Title: 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: (~125 words) Molecular characterization of cell types using single cell transcriptome sequencing is revolutionizing cell biology and enabling new insights into the physiology of human organs. While most work has focused on specific organs, there is a need for a reference which enables comparisons of cell types between organs and tissues. We present a human single cell transcriptomic atlas comprising nearly 500,000 cells from 24 different tissues and organs, many from the same donor. This allows us to control for individual variation and enables characterization of more than 400 cell types, their distribution across tissues and tissue specific variation in gene expression. From this we identify the clonal distribution of T cells between tissues, the tissue specific mutation rate in B cells, and analyze the cell cycle state and proliferative potential of shared cell types across tissues. Finally, we characterize cell type specific RNA splicing and how such splicing varies 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. Despite the fact that nearly every cell in the body has 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 and producing data (*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, which enabled us to analyze 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**). Tissues were processed consistently across all donors. Fresh tissues were collected from consented brain-dead transplant donors through an organ procurement organization (OPO) and transported immediately to tissue experts where each tissue was dissociated. For some tissues the dissociated cells were purified into compartment-level batches (immune, stromal, epithelial and endothelial) and then recombined into balanced cell suspensions in order to enhance sensitivity for rare cell types (*12*).

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**). The raw data was processed to remove low quality cells, projected into a lower-dimensional latent space using scVI (*13*), and visualized with UMAP (*14*) (**fig. S2**). Next, the tissue experts used cellxgene (*15*) to annotate the cells that could be confidently identified by marker gene expression (*12*). These annotations were verified through a combination of automated annotation (*16*)

and manual inspection (*12*). A defined Cell Ontology terminology was used to make the annotations consistent across the different tissues (*17*), 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 a data portal located at (*18*)**.**

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. 59 separate specimens in total were collected, processed, and analyzed, and 481,120 cells passed QC filtering (**figs. S3-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.

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. S8**). We observed surprisingly good correspondence in these frequencies, especially considering that the single cell data was obtained on dissociated tissues and that some compartments were enriched. Dissociation is known to alter the representation of various cell types and generally to

increase the representation of immune cell types at the expense of others. The balancing process used on some tissues creates another distortion to absolute representation, but not to the extent that the compartments become completely equalized and thus some aspects of the original representation statistics are preserved.

To further characterize the relationship between transcriptome data and conventional histologic analysis of tissue, a team of trained pathologists analyzed H&E stained sections prepared from 9 tissues from donor TSP2 and 13 tissues from donor TSP14 (*18*). 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. In some cases, finer cell type classification was also performed. An example of such cellular and compartmental identification is illustrated in the case of the distal small intestine (**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. S9**)**.** We compared the histologically determined abundances with those obtained by single cell sequencing (**fig. S10**). 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 for organs where the compartments were purified, and more generally in every organ since not every cell type survives dissociation with equal efficiency (*19*). The histology images of the tissue are available as part of the online data portal (*18*).

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

The Tabula Sapiens can be used to study subtle differences in the gene expression programs and lineage histories of cell types that are shared across tissues. Importantly, these analyses were performed after correcting counts for potential ambient mRNA contamination and dissociation artifacts (*12*), which would otherwise result in the detection of differentially expressed genes (DEGs) that are specific to a tissue rather than to the cell type of interest. We first examined immune cells, which are born in one niche, circulate through the body, and home into other niches where they stay over time scales ranging from minutes to years. We identified tissue-specific gene expression features for most immune cell ontology classes via classical DEG analysis.

Here we focus on the signatures of tissue similarity and differences in the 36,475 macrophages distributed amongst 20 tissues, as tissue-resident macrophages are known to carry out specialized functions in different tissues and under different conditions. 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 (**fig. S11B**). We also observed a shared signature of elevated EREG expression in solid tissues such as the skin, uterus and mammary compared with EREG expression in circulatory tissues (**fig. S11B**).

Macrophages likely secrete abnormal levels of EREG during cancer-progression and facilitate the tumor micro-environment (*20*), but secretion of such factors also has a role in the homeostatic maintenance of tissues (*21*). We observed an antimicrobial phenotype of macrophage in the lung and lymph node characterized by CHIT1 expression (**fig. S11B**). Interestingly, macrophages in the lymph nodes co-expressed CHIT1 and CTSK, while CTSK was largely not expressed in the lung (**fig. S11B**). Like EREG, CTSK is thought to have roles in cancer metastasis as well as normal tissue regulation and these data provide insight into tissue-specific specializations and functional differences of macrophages (*22*, *23*).

To characterize the lineage relationships between T cells in organs we performed computational assembly of the T cell receptor sequence from T cells sequenced via Smartseq2 from donor TSP2. Multiple T cell lineages were distributed across various tissues in the body, and we mapped their relationships (**Fig. 3A**). Large clones were often found to reside in multiple organs, and several clones of Mucosal Associated Invariant T cells were shared across donors; we identified these cells by their characteristic expression of *TRAV1-2* as they are thought to be innate-like effector cells (*24*).

Lineage information can also be used to measure the level of tissue-specific somatic hyper-mutation in B cells. We computationally assembled the B Cell Receptor (BCR) gene from Smartseq2 data from donor TSP2 and then inferred the germline ancestor of each cell. The mutational load varies dramatically by tissue of residence, with blood

have the lowest mutational load compared to all of the solid tissues (**fig. S11D**); in quantitative terms 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 these functions, 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 application of using the Tabula Sapiens to analyze shared cell types across organs, we focused on endothelial cells (ECs). These cells line the surface of blood vessels and together form a conduit allowing for inter-tissue communication, oxygen, nutrient and waste exchange, and tissue-level homeostasis. While ECs are widely 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 (*25*) (**fig. S13**). Vascular-bed specific genes could provide further insight into tissue-specific homeostatic mechanisms, as well as allow for EC tissue-specificity to be deconvolved in experiments like flow cytometry.

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 (*26*) (**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

The Tabula Sapiens can also be used to understand cell type specific usage of alternative splicing. The GRCH38 RefSeq genome annotation contains 37,344 genes with multiple annotated exons, 21,923 of which have multiple annotated transcripts, totaling 169,061 splice variants derived from 328,603 possible splice junctions (*27*). Yet the function of alternative splicing and the extent to which regulation is cell type specific remains largely unexplored. We used SICILIAN (*28*), a statistical method that removes false positive spliced alignments due to technical artifacts, to identify splice junctions in the Tabula Sapiens corpus. Among other statistical filters, SICILIAN requires each called junction to have at least two supporting reads and has an estimated false positive rate of 2.8%.

SICILIAN detected a total of 955,785 junctions (**fig. S14A-E, table S7**). Of these, 217,855 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 (*29*) 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 (*18*) which uses a statistical approach for detecting cell-type-specific splicing called the SpliZ (*30*). Here we focus on two examples of cell type specific usage of two well studied genes: MYL6 and CD47. Both genes are ubiquitously expressed yet are highly regulated in single cells at the level of splicing (**Fig. 4**). 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 (as opposed to Ca2+-sensitive “regulatory light chains” (RLCs) such as calmodulin), 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**). The two isoforms differ by 5 amino acids in the C-terminal helix, which is in close contact with the myosin lever arm; some studies suggest that the -exon6 isoform confers on myosin a faster shortening velocity (*31*). While the -exon6 isoform has previously been mainly described in phasic smooth muscle (*32*), the Tabula Sapiens atlas shows that 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. Further, 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**) and using both 10X and Smart-Seq2 sequencing technologies.

As another example, 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 (*33*). Targeting the latter function has been promising for treating some myeloid malignancies (*34*). CD47 has complex splicing patterns that include alternative inclusion of at least 4 different exons immediately adjacent to the signaling domain ending at the 3’ splice splice site at exon 11 (**fig. S14F**). Differential use of exons 7-10 (**Fig. 4C and fig. S14F**) compose a variably long cytoplasmic tail (*35*). 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 understanding potentially 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 various forms of dynamic information about the cells from the data. For example, one of the most important transient changes of internal cell state is cell division. We computed a cycling index for each cell type across all organs to identify actively proliferating versus quiescent or post-mitotic cell states. This index was derived on the basis of the log ratio of the number of cycling to non-cycling cells for each cell type, determined by high confidence cell cycle markers for G1-M phases (G1/S markers: CEP57, CDCA7L; S markers: ABHD10, CCDC14, CDKN2AIP, NT5DC1, SVIP, PTAR1; G2 makers: ANKRD36C, YEATS4, DCTPP1; G2/M markers: SMC4, TMPO, LMNB1, HINT3; M markers: HMG20B, HMGB3, HPS4) to indicate cycling and G0 phase markers (CDKN1A, CDKN1B, CDKN1C) for non-cycling (*12*).

We observed that rapidly dividing progenitor cells had among the highest cycling indices, while cell types mostly 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 (*36*). 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**).

As a further example of how the Tabula Sapiens can be used to reveal cell state dynamics, we used RNA velocity (*37*) to study trans-differentiation of bladder mesenchymal cells to myofibroblasts (**Fig. 5B**). This process is important for tissue remodeling and healing, and if left unchecked can result in fibrosis. Myofibroblasts produce different components of the ECM such as collagen and fibronectin. RNA velocity captures the ratio of unspliced to spliced RNA on a per-gene basis, and thus

provides temporal information (*37*, *38*). RNA velocities across many genes are used to infer a latent temporal axis on which cells undergoing a dynamic process such as differentiation can be ordered on. The inferred transitions on this temporal axis as bladder mesenchymal cells transdifferentiate to myofibroblasts are shown. Latent time analysis, which provides an estimate of each cell's internal clock using RNA velocity trajectories (*39*), correctly identified the direction of differentiation without requiring specification of root cells (**Fig. 5C**). Similar trajectories were found across multiple donors. Finally, the ordering of 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 (*40*) – 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 (*41*). 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

Imbalances in the interactions between the gut microbiome and the host immune system impact are linked to many region-specific intestinal diseases (*42*, *43*), and stool is not necessarily representative of the spatially distinct microbial (*44*) and immune (*45*)

niches throughout the intestinal tract. Despite its importance, the spatial heterogeneity of the microbiome remains understudied and largely unknown. 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 ~1 g of digesta were collected from each location using an inoculating loop (23 samples in total from TSP2 and 30 samples from TSP14).

The taxonomic composition of the microbiome was characterized by amplifying the 16S rRNA gene from all samples. Uniformly there was a high (~10-30%) relative abundance of Proteobacteria, particularly Enterobacteriaceae (**Fig. 6B**), even in the colon. Enterobacteriaceae are rarely found at such high abundance in stool, hence this high relative abundance may be due to the postmortem state of the donor. Most samples within the sigmoid and ascending colon were similar to each other, although the relative abundances of the Proteobacteria phylum and Lachnospiraceae family were variable in the sigmoid colon. By contrast, 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 the spatial pattern of TSP2's microbiome (**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. 6C**); 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**). Shannon diversity largely mimicked the number of observed species (**Fig. 6D**). In a comparison of species from adjacent regions across the gut, a large fraction of species was unique to each region (**Fig. 6E**), reflecting the patchiness. The average number of unique ASVs (a proxy for species) in each sample was 41.4±21.1 for the five regions of TSP14, and within each location, samples in the duodenum, jejunum, and ileum exhibited 1.83±1.6, 8.5±4.4, and 11.4±6.9 unique ASVs, respectively, compared with 18.2±8.2 and 16.2±7.9 for the ascending and sigmoid colon, respectively. These numbers indicate that ~25% of the ASVs in each sample were unique when comparing to other samples from the same region. For TSP2, somewhat higher fractions of unique ASVs were observed in the small intestine and very few ASVs were unique to the colon (both in comparisons across regions and within each region, **fig. S17D**), indicating greater patchiness in the SI of this donor.

To compare our findings to existing fecal microbiome data, we analyzed the prevalence of each of the 601 ASVs detected across all TSP14 samples to a set of stool samples collected from healthy individuals in a separate study. Of the 1193 ASVs detected in the stool samples, only 31 were also found in TSP14 samples, of which 7 were specific to the small intestine, 9 were specific to the colon, and 15 were found in both the small and large intestine. The 31 ASVs were typically found in only a few of the stool samples

(4.0±3.5), highlighting the unique aspect of the ASVs that we observed in the small intestine.

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. 6F**). The most significant transcriptional differences in T cells between the small and large intestine were genes associated with trafficking, survival, and activation (**Fig. 6G**, **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 (*46*), 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 to 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 unique in being constructed from live cells from multiple-organ donors and including biological replicates from more than one donor. It 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. For example, choosing a particular cell type and studying it across organs, datasets, and species will yield new biological insights – as was demonstrated recently with fibroblasts (*47*). As another example, one can compare gene expression in fetal human cell types (*48*) to those determined here in adults to glean insight into the effects of aging and the loss of plasticity from early development to maturity. Many of these datasets were created with different technical approaches and having data from cell types shared across many organs may facilitate the development of normalization methods to compare such diverse data – and also may enable better understanding of technical artifacts introduced by the various approaches (*8*, *9*, *49*, *50*). The Tabula Sapiens has enabled several discoveries relating to shared behavior or subtle organ specific differences across a number of cell types. For example, we found specific T cell clones which are shared between organs, and characterized organ dependent hypermutation rates amongst B cells. Cell types which are shared across tissues often show subtle but clear differences in gene expression, as we found with both endothelial cells and macrophages. We found an unexpectedly large and diverse amount of cell-type specific RNA splice variant usage, and discovered and validated many new splices. Finally, we performed direct spatial characterization of microbiome diversity and related that to gene expression in resident immune cells. 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. We expect that, similar to the human genome project, over time the release of updated versions of Tabula Sapiens will incorporate data from additional donors and include further refinements in the cell type annotations.

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 as rapidly as possible to individual tissue expert labs where they were immediately prepared for transcriptome sequencing. When available, a 1-2 cm^2 sample was preserved in Formalin for histological analysis. For each organ/tissue, 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; the details of which are tissue specific and explained in the methods supplement (*12*). To detect as many cell types as possible, the 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. *Annual Review of Biomedical Data Science*. **4** (2021), doi:10.1146/ANNUREV-BIODATASCI-092820-031008.

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

7. A. R. Wu, J. Wang, A. M. Streets, Y. Huang, Single-Cell Transcriptional Analysis. *Annual Review of Analytical Chemistry*. **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 Biology 2020 21:1*. **21**, 1–34 (2020).

10. Tabula Muris Consortium, Overall coordination, Logistical coordination, Organ collection and processing, Library preparation and sequencing, Computational data analysis, Cell type annotation, Writing group, Supplemental text writing group, Principal investigators, 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. R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, N. Yosef, Deep generative modeling for single-cell transcriptomics. *Nature Methods*. **15**, 1053–1058 (2018).

14. E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, E. W. Newell, *Nature Biotechnology*, in press, doi:10.1038/nbt.4314.

15. 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, *bioRxiv*, in press, doi:10.1101/2021.04.05.438318.

16. 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. *Molecular Systems Biology*. **17**, e9620 (2021).

17. 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).

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

19. 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).

20. M. Farooqui, L. R. Bohrer, N. J. Brady, P. Chuntova, S. E. Kemp, C. T. Wardwell, A. C. Nelson, K. L. Schwertfeger, Epiregulin contributes to breast tumorigenesis through regulating matrix metalloproteinase 1 and promoting cell survival. *Molecular Cancer 2015 14:1*. **14**, 1–13 (2015).

21. D. J. Riese, II, R. L. Cullum, Epiregulin: Roles in Normal Physiology and Cancer. *Seminars in cell & developmental biology*. **0**, 49 (2014).

22. M. K. Herroon, E. Rajagurubandara, D. L. Rudy, A. Chalasani, A. L. Hardaway, I. Podgorski, Macrophage cathepsin K promotes prostate tumor progression in bone. *Oncogene*. **32**, 1580–1593 (2013).

23. F. K. Leusink, E. Koudounarakis, M. H. Frank, R. Koole, P. J. van Diest, S. M. Willems, Cathepsin K associates with lymph node metastasis and poor prognosis in oral squamous cell carcinoma. *BMC Cancer*. **18**, 385 (2018).

24. D. I. Godfrey, H.-F. Koay, J. McCluskey, N. A. Gherardin, The biology and functional importance of MAIT cells. *Nature Immunology 2019 20:9*. **20**, 1110–1128 (2019).

25. 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. *Science Advances*. **7**, eabh2169.

26. 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. *Nature 2020 586:7831*. **586**, 785–789 (2020).

27. N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, K. D. Pruitt, Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*. **44**, D733–D745 (2016).

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

29. 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).

30. 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).

31. H. Sugi, *Muscle Contraction and Cell Motility* (Jenny Stanford Publishing, 2016).

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

33. 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).

34. R. Majeti, M. P. Chao, A. A. Alizadeh, W. W. Pang, S. Jaiswal, K. D. Gibbs, N. van Rooijen, I. L. Weissman, CD47 Is an Adverse Prognostic Factor and Therapeutic Antibody Target on Human Acute Myeloid Leukemia Stem Cells. *Cell*. **138**, 286–299 (2009).

35. 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. *Cellular Oncology 2019 43:1*. **43**, 19–30 (2019).

36. H. Gehart, H. Clevers, Tales from the crypt: new insights into intestinal stem cells. *Nature Reviews Gastroenterology and Hepatology*. **16** (2019), pp. 19–34.

37. 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).

38. V. Bergen, R. A. Soldatov, P. V. Kharchenko, F. J. Theis, RNA velocity—current challenges and future perspectives. *Mol Syst Biol*. **17** (2021), doi:10.15252/msb.202110282.

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

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

41. R. Li, X. Li, J. Hagood, M.-S. Zhu, X. Sun, Myofibroblast contraction is essential for generating and regenerating the gas-exchange surface. *The Journal of Clinical Investigation*. **130**, 2859–2871 (2020).

42. A. M. Mowat, W. W. Agace, Regional specialization within the intestinal immune system. *Nat Rev Immunol*. **14**, 667–685 (2014).

43. J. Halfvarson, C. J. Brislawn, R. Lamendella, Y. Vázquez-Baeza, W. A. Walters, L. M. Bramer, M. D'Amato, F. Bonfiglio, D. McDonald, A. Gonzalez, E. E. McClure, M. F. Dunklebarger, R. Knight, J. K. Jansson, Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat Microbiol*. **2**, 1–7 (2017).

44. G. P. Donaldson, S. M. Lee, S. K. Mazmanian, Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol*. **14**, 20–32 (2016).

45. K. R. James, T. Gomes, R. Elmentaite, N. Kumar, E. L. Gulliver, H. W. King, M. D. Stares, B. R. Bareham, J. R. Ferdinand, V. N. Petrova, K. Polański, S. C. Forster, L. B. Jarvis, O. Suchanek, S. Howlett, L. K. James, J. L. Jones, K. B. Meyer, M. R. Clatworthy, K. Saeb-Parsy, T. D. Lawley, S. A. Teichmann, Distinct microbial and immune niches of the human colon. *Nat Immunol*. **21**, 343–353 (2020).

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

47. 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).

48. 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.

49. 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.

50. 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.

51. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods. 2013;10(11):nmeth.2639.

52. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15.

53. Griffiths JA, Richard AC, Bach K, Lun ATL, Marioni JC. Detection and removal of barcode swapping in single-cell RNA-seq data. Nat Commun. 2018 Dec;9(1):2667.

54. Yang S, Corbett SE, Koga Y, Wang Z, Johnson WE, Yajima M, et al. Decontamination of ambient RNA in single-cell RNA-seq with DecontX. Genome Biol. 2020 Dec;21(1):57.

55. Gayoso A, Lopez R, Xing G, Boyeau P, Wu K, Jayasuriya M, et al. scvi-tools: a library for deep probabilistic analysis of single-cell omics data [Internet]. Bioinformatics; 2021 Apr [cited 2021 Jul 9]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.04.28.441833

56. Diehl AD, Meehan TF, Bradford YM, Brush MH, Dahdul WM, Dougall DS, et al. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. J Biomed Semant. 2016;7(1):44.

57. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12(85):2825–30.

58. Polański K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park J-E. BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics. 2019;

59. Hie B, Bryson B, Berger B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat Biotechnol. 2019 Jun;37(6):685–91.

60. Sofroniew N, Talley Lambert, Evans K, Nunez-Iglesias J, Winston P, Bokota G, et al. napari/napari: 0.4.10rc0 [Internet]. Zenodo; 2021 [cited 2021 Jul 10]. Available from: https://zenodo.org/record/4968798

61. Graham S, Vu QD, Raza SEA, Azam A, Tsang YW, Kwak JT, et al. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal. 2019 Dec;58:101563.

62. Core Team R. R: A Language and Environment for Statistical Computing. 2017 [cited 2021 Jul 10]; Available from: https://www.r-project.org/

63. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016 Jul;13(7):581–3.

64. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, et al. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Res. 2014 Jan;42(D1):D643–8.

65. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011 Feb 15;27(4):592–3.

66. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. Watson M, editor. PLoS ONE. 2013 Apr 22;8(4):e61217.

67. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, et al. Identification of Genes Periodically Expressed in the Human Cell Cycle and Their Expression in Tumors. Solomon MJ, editor. Mol Biol Cell. 2002 Jun;13(6):1977–2000.

68. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015 May;161(5):1202–14.

69. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016 Apr 8;352(6282):189–96.

70. Santos A, Wernersson R, Jensen LJ. Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. Nucleic Acids Res. 2015 Jan 28;43(D1):D1140–4.

71. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;48(5):1070 420.

72. van den Brink SC, Sage F, Vértesy Á, Spanjaard B, Peterson-Maduro J, Baron CS, et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat Methods. 2017 Oct;14(10):935–6.

73. Ferreira PG, Muñoz-Aguirre M, Reverter F, Sá Godinho CP, Sousa A, Amadoz A, et al. The effects of death and post-mortem cold ischemia on human tissue transcriptomes. Nat Commun. 2018 Feb 13;9(1):490.

74. Massoni-Badosa R, Iacono G, Moutinho C, Kulis M, Palau N, Marchese D, et al. Sampling time-dependent artifacts in single-cell genomics studies. Genome Biol. 2020 Dec;21(1):112.

75. van Velthoven CTJ, de Morree A, Egner IM, Brett JO, Rando TA. Transcriptional Profiling of Quiescent Muscle Stem Cells In Vivo. Cell Rep. 2017 Nov;21(7):1994–2004.

76. Savage AK, Gutschow MV, Chiang T, Henderson K, Green R, Chaudhari M, et al. Multimodal Analysis for Human ex vivo Studies Shows Extensive Molecular Changes from Delays in Blood Processing [Internet]. Immunology; 2020 Oct [cited 2021 Jul 10]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.10.18.344663

77. Stubbington MJT, Lönnberg T, Proserpio V, Clare S, Speak AO, Dougan G, et al. T cell fate and clonality inference from single-cell transcriptomes. Nat Methods. 2016;13(4):nmeth.3800.

78. Lindeman I, Emerton G, Mamanova L, Snir O, Polanski K, Qiao S-W, et al. BraCeR: B-cell-receptor reconstruction and clonality inference from single-cell RNA-seq. Nat Methods. 2018;15(8):563–5.

79. Koster J, Rahmann S. Snakemake--a scalable bioinformatics workflow engine. Bioinformatics. 2012 Oct 1;28(19):2520–2.

80. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. 2013 Jul 1;41(W1):W34–40.

81. Sturm G, Szabo T, Fotakis G, Haider M, Rieder D, Trajanoski Z, et al. Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor-sequencing data. Mathelier A, editor. Bioinformatics. 2020 Sep 15;36(18):4817–8.

82. Peixoto TP. The graph-tool python library [Internet]. figshare; 2017 [cited 2021 Jul 10]. p. 136152697 Bytes. Available from: https://figshare.com/articles/dataset/graph_tool/1164194

83. Waskom M. seaborn: statistical data visualization. J Open Source Softw. 2021 Apr 6;6(60):3021.

84. Gehart H, Clevers H. Tales from the crypt: new insights into intestinal stem cells. Nat Rev Gastroenterol Hepatol. 2019 Jan;16(1):19–34.

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 (*18*)**.**

Data and materials availability: The entire dataset can be explored interactively at http://tabula-sapiens-portal.ds.czbiohub.org/. The code used for the analysis is available from GitHub at https://github.com/czbiohub/tabula-sapiens. Gene counts and metadata are available from figshare (https://doi.org/10.6084/m9.figshare.14267219) 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). SpliZ scores are available from figshare (https://doi.org/10.6084/m9.figshare.14977281).

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], Rob Bierman[3,6], 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. Only shared cell types with more than 10 cells with spliced reads mapping to MYL6 are shown. Horizontal box plots in **B** show the distribution of exon inclusion in each cell type (point outside of 1.5*IQR are plotted individually). **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 (based on read counts and using exon numbers) 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. The average splice site rank was calculated for each cell with more than one read corresponding to the splice site. Only

shared cell types with more than 5 cells with spliced reads mapping to CD47 are shown. Both genes were found significant (controlled FDR< 0.05) by the SpliZ analysis (*30*).

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-D**. Relative abundances (**B**) and richness (number of observed species, **C**) at the family level in each sampling location, as determined by 16S rRNA sequencing. The Shannon diversity (**D**), 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. **E**. 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. **F**. UMAP clustering of T cells reveals distinct transcriptome profiles in the distal and proximal small and large intestines. **G**. 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 (51-84)

Resources

Data portal for Tabula Sapiens (*18*)

Fig. 1

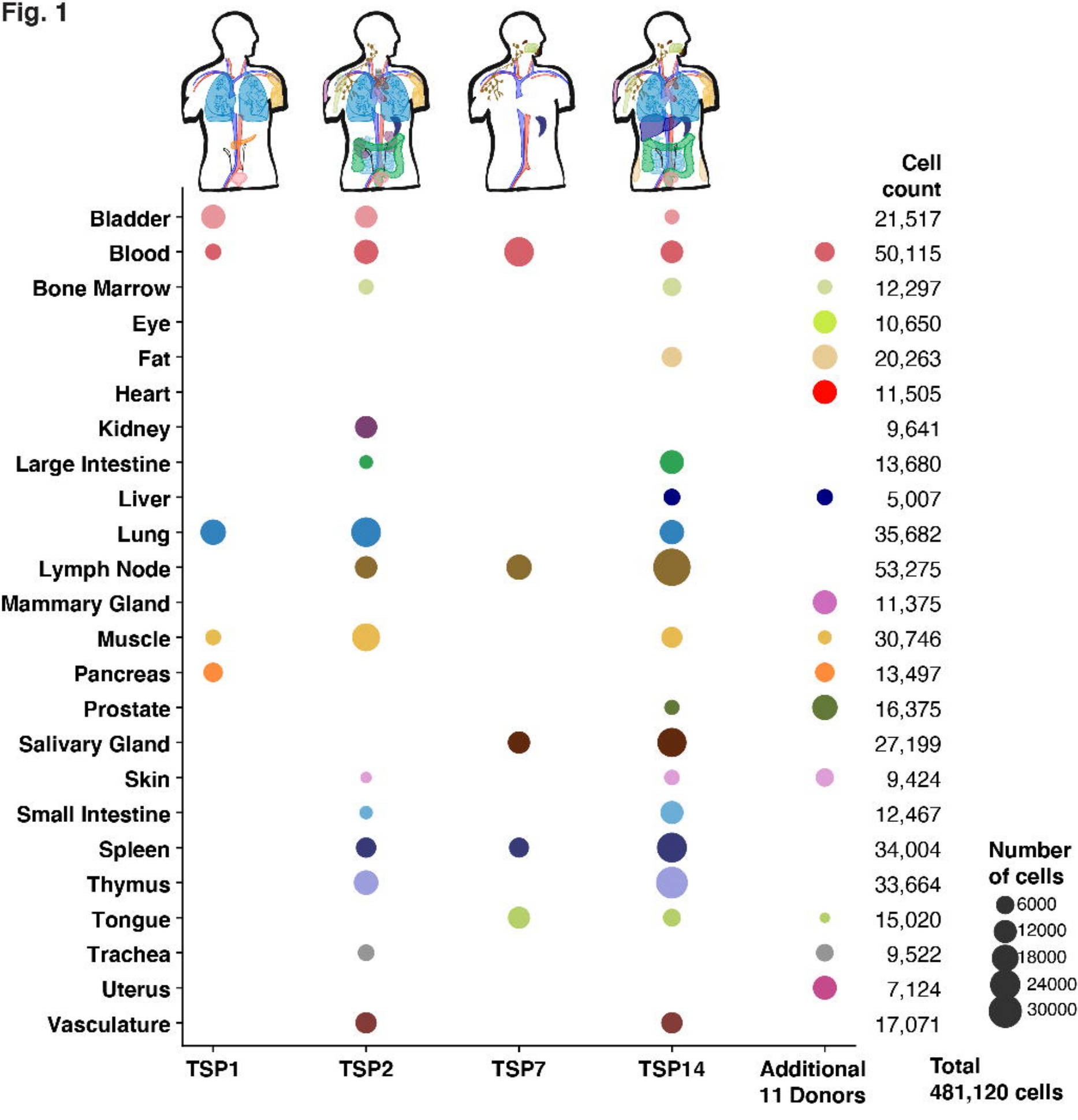
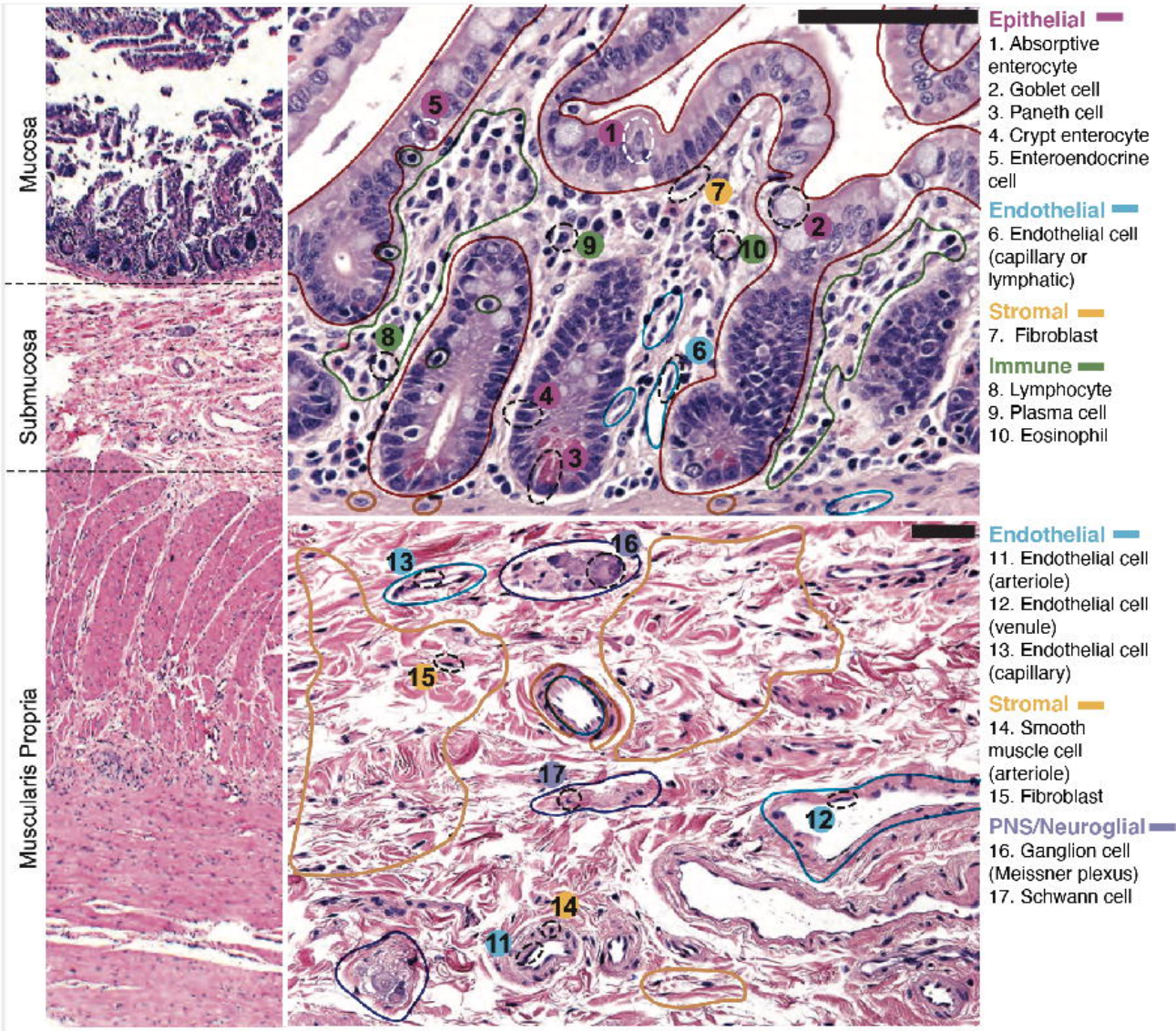

Cell count
Bladder 21,517
Blood 50,115
Bone Marrow 12,297
Eye 10,650
Fat 20,263
Heart 11,505
Kidney 9,641
Large Intestine 13,680
Liver 5,007
Lung 35,682
Lymph Node 53,275
Mammary Gland 11,375
Muscle 30,746
Pancreas 13,497
Prostate 16,375
Salivary Gland 27,199
Skin 9,424
Small Intestine 12,467
Spleen 34,004
Thymus 33,664
Tongue 15,020
Trachea 9,522
Uterus 7,124
Vasculature 17,071
TSP1
TSP2
TSP7
TSP14
Additional 11 Donors
Total 481,120 cells
Number of cells
6000
12000
18000
24000
30000

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

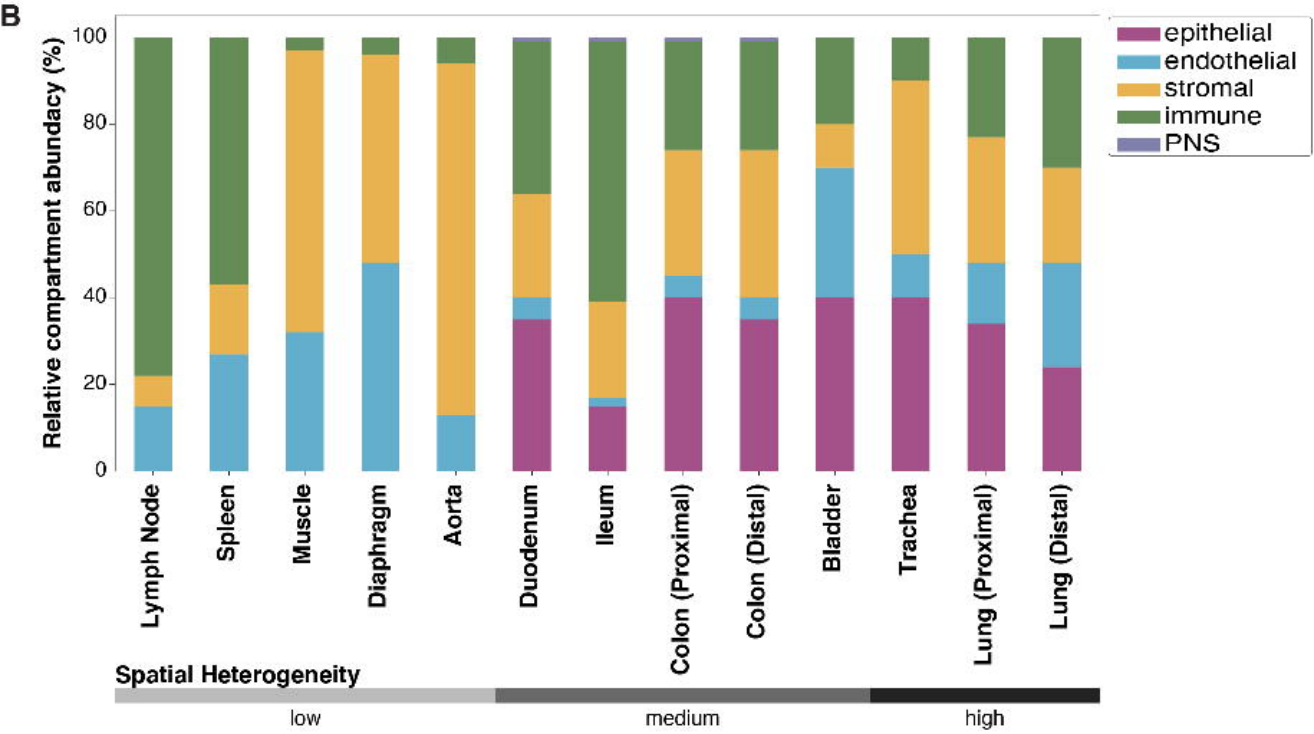

Relative compartment abundacy (%)
0
20
40
60
80
100
epithelial
endothelial
stromal
immune
PNS
Lymph Node
Spleen
Muscle
Diaphragm
Aorta
Duodenum
Ileum
Colon (Proximal)
Colon (Distal)
Bladder
Trachea
Lung (Proximal)
Lung (Distal)
Spatial Heterogeneity
low
medium
high

Fig. 3

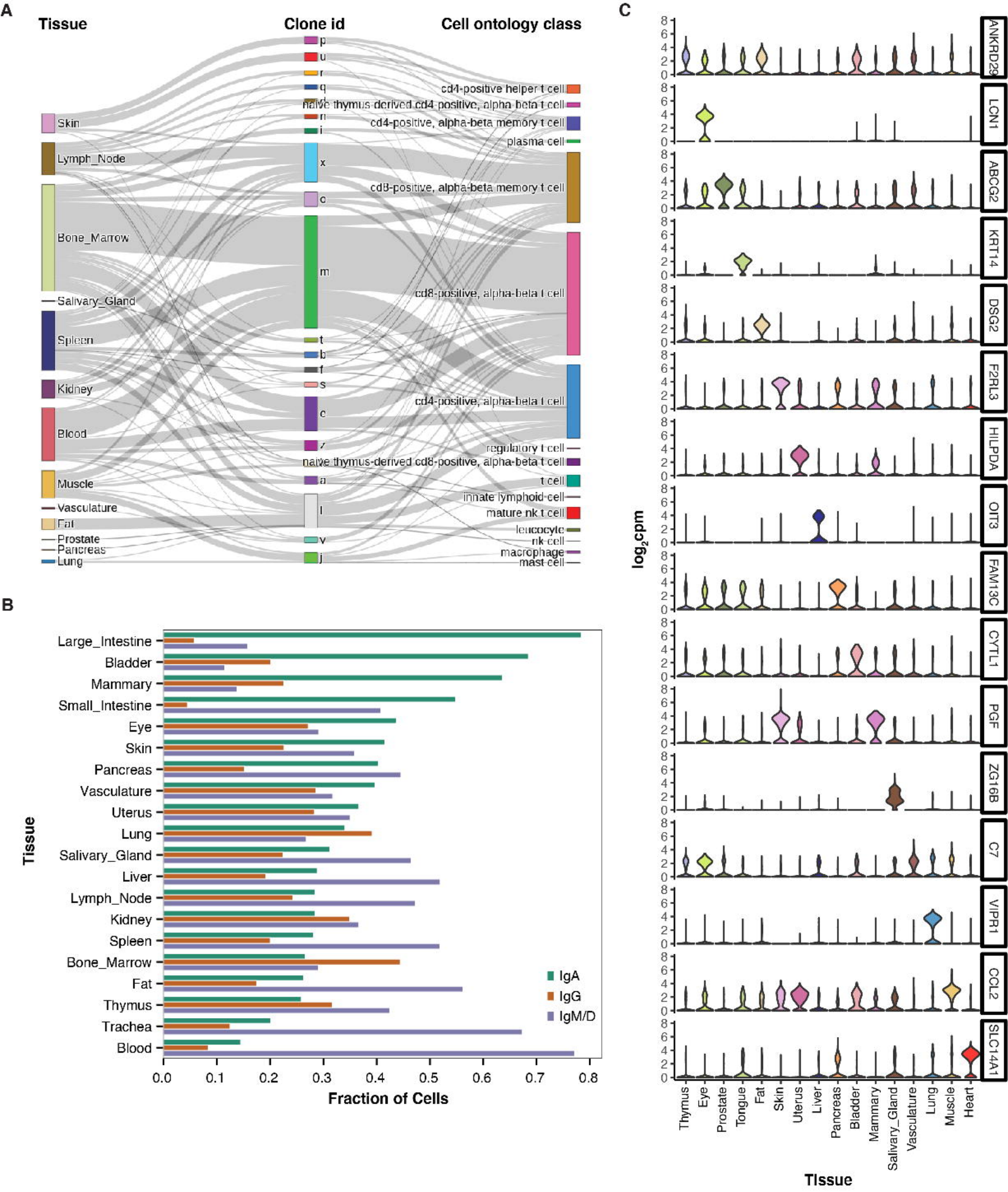
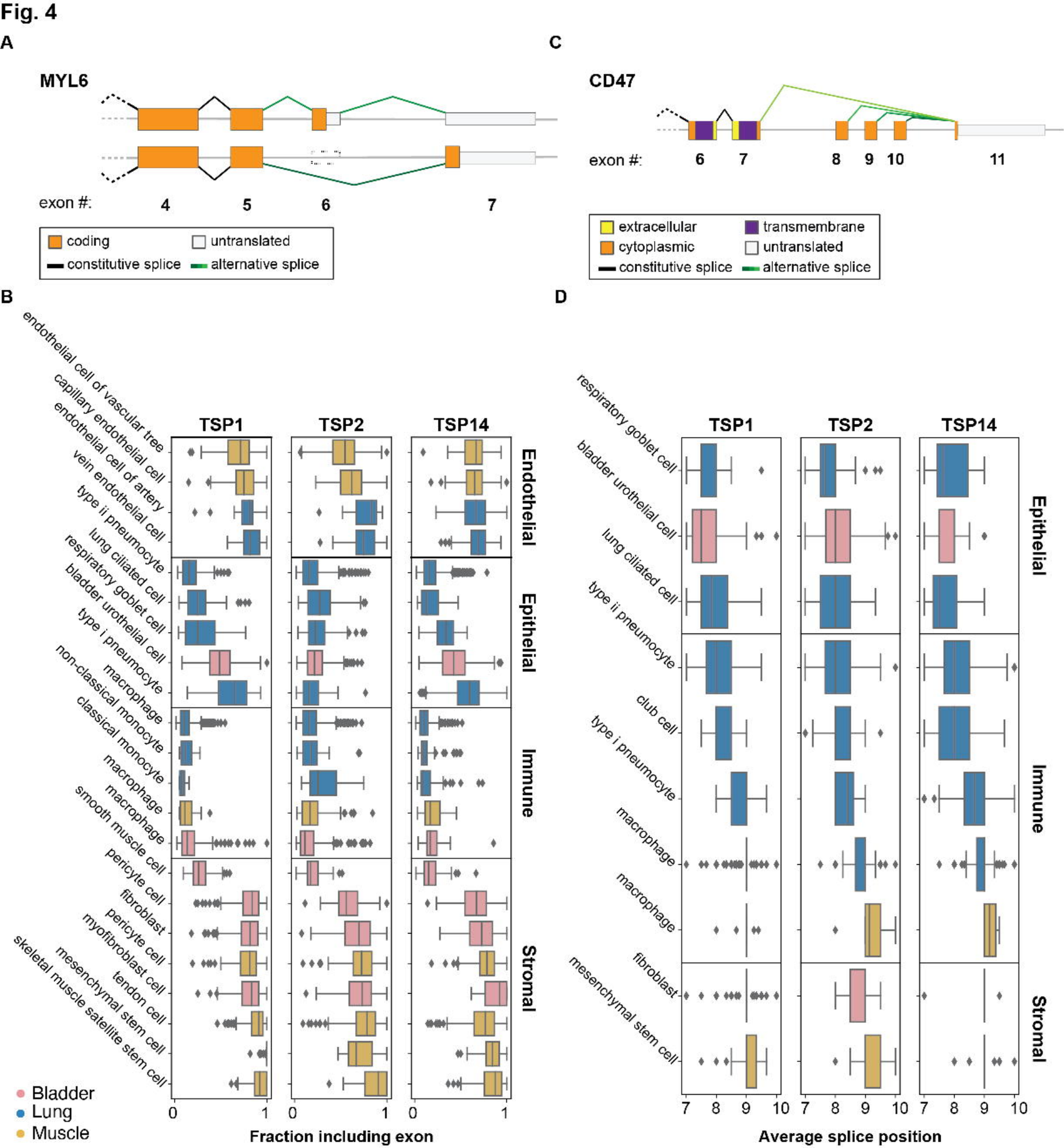

A
Tissue
Clone id
Cell ontology class
Skin
Lymph_Node
Bone_Marrow
Salivary_Gland
Spleen
Kidney
Blood
Muscle
Vasculature
Fat
Prostate
Pancreas
Lung
cd4-positive helper t cell
naïve thymus-derived cd4-positive, alpha-beta t cell
cd4-positive, alpha-beta memory t cell
plasma cell
cd8-positive, alpha-beta memory t cell
cd8-positive, alpha-beta t cell
cd4-positive, alpha-beta t cell
regulatory t cell
naïve 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
Fraction of Cells
C
ANKRD29
LCN1
ABCG2
KRT14
DSG2
F2RL3
HILPDA
OIT3
FAM13C
CYTL1
PGF
ZG16B
C7
VIPR1
CCL2
SLC14A1
log2cpm
Thymus
Eye
Prostate
Tongue
Fat
Skin
Uterus
Liver
Pancreas
Bladder
Mammary
Salivary_Gland
Vasculature
Lung
Muscle
Heart
Tissue

Fig. 4


A
MYL6
exon #:
4
5
6
7
coding
untranslated
constitutive splice
alternative splice
C
CD47
exon #:
6
7
8
9
10
11
extracellular
transmembrane
cytoplasmic
untranslated
constitutive splice
alternative splice
B
TSP1
TSP2
TSP14
endothelial cell of vascular tree
capillary endothelial cell
endothelial cell of artery
vein endothelial cell
type ii pneumocyte
lung ciliated cell
respiratory goblet cell
bladder urothelial cell
type i pneumocyte
macrophage
non-classical monocyte
classical monocyte
macrophage
macrophage
smooth muscle cell
pericyte cell
fibroblast
pericyte cell
myofibroblast cell
tendon cell
mesenchymal stem cell
skeletal muscle satellite stem cell
Endothelial
Epithelial
Immune
Stromal
Bladder
Lung
Muscle
0
1
Fraction including exon
D
TSP1
TSP2
TSP14
respiratory goblet cell
bladder urothelial cell
lung ciliated cell
type ii pneumocyte
club cell
type i pneumocyte
macrophage
macrophage
fibroblast
mesenchymal stem cell
Epithelial
Immune
Stromal
7
8
9
10
Average splice position

Mesenchymal to myofibroblas transdifferentiation (bladder)

Fig. 5

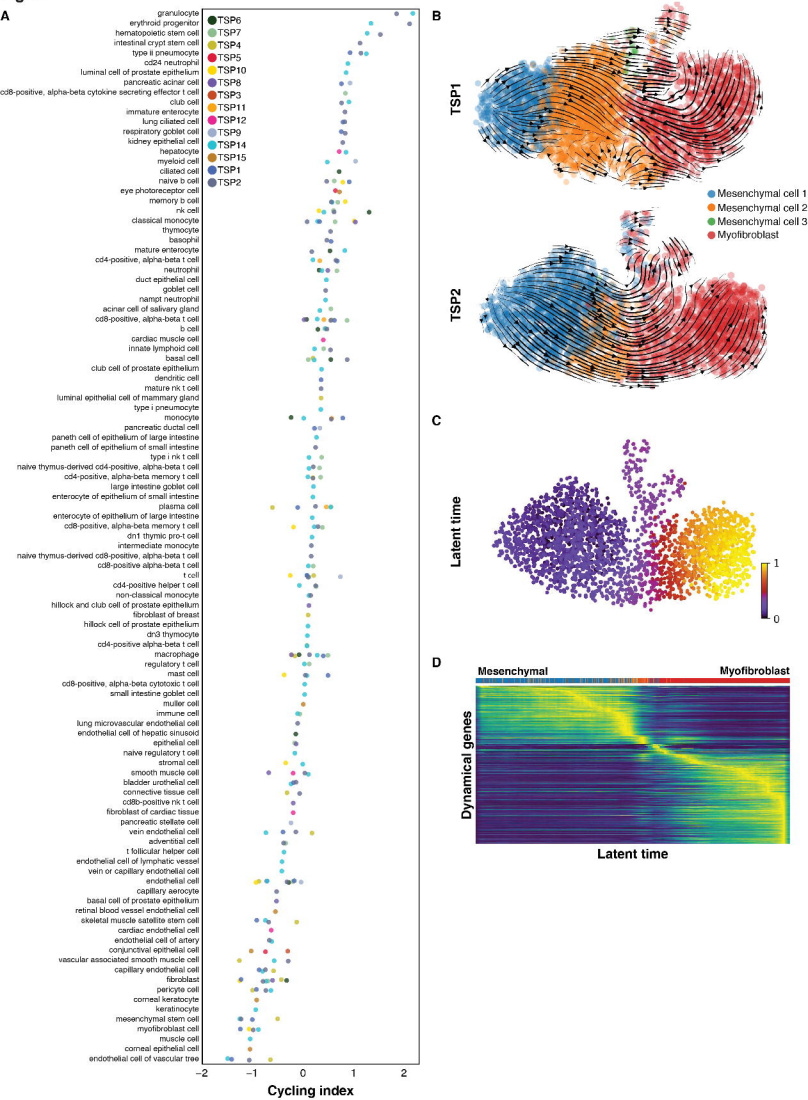


Fig. 6

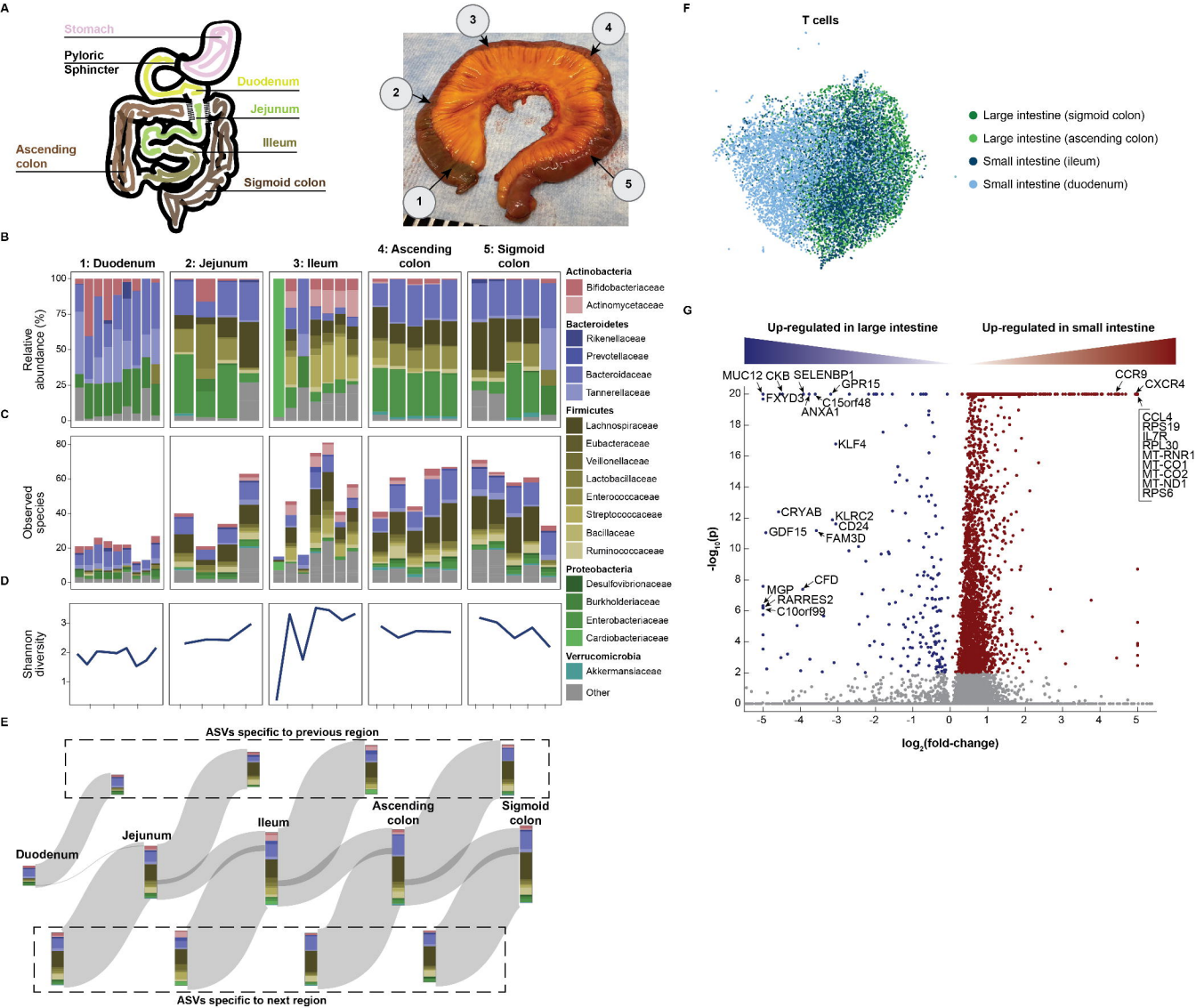

A
Stomach
Pyloric Sphincter
Duodenum
Jejunum
Ileum
Ascending colon
Sigmoid colon
1
2
3
4
5
B
1: Duodenum
2: Jejunum
3: Ileum
4: Ascending colon
5: Sigmoid colon
Relative abundance (%)
C
Observed species
D
Shannon diversity
Actinobacteria
Bifidobacteriaceae
Actinomycetaceae
Bacteroidetes
Rikenellaceae
Prevotellaceae
Bacteroidaceae
Tannerellaceae
Firmicutes
Lachnospiraceae
Eubacteraceae
Veillonellaceae
Lactobacillaceae
Enterococcaceae
Streptococcaceae
Bacillaceae
Ruminococcaceae
Proteobacteria
Desulfovibrionaceae
Burkholderiaceae
Enterobacteriaceae
Cardiobacteriaceae
Verrucomicrobia
Akkermansiaceae
Other
E
ASVs specific to previous region
Duodenum
Jejunum
Ileum
Ascending colon
Sigmoid colon
ASVs specific to next region
F
T cells
Large intestine (sigmoid colon)
Large intestine (ascending colon)
Small intestine (ileum)
Small intestine (duodenum)
G
Up-regulated in large intestine
Up-regulated in small intestine
MUC12
CKB
SELENBP1
GPR15
FXYD3
C15orf48
ANXA1
KLF4
CRYAB
KLRC2
CD24
GDF15
FAM3D
CFD
MGP
RARRES2
C10orf99
CCR9
CXCR4
CCL4
RPS19
IL7R
RPL30
MT-RNR1
MT-CO1
MT-CO2
MT-ND1
RPS6
-log10(p)
log2(fold-change)

Fig. S1

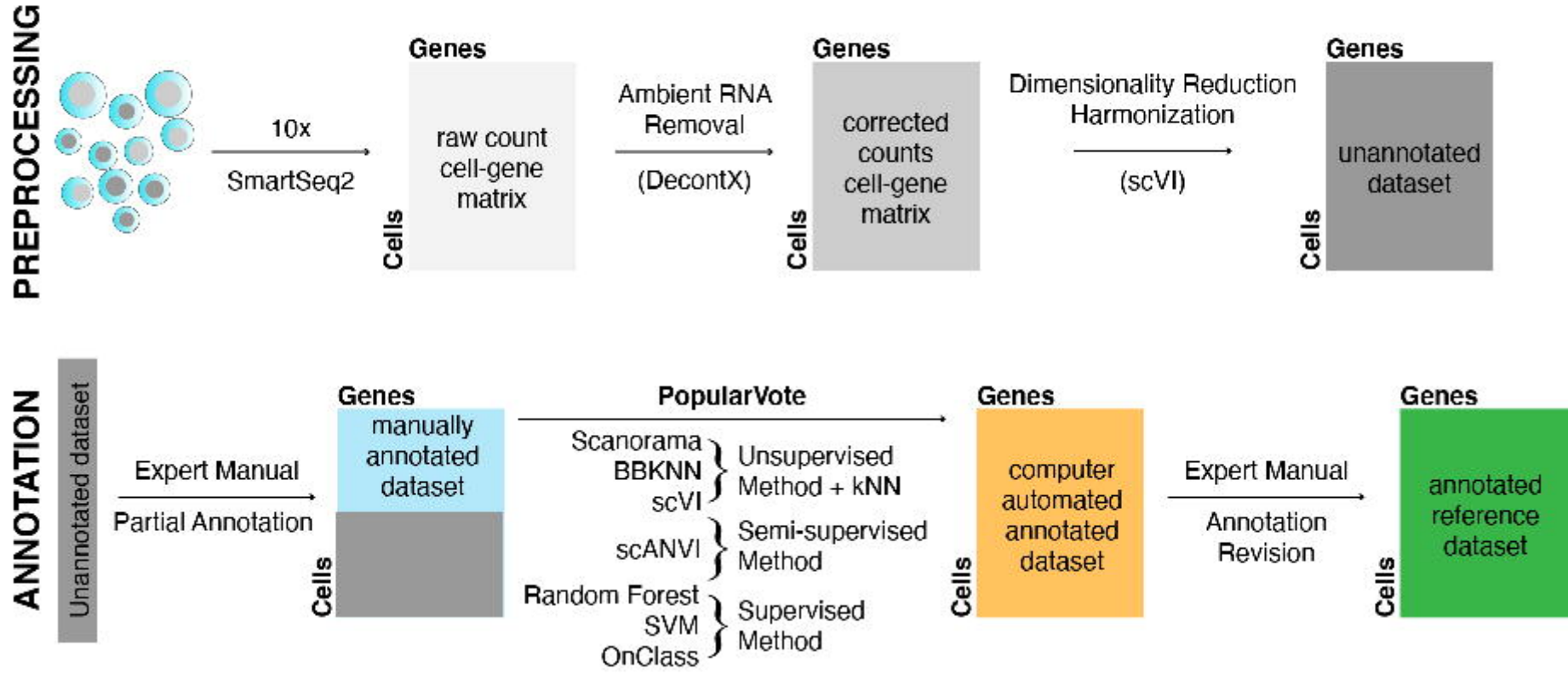

PREPROCESSING
10x
SmartSeq2
Genes
Cells
raw count cell-gene matrix
Ambient RNA Removal
(DecontX)
Genes
Cells
corrected counts cell-gene matrix
Dimensionality Reduction Harmonization
(scVI)
Genes
Cells
unannotated dataset
ANNOTATION
Unannotated dataset
Expert Manual
Partial Annotation
Genes
Cells
manually 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

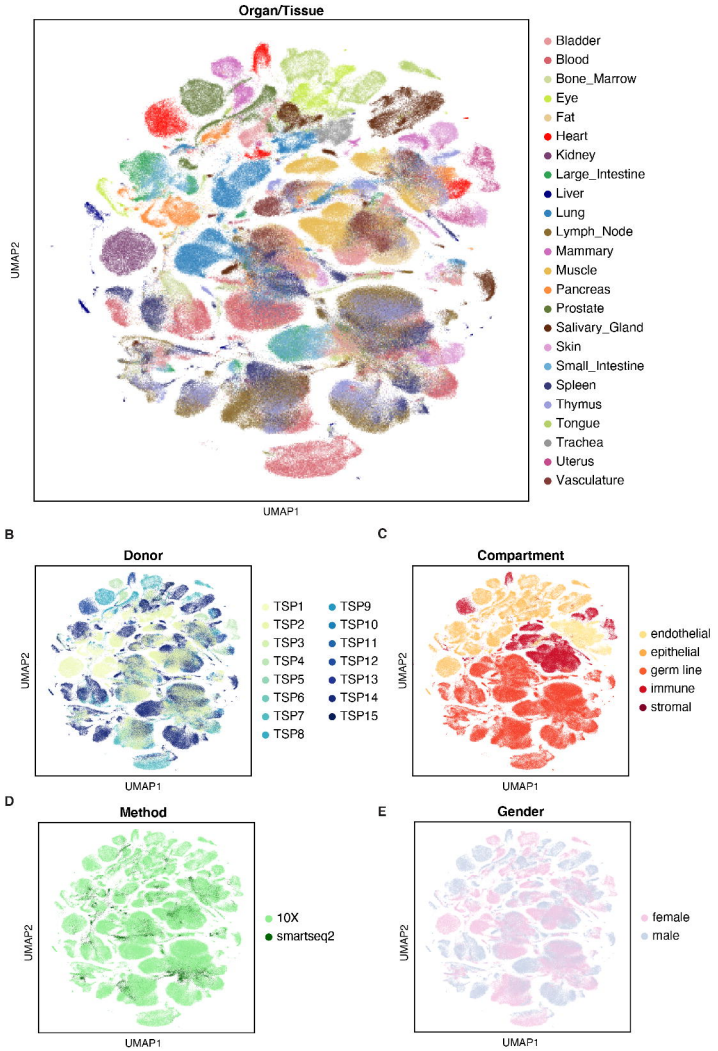

A
Organ/Tissue
Bladder
Blood
Bone_Marrow
Eye
Fat
Heart
Kidney
Large_Intestine
Liver
Lung
Lymph_Node
Mammary
Muscle
Pancreas
Prostate
Salivary_Gland
Skin
Small_Intestine
Spleen
Thymus
Tongue
Trachea
Uterus
Vasculature
UMAP2
UMAP1
B
Donor
TSP1
TSP2
TSP3
TSP4
TSP5
TSP6
TSP7
TSP8
TSP9
TSP10
TSP11
TSP12
TSP13
TSP14
TSP15
UMAP2
UMAP1
C
Compartment
endothelial
epithelial
germ line
immune
stromal
UMAP2
UMAP1
D
Method
10X
smartseq2
UMAP2
UMAP1
E
Gender
female
male
UMAP2
UMAP1

Fig. S3

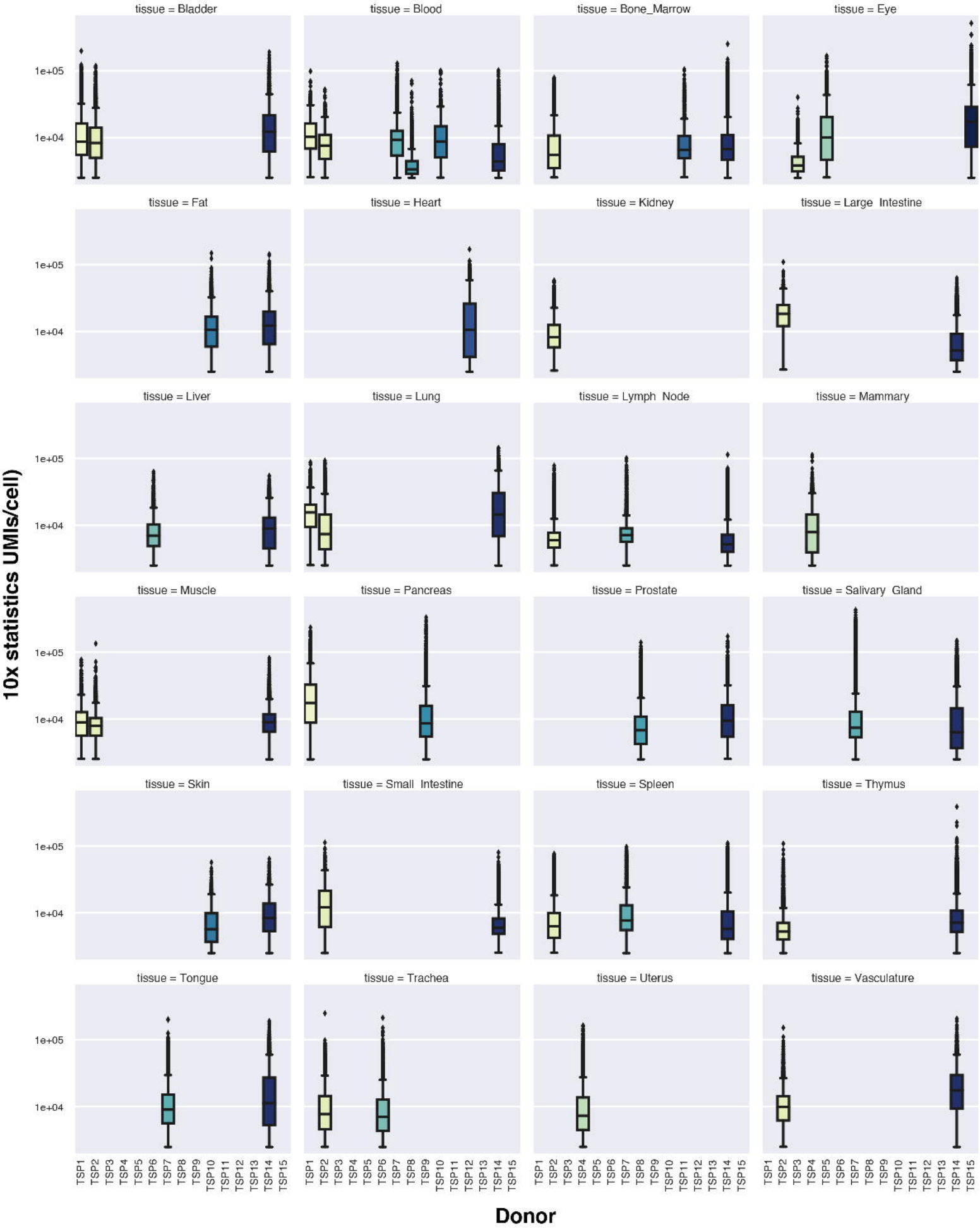

tissue = Bladder
tissue = Blood
tissue = Bone_Marrow
tissue = Eye
tissue = Fat
tissue = Heart
tissue = Kidney
tissue = Large Intestine
tissue = Liver
tissue = Lung
tissue = Lymph Node
tissue = Mammary
tissue = Muscle
tissue = Pancreas
tissue = Prostate
tissue = Salivary Gland
tissue = Skin
tissue = Small Intestine
tissue = Spleen
tissue = Thymus
tissue = Tongue
tissue = Trachea
tissue = Uterus
tissue = Vasculature
1e+05
1e+04
10x statistics UMIs/cell)
TSP1
TSP2
TSP3
TSP4
TSP5
TSP6
TSP7
TSP8
TSP9
TSP10
TSP11
TSP12
TSP13
TSP14
TSP15
Donor

Fig. S4

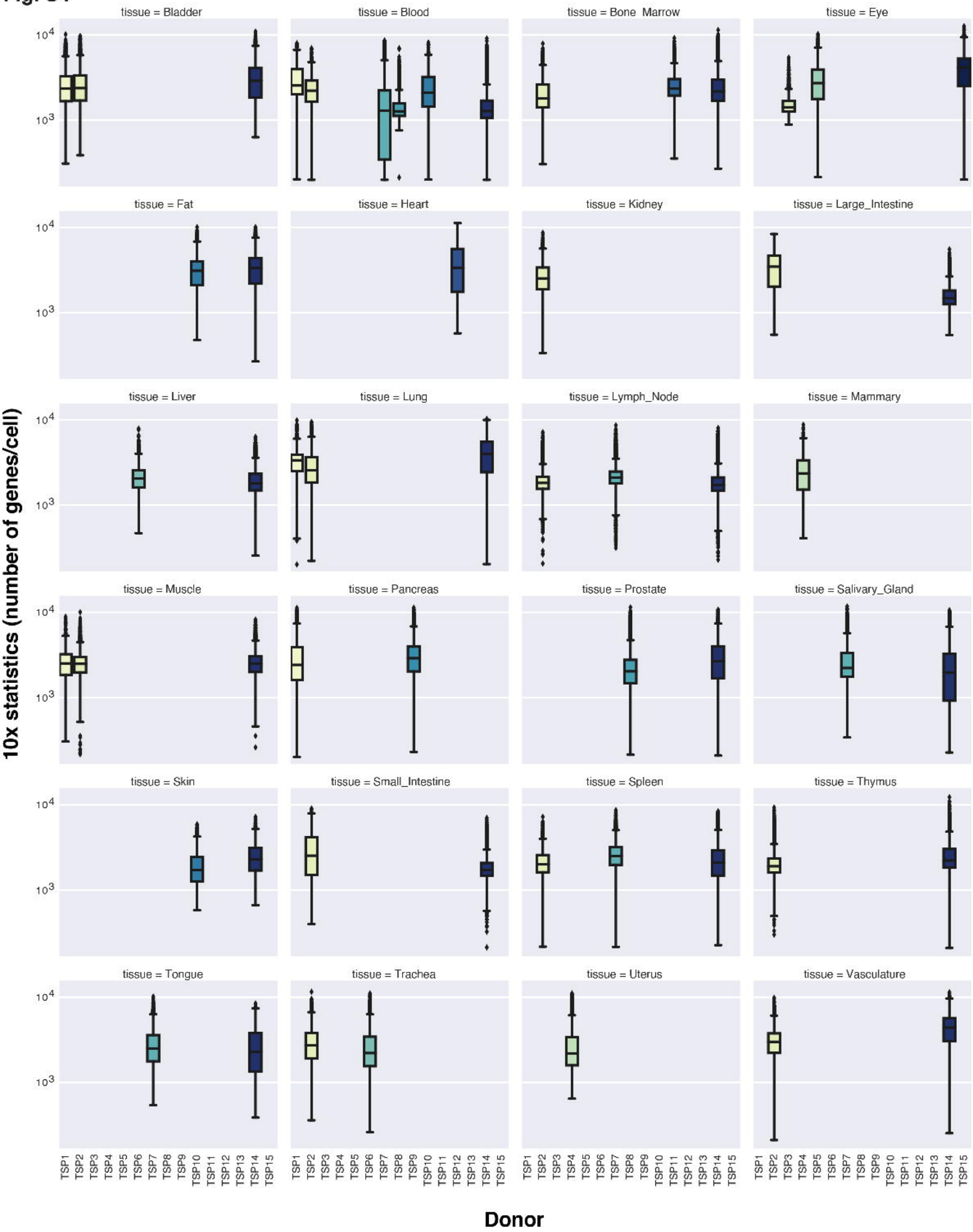

tissue = Bladder
tissue = Blood
tissue = Bone_Marrow
tissue = Eye
tissue = Fat
tissue = Heart
tissue = Kidney
tissue = Large_Intestine
tissue = Liver
tissue = Lung
tissue = Lymph_Node
tissue = Mammary
tissue = Muscle
tissue = Pancreas
tissue = Prostate
tissue = Salivary_Gland
tissue = Skin
tissue = Small_Intestine
tissue = Spleen
tissue = Thymus
tissue = Tongue
tissue = Trachea
tissue = Uterus
tissue = Vasculature
10x statistics (number of genes/cell)
10^4
10^3
TSP1 TSP2 TSP3 TSP4 TSP5 TSP6 TSP7 TSP8 TSP9 TSP10 TSP11 TSP12 TSP13 TSP14 TSP15
Donor

Fig. S5

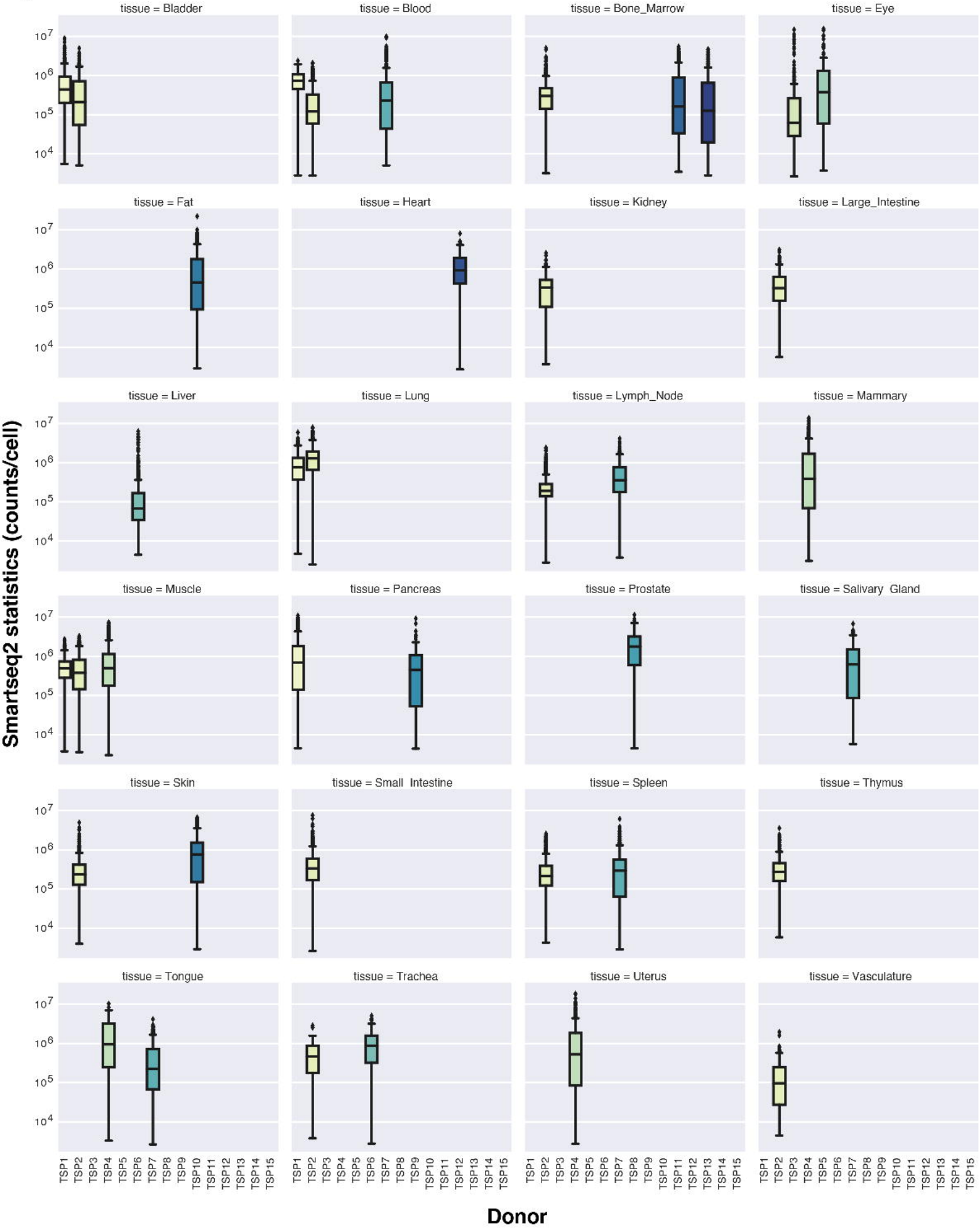

tissue = Bladder
tissue = Blood
tissue = Bone_Marrow
tissue = Eye
tissue = Fat
tissue = Heart
tissue = Kidney
tissue = Large_Intestine
tissue = Liver
tissue = Lung
tissue = Lymph_Node
tissue = Mammary
tissue = Muscle
tissue = Pancreas
tissue = Prostate
tissue = Salivary Gland
tissue = Skin
tissue = Small Intestine
tissue = Spleen
tissue = Thymus
tissue = Tongue
tissue = Trachea
tissue = Uterus
tissue = Vasculature
Smartseq2 statistics (counts/cell)
10^7
10^6
10^5
10^4
TSP1 TSP2 TSP3 TSP4 TSP5 TSP6 TSP7 TSP8 TSP9 TSP10 TSP11 TSP12 TSP13 TSP14 TSP15
Donor

Fig. S6

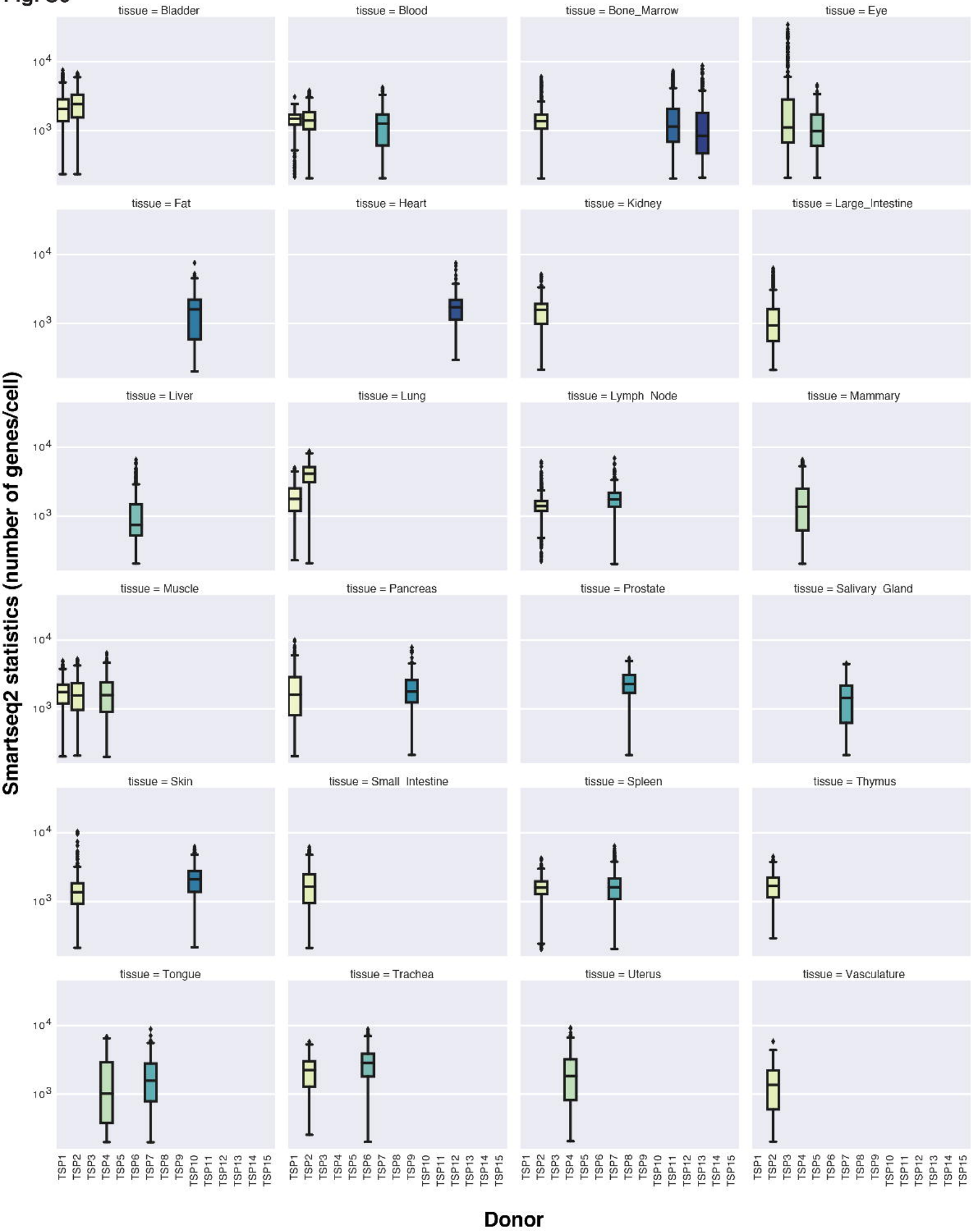

tissue = Bladder
tissue = Blood
tissue = Bone_Marrow
tissue = Eye
tissue = Fat
tissue = Heart
tissue = Kidney
tissue = Large_Intestine
tissue = Liver
tissue = Lung
tissue = Lymph Node
tissue = Mammary
tissue = Muscle
tissue = Pancreas
tissue = Prostate
tissue = Salivary Gland
tissue = Skin
tissue = Small Intestine
tissue = Spleen
tissue = Thymus
tissue = Tongue
tissue = Trachea
tissue = Uterus
tissue = Vasculature
Smartseq2 statistics (number of genes/cell)
10^4
10^3
TSP1 TSP2 TSP3 TSP4 TSP5 TSP6 TSP7 TSP8 TSP9 TSP10 TSP11 TSP12 TSP13 TSP14 TSP15
Donor

Fig. S7

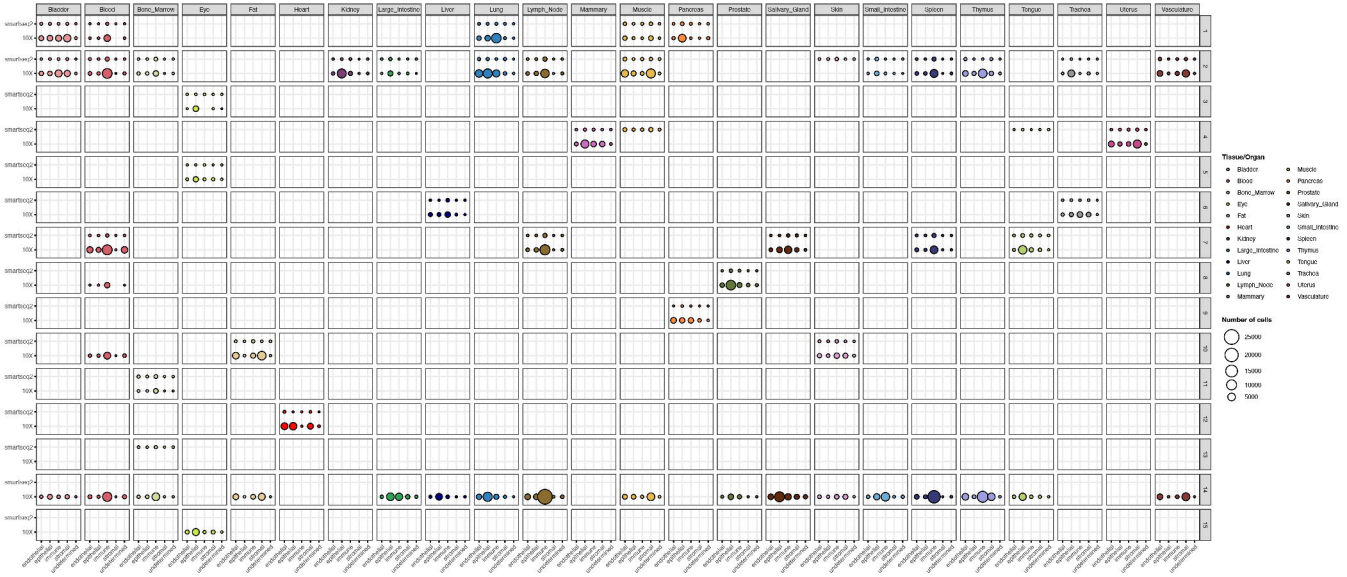

Bladder
Blood
Bone_Marrow
Eye
Fat
Heart
Kidney
Large_Intestine
Liver
Lung
Lymph_Node
Mammary
Muscle
Pancreas
Prostate
Salivary_Gland
Skin
Small_Intestine
Spleen
Thymus
Tongue
Trachea
Uterus
Vasculature
smartseq2
10X
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
endothelial
epithelial
immune
other
stromal
undetermined
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

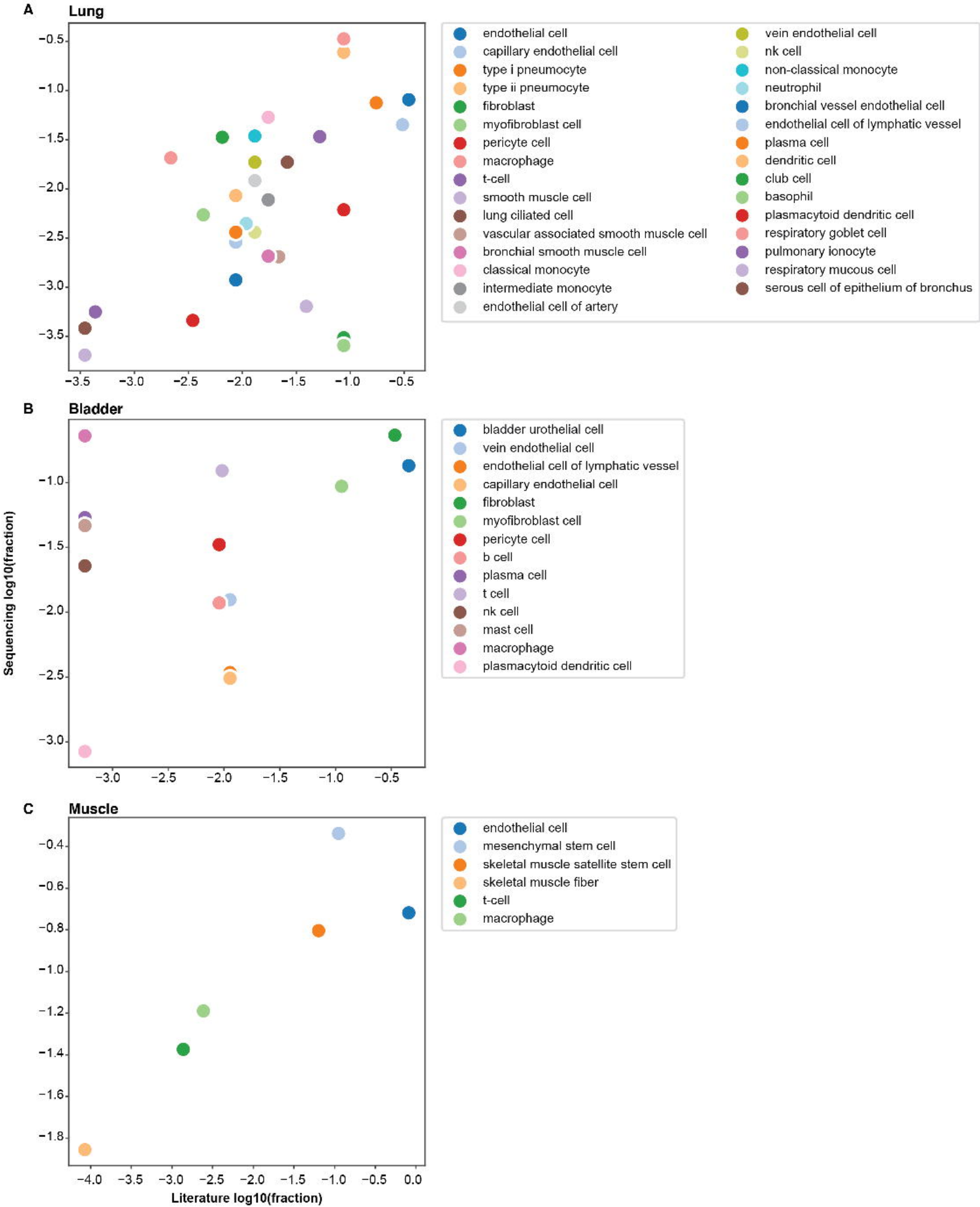

A
Lung
endothelial cell
capillary endothelial cell
type i pneumocyte
type ii pneumocyte
fibroblast
myofibroblast cell
pericyte cell
macrophage
t-cell
smooth muscle cell
lung ciliated cell
vascular associated smooth muscle cell
bronchial smooth muscle cell
classical monocyte
intermediate monocyte
endothelial cell of artery
vein endothelial cell
nk cell
non-classical monocyte
neutrophil
bronchial vessel endothelial cell
endothelial cell of lymphatic vessel
plasma cell
dendritic cell
club cell
basophil
plasmacytoid dendritic cell
respiratory goblet cell
pulmonary ionocyte
respiratory mucous cell
serous cell of epithelium of bronchus
B
Bladder
bladder urothelial cell
vein endothelial cell
endothelial cell of lymphatic vessel
capillary endothelial cell
fibroblast
myofibroblast cell
pericyte cell
b cell
plasma cell
t cell
nk cell
mast cell
macrophage
plasmacytoid dendritic cell
Sequencing log10(fraction)
C
Muscle
endothelial cell
mesenchymal stem cell
skeletal muscle satellite stem cell
skeletal muscle fiber
t-cell
macrophage
Literature log10(fraction)

Fig. S9

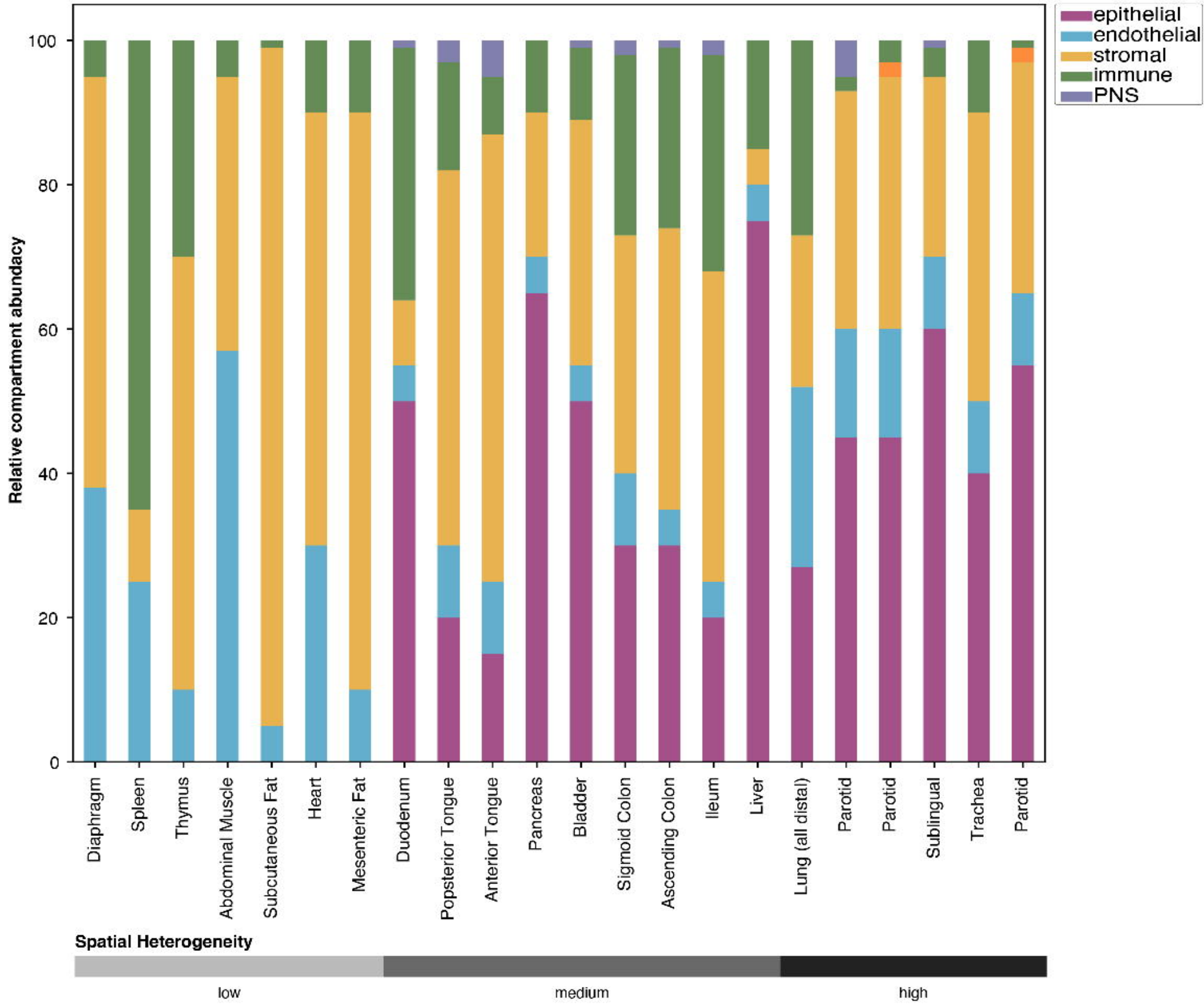

epithelial
endothelial
stromal
immune
PNS
Relative compartment abundacy
100
80
60
40
20
0
Diaphragm
Spleen
Thymus
Abdominal Muscle
Subcutaneous Fat
Heart
Mesenteric Fat
Duodenum
Popsterior Tongue
Anterior Tongue
Pancreas
Bladder
Sigmoid Colon
Ascending Colon
Ileum
Liver
Lung (all distal)
Parotid
Parotid
Sublingual
Trachea
Parotid
Spatial Heterogeneity
low
medium
high

Fig. S10

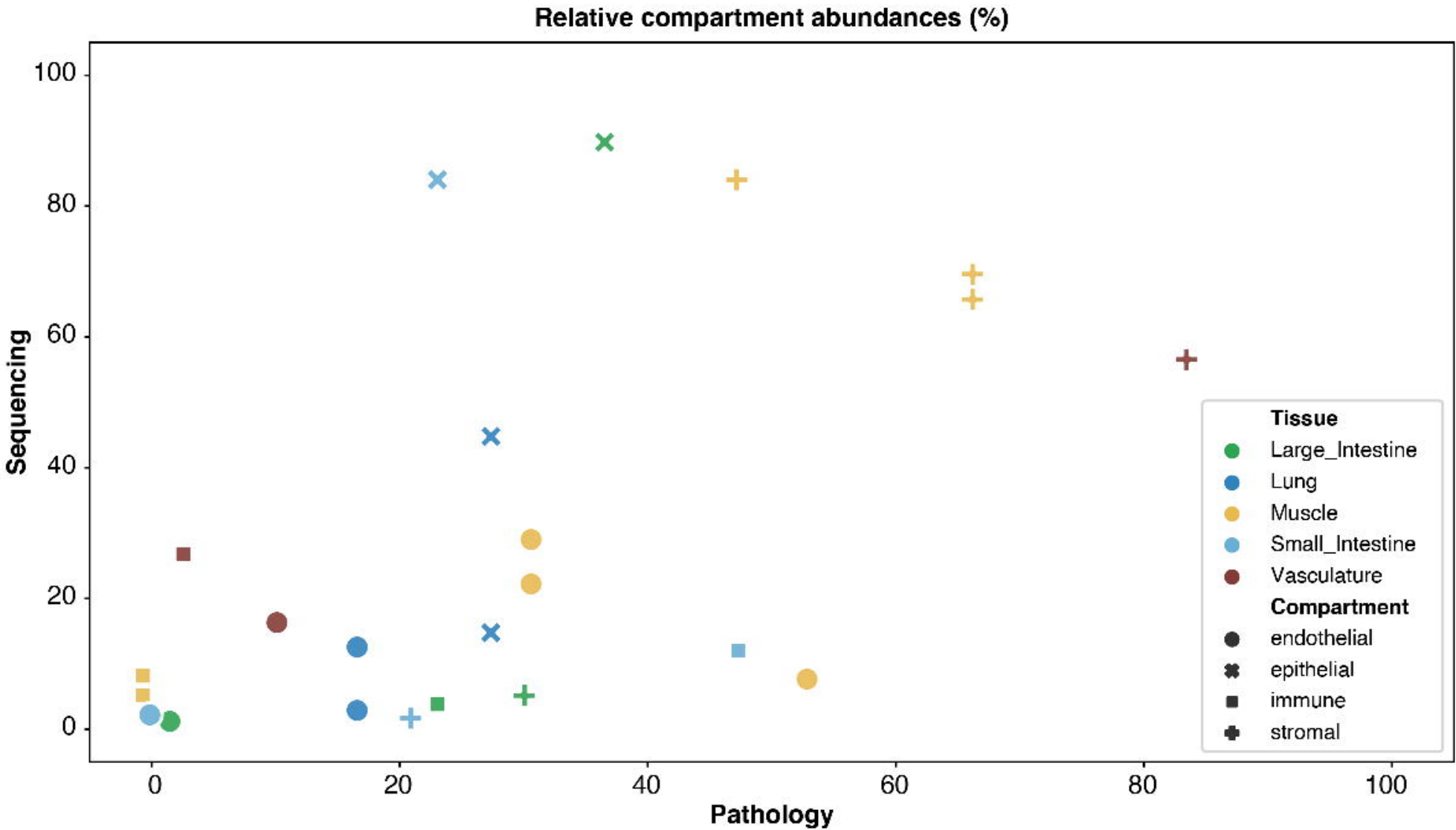

Relative compartment abundances (%)
Sequencing
Pathology
0
20
40
60
80
100
Tissue
Large_Intestine
Lung
Muscle
Small_Intestine
Vasculature
Compartment
endothelial
epithelial
immune
stromal

Fig. S11

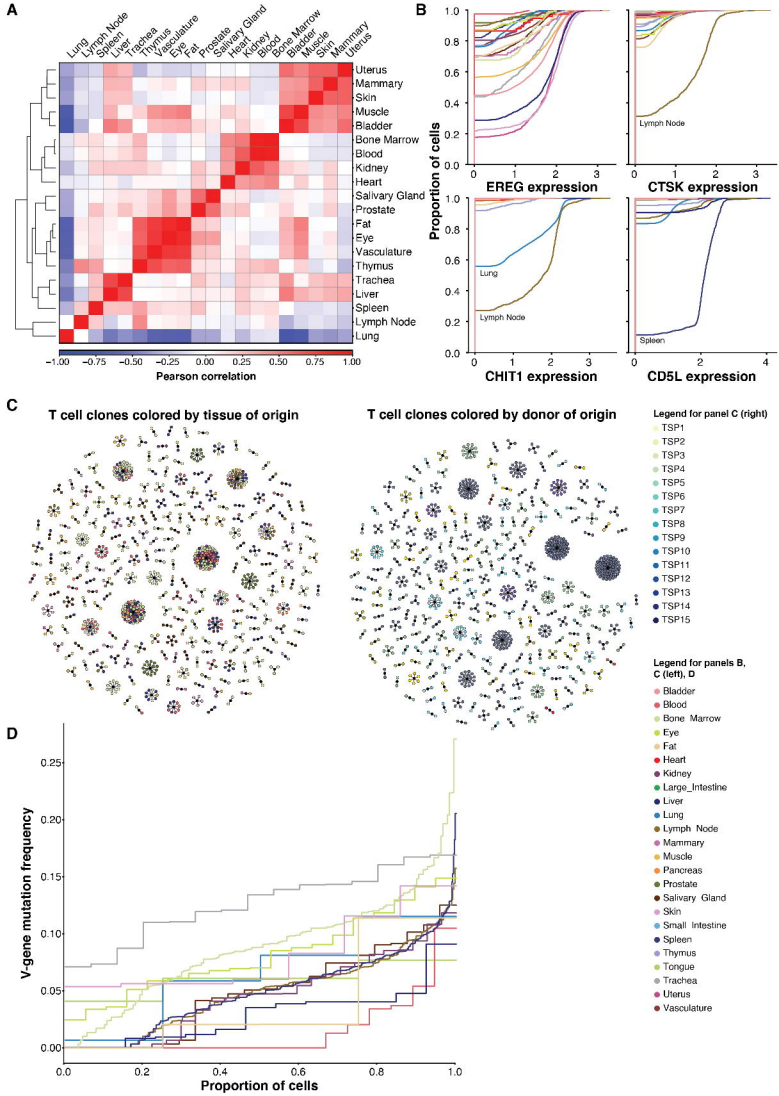

A
Lung
Lymph Node
Spleen
Liver
Trachea
Thymus
Vasculature
Eye
Fat
Prostate
Salivary Gland
Heart
Kidney
Blood
Bone Marrow
Bladder
Muscle
Skin
Mammary
Uterus
Pearson correlation
B
Proportion of cells
EREG expression
CTSK expression
CHIT1 expression
CD5L expression
Lymph Node
Lung
Spleen
C
T cell clones colored by tissue of origin
T cell clones colored by donor of origin
Legend for panel C (right)
TSP1
TSP2
TSP3
TSP4
TSP5
TSP6
TSP7
TSP8
TSP9
TSP10
TSP11
TSP12
TSP13
TSP14
TSP15
Legend for panels B, C (left), D
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
D
V-gene mutation frequency
Proportion of cells

Fig. S12

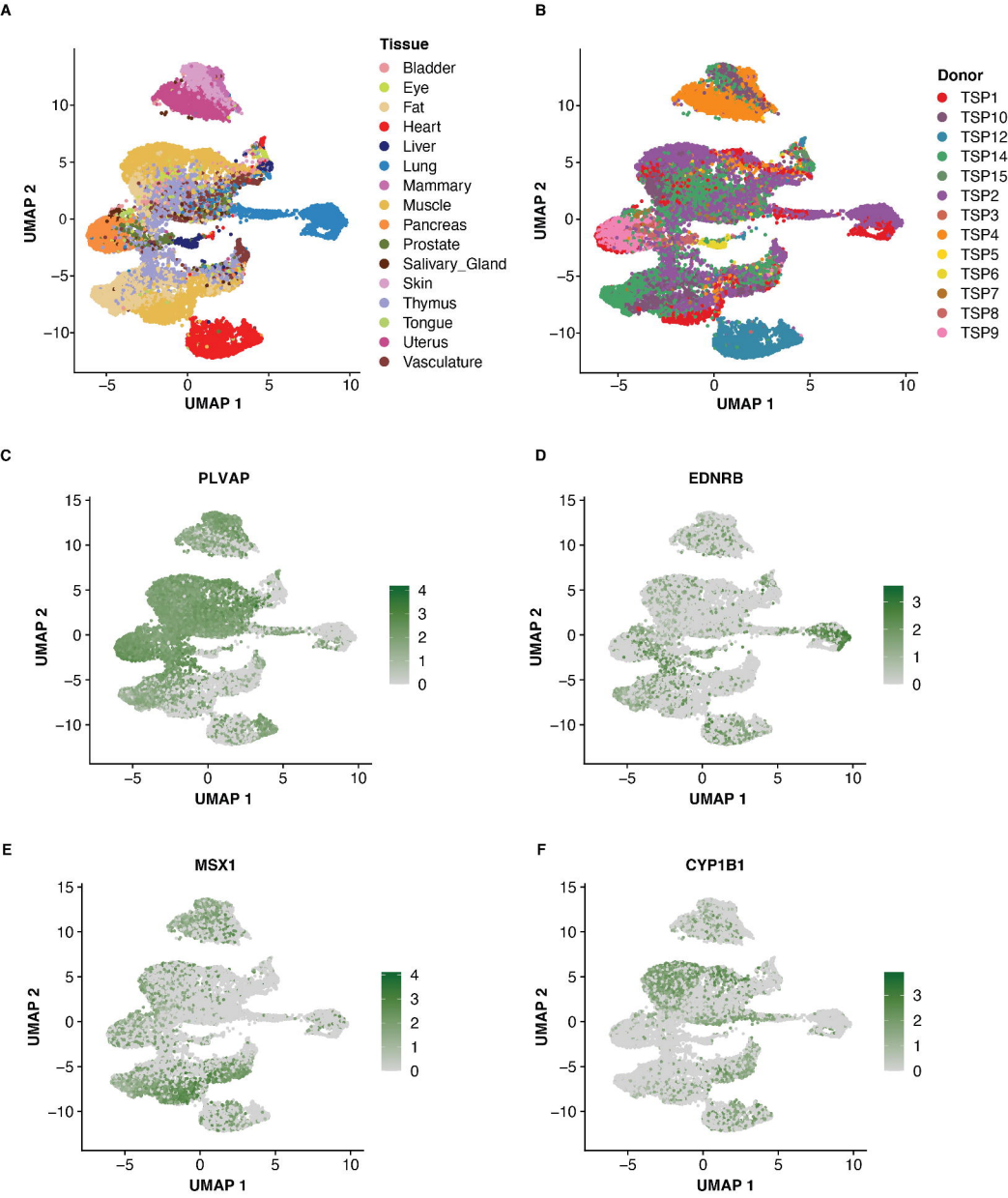
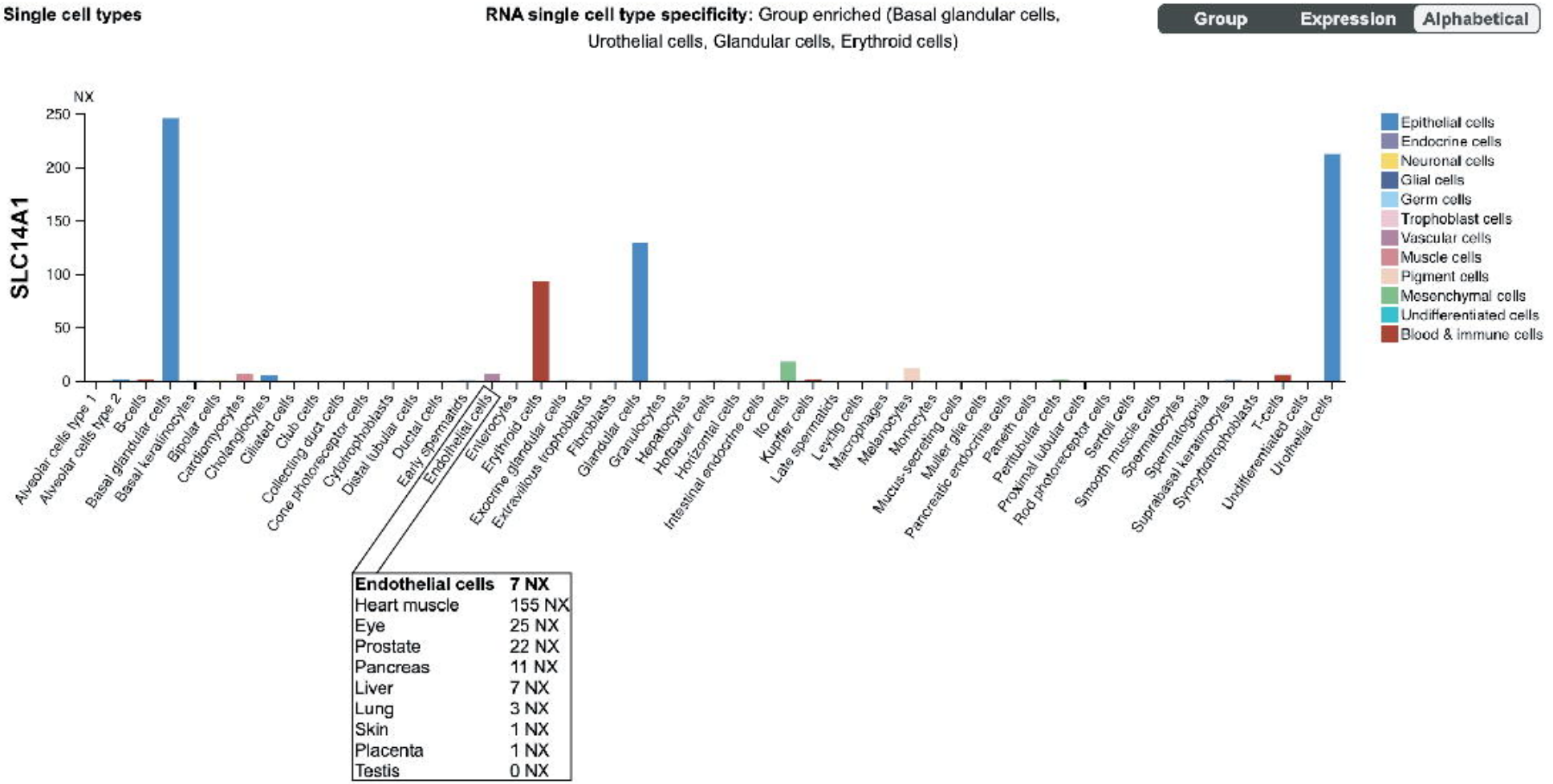

A
Tissue
Bladder
Eye
Fat
Heart
Liver
Lung
Mammary
Muscle
Pancreas
Prostate
Salivary_Gland
Skin
Thymus
Tongue
Uterus
Vasculature
UMAP 1
UMAP 2
B
Donor
TSP1
TSP10
TSP12
TSP14
TSP15
TSP2
TSP3
TSP4
TSP5
TSP6
TSP7
TSP8
TSP9
C
PLVAP
D
EDNRB
E
MSX1
F
CYP1B1

Fig. S13

A


Single cell types
RNA single cell type specificity: Group enriched (Basal glandular cells, Urothelial cells, Glandular cells, Erythroid cells)
Group
Expression
Alphabetical
NX
SLC14A1
250
200
150
100
50
0
Epithelial cells
Endocrine cells
Neuronal cells
Glial cells
Germ cells
Trophoblast cells
Vascular cells
Muscle cells
Pigment cells
Mesenchymal cells
Undifferentiated cells
Blood & immune cells
Alveolar cells type 1
Alveolar cells type 2
B-cells
Basal glandular cells
Basal keratinocytes
Bipolar cells
Cardiomyocytes
Cholangiocytes
Ciliated cells
Club cells
Collecting duct cells
Cone photoreceptor cells
Cytotrophoblasts
Distal tubular cells
Ductal cells
Early spermatids
Endothelial cells
Enterocytes
Erythroid cells
Exocrine glandular cells
Extravillous trophoblasts
Fibroblasts
Glandular cells
Granulocytes
Hepatocytes
Hofbauer cells
Horizontal cells
Intestinal endocrine cells
Ito cells
Kupffer cells
Late spermatids
Leydig cells
Macrophages
Melanocytes
Monocytes
Mucus-secreting cells
Muller glia cells
Pancreatic endocrine cells
Paneth cells
Peritubular cells
Proximal tubular cells
Rod photoreceptor cells
Sertoli cells
Smooth muscle cells
Spermatocytes
Spermatogonia
Suprabasal keratinocytes
Syncytiotrophoblasts
T-cells
Undifferentiated cells
Urothelial cells
Endothelial cells 7 NX
Heart muscle 155 NX
Eye 25 NX
Prostate 22 NX
Pancreas 11 NX
Liver 7 NX
Lung 3 NX
Skin 1 NX
Placenta 1 NX
Testis 0 NX

B

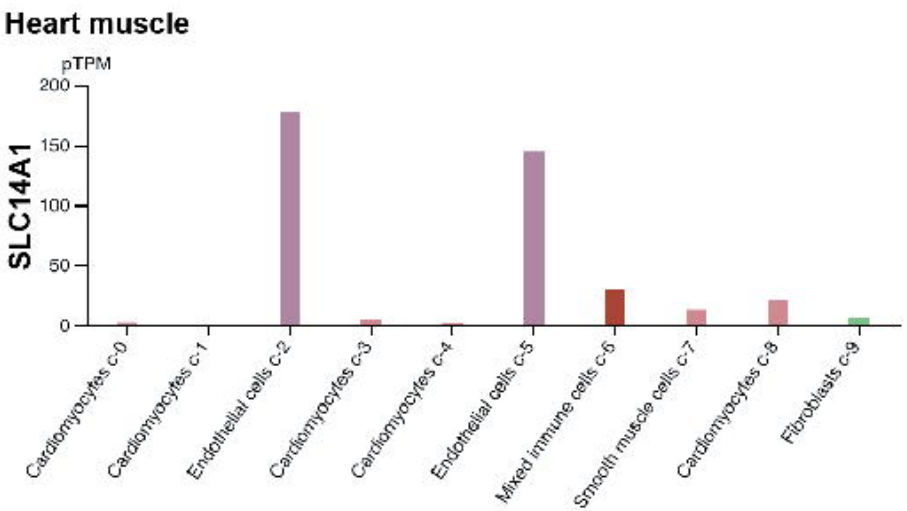

Heart muscle
pTPM
SLC14A1
200
150
100
50
0
Cardiomyocytes c-0
Cardiomyocytes c-1
Endothelial cells c-2
Cardiomyocytes c-3
Cardiomyocytes c-4
Endothelial cells c-5
Mixed immune cells c-6
Smooth muscle cells c-7
Cardiomyocytes c-8
Fibroblasts c-9

Fig. S14

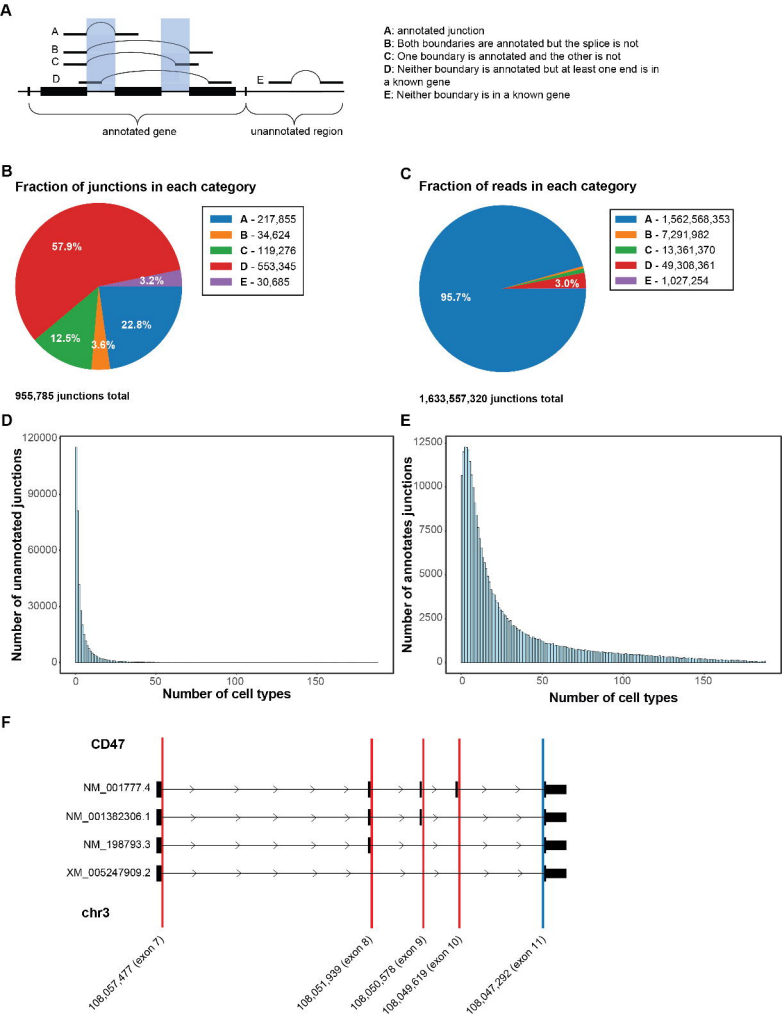
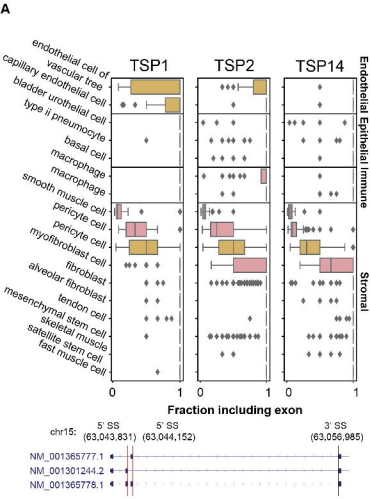
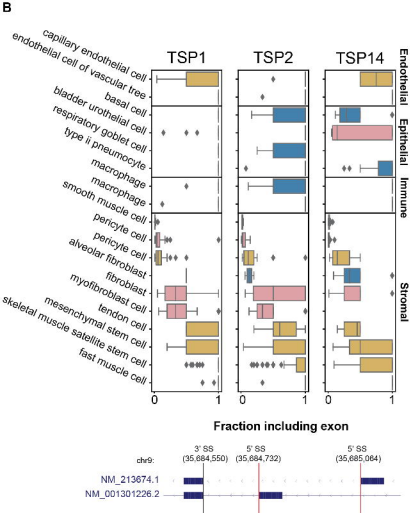

A
A
B
C
D
E
annotated gene
unannotated region
A: annotated junction
B: Both boundaries are annotated but the splice is not
C: One boundary is annotated and the other is not
D: Neither boundary is annotated but at least one end is in a known gene
E: Neither boundary is in a known gene
B
Fraction of junctions in each category
A - 217,855
B - 34,624
C - 119,276
D - 553,345
E - 30,685
57.9%
3.2%
22.8%
12.5%
3.6%
955,785 junctions total
C
Fraction of reads in each category
A - 1,562,568,353
B - 7,291,982
C - 13,361,370
D - 49,308,361
E - 1,027,254
95.7%
3.0%
1,633,557,320 junctions total
D
Number of unannotated junctions
Number of cell types
E
Number of annotates junctions
Number of cell types
F
CD47
NM_001777.4
NM_001382306.1
NM_198793.3
XM_005247909.2
chr3
108,057,477 (exon 7)
108,051,939 (exon 8)
108,050,578 (exon 9)
108,049,619 (exon 10)
108,047,292 (exon 11)

Fig. S15


A
TSP1
TSP2
TSP14
endothelial cell of vascular tree
capillary endothelial cell
bladder urothelial cell
type ii pneumocyte
basal cell
macrophage
macrophage
smooth muscle cell
pericyte cell
pericyte cell
myofibroblast cell
fibroblast
alveolar fibroblast
tendon cell
mesenchymal stem cell
skeletal muscle satellite stem cell
fast muscle cell
Endothelial
Epithelial
Immune
Stromal
0
1
Fraction including exon
chr15:
5' SS (63,043,831)
5' SS (63,044,152)
3' SS (63,056,985)
NM_001365777.1
NM_001301244.2
NM_001365778.1


B
TSP1
TSP2
TSP14
capillary endothelial cell
endothelial cell of vascular tree
basal cell
bladder urothelial cell
respiratory goblet cell
type ii pneumocyte
macrophage
macrophage
smooth muscle cell
pericyte cell
pericyte cell
alveolar fibroblast
fibroblast
myofibroblast cell
tendon cell
mesenchymal stem cell
skeletal muscle satellite stem cell
fast muscle cell
Endothelial
Epithelial
Immune
Stromal
0
1
Fraction including exon
chr9:
3' SS (35,684,550)
5' SS (35,684,732)
5' SS (35,685,084)
NM_213674.1
NM_001301226.2

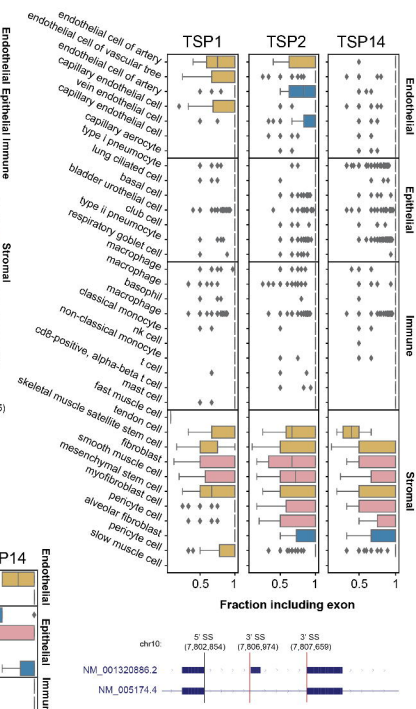

C
TSP1
TSP2
TSP14
endothelial cell of artery
endothelial cell of vascular tree
endothelial cell of artery
vein endothelial cell
capillary endothelial cell
capillary endothelial cell
capillary aerocyte
type i pneumocyte
lung ciliated cell
basal cell
bladder urothelial cell
club cell
type ii pneumocyte
respiratory goblet cell
macrophage
macrophage
basophil
macrophage
classical monocyte
non-classical monocyte
nk cell
cd8-positive, alpha-beta t cell
t cell
mast cell
fast muscle cell
skeletal muscle satellite stem cell
tendon cell
smooth muscle cell
fibroblast
mesenchymal stem cell
myofibroblast cell
pericyte cell
alveolar fibroblast
pericyte cell
slow muscle cell
Endothelial
Epithelial
Immune
Stromal
0.5
1
Fraction including exon
chr10:
5' SS (7,802,854)
3' SS (7,806,974)
3' SS (7,807,659)
NM_001320886.2
NM_005174.4

Fig. S16

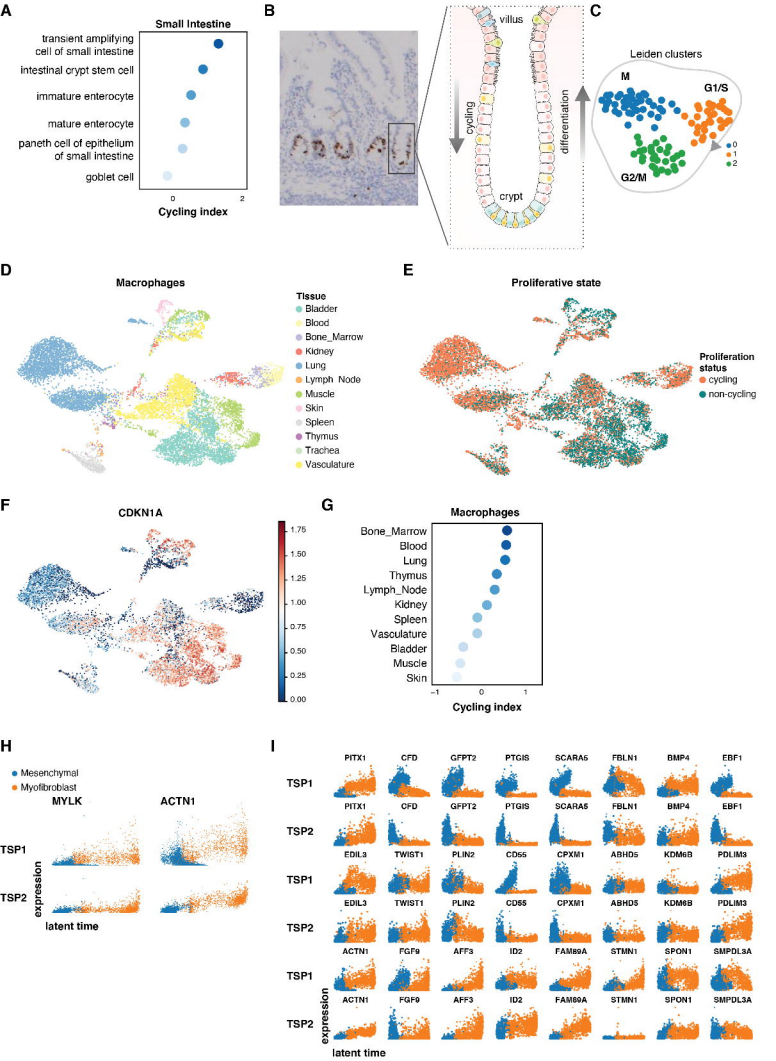
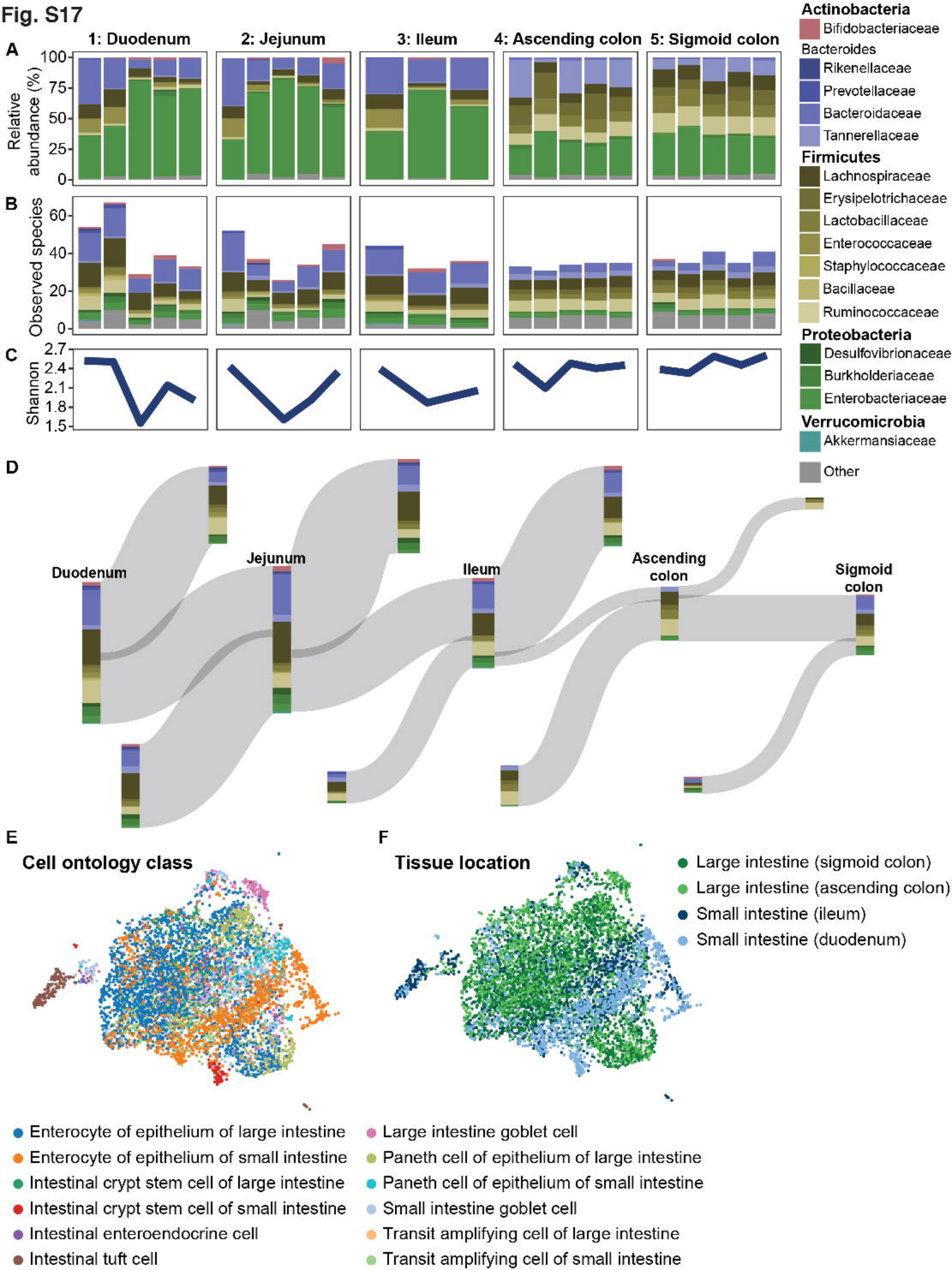

A
Small Intestine
transient amplifying cell of small intestine
intestinal crypt stem cell
immature enterocyte
mature enterocyte
paneth cell of epithelium of small intestine
goblet cell
0
2
Cycling index
B
villus
cycling
differentiation
crypt
C
Leiden clusters
M
G1/S
G2/M
0
1
2
D
Macrophages
Tissue
Bladder
Blood
Bone_Marrow
Kidney
Lung
Lymph_Node
Muscle
Skin
Spleen
Thymus
Trachea
Vasculature
E
Proliferative state
Proliferation status
cycling
non-cycling
F
CDKN1A
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
G
Macrophages
Bone_Marrow
Blood
Lung
Thymus
Lymph_Node
Kidney
Spleen
Vasculature
Bladder
Muscle
Skin
−1
0
1
Cycling index
H
Mesenchymal
Myofibroblast
MYLK
ACTN1
TSP1
TSP2
expression
latent time
I
PITX1
CFD
GFPT2
PTGIS
SCARA5
FBLN1
BMP4
EBF1
EDIL3
TWIST1
PLIN2
CD55
CPXM1
ABHD5
KDM6B
PDLIM3
ACTN1
FGF9
AFF3
ID2
FAM89A
STMN1
SPON1
SMPDL3A
TSP1
TSP2
expression
latent time

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