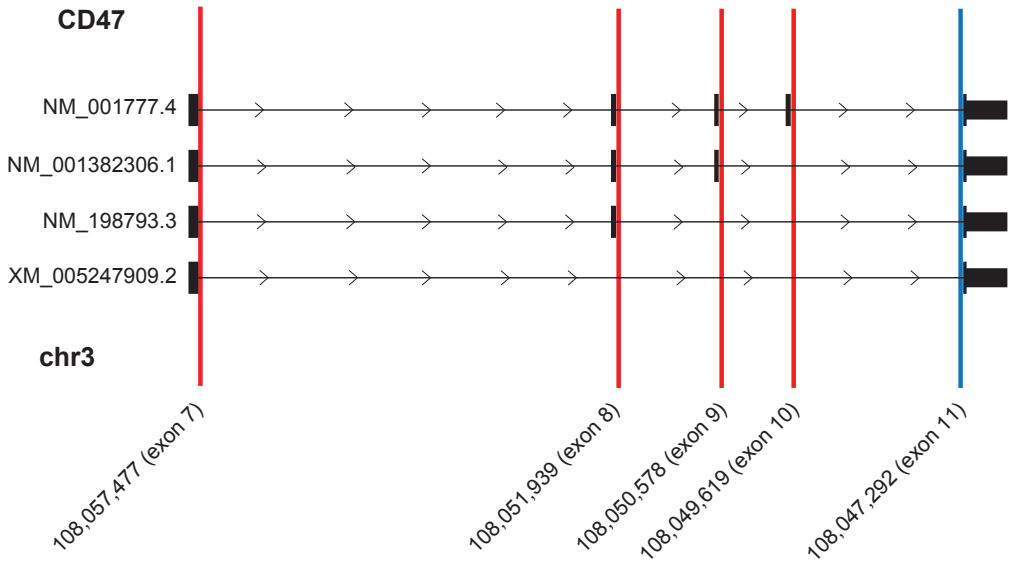


Fig. S15

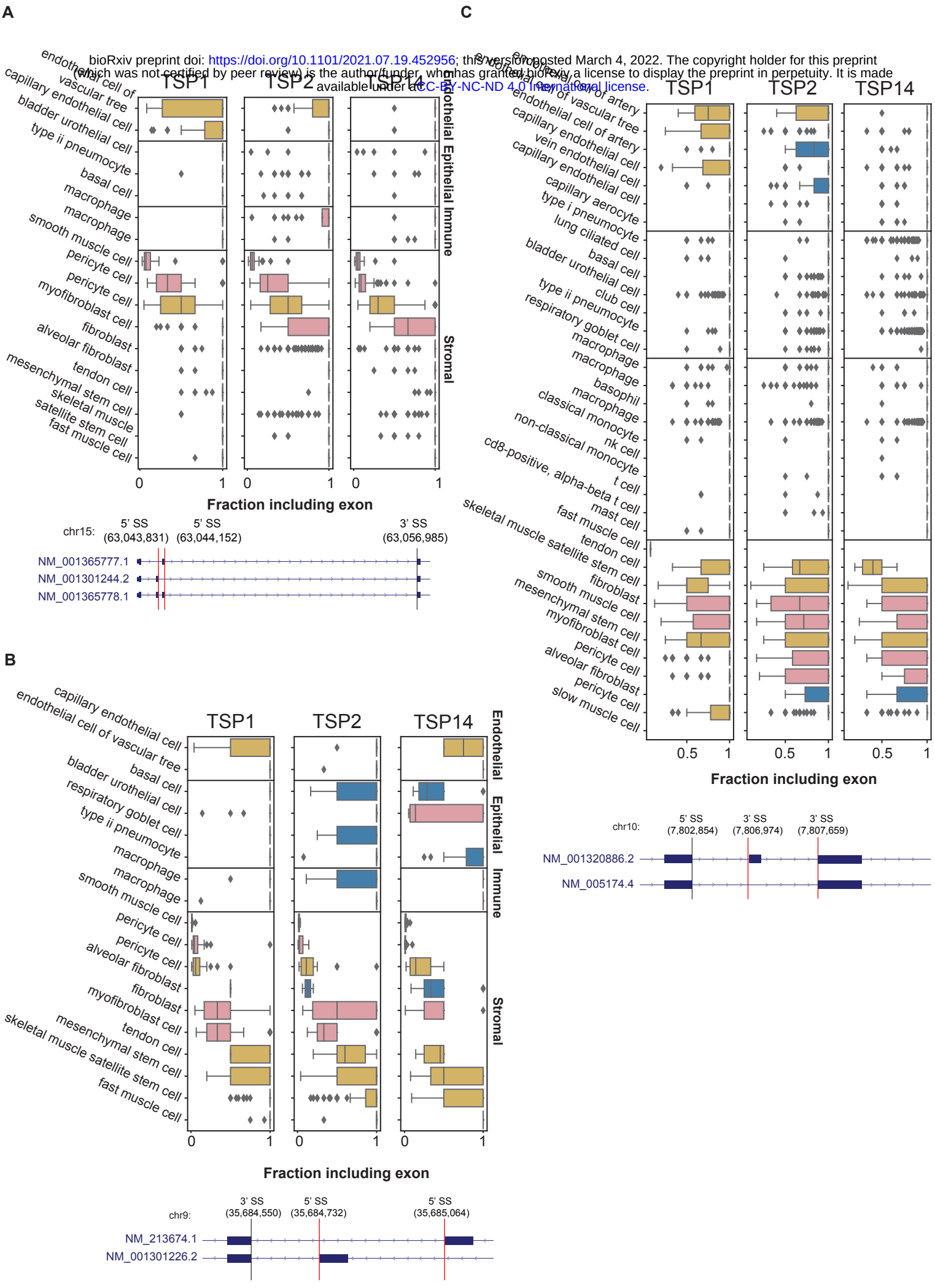


Fig. S16

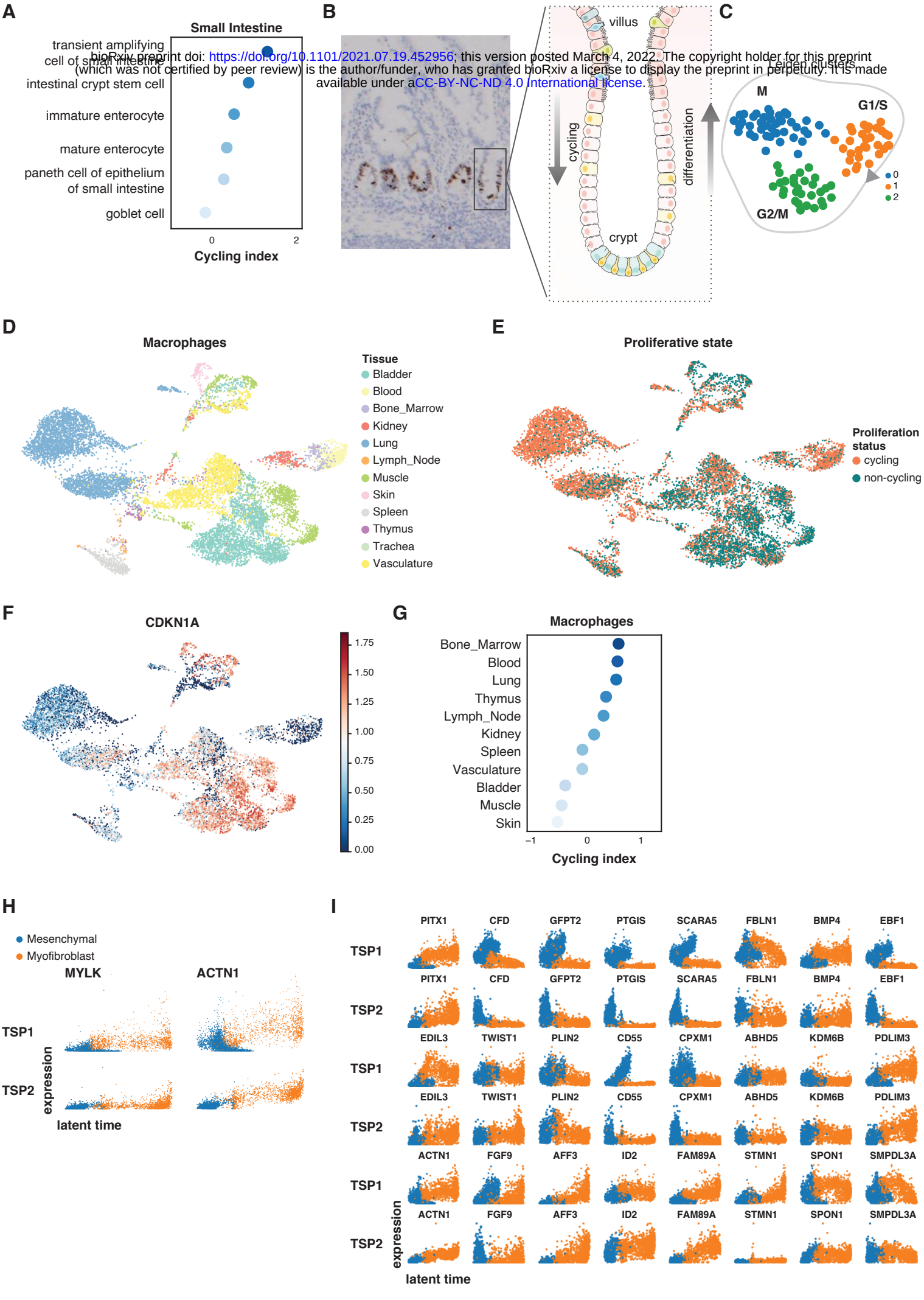
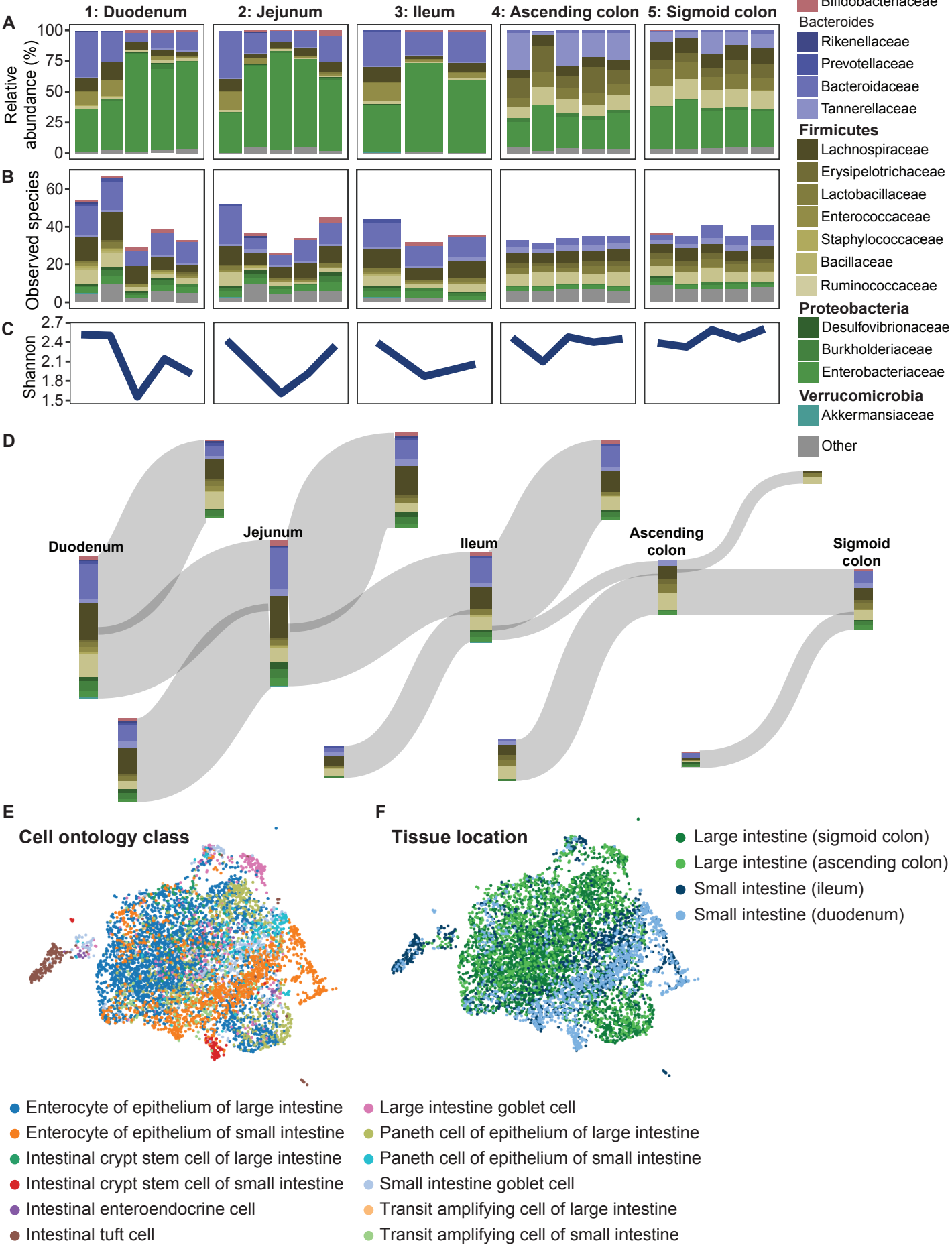


Fig. S17


1: Duodenum
2: Jejunum
3: Ileum
4: Ascending colon
5: Sigmoid colon
A
Relative abundance (%)
100
75
50
25
0
B
Observed species
60
40
20
0
C
Shannon
2.7
2.4
2.1
1.8
1.5
Actinobacteria
Bifidobacteriaceae
Bacteroides
Rikenellaceae
Prevotellaceae
Bacteroidaceae
Tannerellaceae
Firmicutes
Lachnospiraceae
Erysipelotrichaceae
Lactobacillaceae
Enterococcaceae
Staphylococcaceae
Bacillaceae
Ruminococcaceae
Proteobacteria
Desulfovibrionaceae
Burkholderiaceae
Enterobacteriaceae
Verrucomicrobia
Akkermansiaceae
Other
D
Duodenum
Jejunum
Ileum
Ascending colon
Sigmoid colon
E
Cell ontology class
F
Tissue location
Large intestine (sigmoid colon)
Large intestine (ascending colon)
Small intestine (ileum)
Small intestine (duodenum)
Enterocyte of epithelium of large intestine
Enterocyte of epithelium of small intestine
Intestinal crypt stem cell of large intestine
Intestinal crypt stem cell of small intestine
Intestinal enteroendocrine cell
Intestinal tuft cell
Large intestine goblet cell
Paneth cell of epithelium of large intestine
Paneth cell of epithelium of small intestine
Small intestine goblet cell
Transit amplifying cell of large intestine
Transit amplifying cell of small intestine