The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans

The Tabula Sapiens Consortium[*]

Abstract

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 that may or may not be used in the different cell types of a multicellular organism. Although nearly every cell in the body has essentially the same genome, each cell type makes different use of that genome and expresses a subset of all possible genes. This has motivated efforts to characterize the molecular composition of various cell types within humans and multiple model organisms, both by transcriptional and proteomic approaches. We created a human reference atlas comprising nearly 500,000 cells from 24 different tissues and organs, many from the same donor. This atlas enabled molecular characterization of more than 400 cell types, their distribution across tissues, and tissue-specific variation in gene expression.

RATIONALE: One caveat to current approaches to make cell atlases is that individual organs are often collected at different locations, collected from different donors, and processed using different protocols. 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.

RESULTS: We collected multiple tissues from individual human donors and performed coordinated single-cell transcriptome analyses on live cells. The donors come from a range of ethnicities, are balanced by gender, have a mean age of 51 years, and have a variety of medical backgrounds. Tissue experts used a defined cell ontology terminology to annotate cell types consistently across the different tissues, leading to a total of 475 distinct cell types with reference transcriptome profiles. The full dataset can be explored online with the cellxgene tool.

Data were collected for the bladder, blood, bone marrow, eye, fat, heart, kidney, large intestine, liver, lung, lymph node, mammary, muscle, pancreas, prostate, salivary gland, skin, small intestine,

[*]Corresponding author: Stephen R. Quake (steve@quake-lab.org).

Author contributions: See below, where authors are arranged by working group and contributions.

spleen, thymus, tongue, trachea, uterus, and vasculature. Fifty-nine separate specimens in total were collected, processed, and analyzed, and 483,152 cells passed quality control filtering. On a per-compartment basis, the dataset includes 264,824 immune cells, 104,148 epithelial cells, 31,691 endothelial cells, and 82,478 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.

The Tabula Sapiens also provided an opportunity to densely and directly sample the human microbiome throughout the gastrointestinal tract. The intestines from two donors were sectioned into five regions: the duodenum, jejunum, ileum, and ascending and sigmoid colon. Each section was transected, and three to nine samples were collected from each location, followed by amplification and sequencing of the 16S ribosomal RNA gene.

CONCLUSION: The Tabula Sapiens has revealed discoveries relating to shared behavior and subtle, organ-specific differences across cell types. We found T cell clones shared between organs and characterized organ-dependent hypermutation rates among B cells. Endothelial cells and macrophages are shared across tissues, often showing subtle but clear differences in gene expression. We found an unexpectedly large and diverse amount of cell type–specific RNA splice variant usage and discovered and validated many previously undefined splices. The intestinal microbiome was revealed to have nonuniform species distributions down to the 3-inch (7.62-cm) length scale. These are but a few examples of how the Tabula Sapiens represents a broadly useful reference to deeply understand and explore human biology at cellular resolution.

Abstract

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

Graphical Abstract

Overview of Tabula Sapiens. Molecular characterization of cell types using single-cell transcriptome sequencing is revolutionizing cell biology and enabling new insights into the physiology of human organs. We created a human reference atlas comprising nearly 500,000 cells from 24 different tissues and organs, many from the same donor. This multimodal atlas enabled molecular characterization of more than 400 cell types.
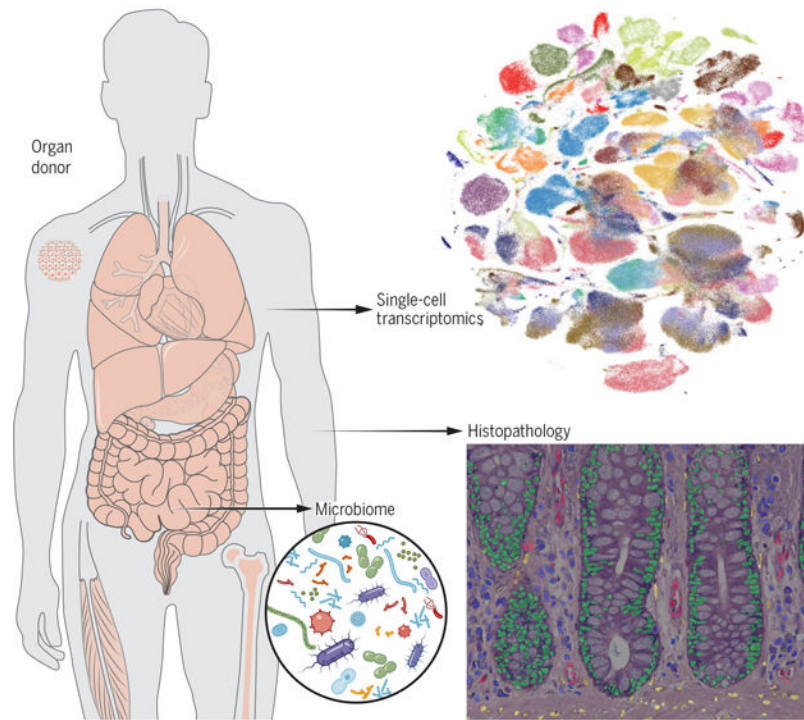
Although the genome is often called the blueprint of an organism, it is perhaps more accurate to describe it as a parts list composed of the various genes that may or may not be used in the different cell types of a multicellular organism. Although nearly every cell in the body has essentially the same genome, each cell type makes different use of that genome and expresses a subset of all possible genes (1). Therefore, the genome in and of itself does not provide an understanding of the molecular complexity of the various cell types of that organism. This has motivated efforts to characterize the molecular composition of various cell types within humans and multiple model organisms, both by transcriptional (2) and proteomic (3, 4) approaches.

Although such efforts are yielding insights (5–7), one caveat to current approaches is that individual organs are often collected at different locations or from different donors (8), are 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 analyze 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 TSP1 to TSP15) 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 five tissues from two other donors (Fig. 1). We also collected smaller numbers of tissues from a further 11

donors, creating biological replicates for nearly all tissues. The donors come from a range of ethnicities, are balanced by gender, have a mean age of 51 years, and have a variety of medical backgrounds (table S1). Single-cell transcriptome sequencing was performed with both fluorescence-activated cell sorting (FACS)–sorted cells in well plates with smart-seq2 amplification as well as 10x microfluidic droplet capture and amplification for each tissue (fig. S1). Tissue experts used a defined cell ontology terminology to annotate cell types consistently across the different tissues (13), leading to a total of 475 distinct cell types with reference transcriptome profiles (tables S2 and S3). The full dataset can be explored online with the cellxgene tool through the Tabula Sapiens data portal (14).

Data were collected for the bladder, blood, bone marrow, eye, fat, heart, kidney, large intestine, liver, lung, lymph node, mammary, muscle, pancreas, prostate, salivary gland, skin, small intestine, spleen, thymus, tongue, trachea, uterus, and vasculature. Fifty-nine separate specimens in total were collected, processed, and analyzed, and 483,152 cells passed quality control (QC) filtering (figs. S2 to S7 and table S2). On a per-compartment basis, the dataset includes 264,824 immune cells, 104,148 epithelial cells, 31,691 endothelial cells, and 82,478 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 insights into variations in alternative splicing.

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

Immune cells: Variation in gene expression across tissues and a shared lineage history

The Tabula Sapiens can be used to study differences in the gene expression programs and lineage histories of cell types that are shared across tissues. We analyzed tissue differences in the 36,475 macrophages distributed among 20 tissues because tissue-resident

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

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

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

Endothelial cell subtypes with tissue-specific gene expression programs

As another example of analyzing shared cell types across organs, we focused on endothelial cells (ECs). Although ECs are often categorized as a single cell type, they exhibit differences in morphology, structure, and immunomodulatory and metabolic phenotypes depending on their tissue of origin. We discovered that tissue specificity is also reflected in their transcriptomes because ECs mainly cluster by tissue of origin (table S6). Uniform manifold approximation and projection (UMAP) analysis (fig. S12A) revealed that the 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).

Notably, ECs from the thymus, vasculature, prostate, and eye were similarly distributed across several clusters, which suggests not only similarity in transcriptional profiles but in

their sources of heterogeneity. Differential gene expression analysis between ECs from 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 previously undescribed markers determined by this analysis, SLC14A1 (solute carrier family 14 member 1) appears to be a specific marker for endothelial cells in the heart, whose expression was independently validated with data from the Human Protein Atlas (18) (fig. S13).

Lung ECs formed two distinct populations, which is in line with the aerocyte (aCap - EDNRB+) and general capillary (gCap - PLVAP+) cells described in the mouse and human lung (19) (fig. S12, C and 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 CYPIB1+ population enriched in metabolic genes, which suggests the presence of functional specialization in the muscle vasculature (fig. S12, E and F).

Alternative splice variants are cell type specific

We used SICILIAN (20) to identify alternative splice junctions in Tabula Sapiens using both 10× and smart-seq2 sequencing technologies and found a total of 955,785 junctions (fig. S14, A to E, and table S7). Of these, 217,855 were previously annotated, so our data provide independent validation of 61% of the total junctions cataloged in the entire RefSeq database. Although annotated junctions made up only 22.8% of the total junctions, they represent 93% of total reads, which indicates that previously annotated junctions tend to be expressed at higher levels than previously unidentified junctions. We additionally found 34,624 junctions between previously annotated 3′ and 5′ splice sites (3.6%). We identified 119,276 junctions between a previously annotated site and a previously unannotated site in the gene (12.4%). This leaves 584,030 putative junctions for which both splice sites were previously unannotated—i.e., ~61% of the total detected junctions. Most of these have at least one end in a known gene (94.7%), whereas the remainder represent potential previously undescribed splice variants from unannotated regions (5.3%). In the absence of independent validation, we conservatively characterized all of the unannotated splices as putative previously unknown junctions. We then used the GTEx database (21) to seek independent corroborating evidence of these putative junctions and found that reads corresponding to nearly one-third of these previously unknown junctions can be found within the GTEx data (table S7); this corresponds to >300,000 previously undefined validated splice valiants revealed by the Tabula Sapiens.

Hundreds of splice variants are used in a highly cell type–specific fashion; these can be explored in the cellxgene browser (14), which uses a statistic called SpliZ (22). We focus on two examples of cell type–specific splicing of two well-studied genes: MYL6 and CD47. Similar cell type–specific splice usage was also observed with TPM1, TPM2, and ATP5F1C, three other genes with well-characterized splice variants (fig. S15).

MYL6 is an essential light chain (ELC) for myosin and is highly expressed in all tissues and compartments. Yet, splicing of MYL6, particularly that involving the inclusion or exclusion of exon 6 (Fig. 4A), varies in a cell type- and compartment-specific manner (Fig. 4B). Although the isoform excluding exon 6 has previously been mainly described in phasic smooth muscle (23), we discovered 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. These previously unknown, compartment-specific expression patterns of the two MYL6 isoforms are reproduced in multiple individuals from the Tabula Sapiens dataset (Fig. 4, A and B).

CD47 is a multispanning membrane protein involved in many cellular processes, including angiogenesis and cell migration and as a "do not eat me" signal to macrophages (24). Differential use of exons 7 to 10 (Fig. 4C and fig. S14F) composes a variably long cytoplasmic tail (25). Immune cells—but also stromal and endothelial cells—have a distinct, consistent splicing pattern in CD47 that dominantly excludes two proximal exons and splicing directly to exon 8. In contrast to other compartments, epithelial cells exhibit a different splicing pattern that increases the length of the cytoplasmic tail by splicing more commonly to exon 9 and exon 10 (Fig. 4D). Characterization of the splicing programs of CD47 in single cells may have implications for understanding the differential signaling activities of CD47 and for therapeutic manipulation of CD47 function.

Cell state dynamics can be inferred from a single time point

Although the Tabula Sapiens was created from a single moment in time for each donor, it is possible to infer dynamic information from the data. Cell division is an important transient change of internal cell state, and we computed a cycling index for each cell type to identify actively proliferating versus quiescent or postmitotic cell states. Rapidly dividing progenitor cells had among the highest cycling indices, whereas cell types from the endothelial and stromal compartments, which are known to be largely quiescent, had low cycling indices (Fig. 5A). In intestinal tissue, transient amplifying cells and the crypt stem cells divide rapidly in the intestinal crypts to give rise to terminally differentiated cell types of the villi (26). These cells were ranked with the highest cycling indices, whereas terminally differentiated cell types, such as the goblet cells, had the lowest ranks (fig. S16A). To complement the computational analysis of cell cycling, we performed immunostaining of intestinal tissue for the MKI67 protein (commonly referred to as K_i-67) and confirmed that transient amplifying cells abundantly express this proliferation marker (fig. S16, B and C), which supports the conclusion 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 with macrophages found in the bladder, skin, and muscle (fig. S16, D to G). Consistent with this finding, the expression values of cyclin-dependent kinase (CDK) inhibitors (in particular the gene CDKN1A), which block the cell cycle, have the lowest overall expression in macrophages from tissues with high cycling indices (fig. S16F).

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

Unexpected spatial variation in the microbiome

The Tabula Sapiens provided an opportunity to densely and directly sample the human microbiome throughout the gastrointestinal tract. The intestines from donors TSP2 and TSP14 were sectioned into five regions: the duodenum, jejunum, ileum, and ascending and sigmoid colon (Fig. 6A). Each section was transected, and three to nine samples were collected from each location, followed by amplification and sequencing of the $16S$ ribosomal RNA (rRNA) gene. Uniformly, there was a high (~10 to 30%) relative abundance of Proteobacteria, particularly Enterobacteriaceae (Fig. 6B), even in the colon. Samples from each of the duodenum, jejunum, and ileum were largely distinct from one another, with samples exhibiting individual patterns of blooming or absence of certain families (Fig. 6B). These data reveal that the microbiota are patchy, even at a 3-inch (7.62-cm) length scale. We observed similar heterogeneity in both donors (fig. S17, A to C). In the small intestine, richness (number of observed species) was also variable and was negatively correlated with the relative abundance of Burkholderiaceae (Fig. 6B); in TSP2, the Proteobacteria phylum was dominated by Enterobacteriaceae, which was present at >30% in all samples at a level negatively correlated with richness (fig. S17, A to C). In a comparison of species from adjacent regions across the gut, a large fraction of species was specific to each region (Fig. 6C), reflecting the patchiness. These data are derived from only two donor samples, and further conclusions about the statistics and extent of microbial patchiness will require larger studies.

We analyzed host immune cells in conjunction with the spatial microbiome data; UMAP clustering analysis revealed that the small intestine T cell pool from TSP14 contained a population with distinct transcriptomes (Fig. 6D). The most significant transcriptional differences in T cells between the small and large intestine were genes associated with trafficking, survival, and activation (Fig. 6E and table S8). For example, expression of the long noncoding RNA MALAT1, which affects the regulatory function of T cells, and CCR9, which mediates T lymphocyte development and migration to the intestine (31), were high only in the small intestine, whereas 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, whereas there was some degree of overlap for each of the two cell types for either location (fig. S17, E and F). The site-specific composition of the microbiome in the intestine, paired with distinct T cell populations at each site, helps define local host-microbe interactions that occur in the gastrointestinal tract and is likely reflective of a gradient of physiological conditions that influence hostmicrobe dynamics.

Conclusion

The Tabula Sapiens is part of a growing set of data that, when analyzed together, will enable many interesting comparisons of both a biological and technical nature. Studying particular cell types across organs, datasets, and species will yield new biological insights—as shown with fibroblasts (32). Similarly, comparing fetal human cell types (33) with those determined in this work in adults may give insight into the loss of plasticity from early development to maturity. Having multiorgan data from individual donors may facilitate the development of methods to compare diverse datasets and yield understanding of technical artifacts from various approaches (8, 9, 34, 35). The Tabula Sapiens has enabled discoveries relating to shared behavior and organ-specific differences across cell types. For example, we found T cell clones shared between organs and characterized organ-dependent hypermutation rates among resident B cells. Endothelial cells and macrophages are cell types that are shared across tissues but often show subtle tissue-specific differences in gene expression. We found an unexpectedly large and diverse amount of cell type–specific RNA splice variant usage and discovered and validated many previously undiscovered splices. These are but a few examples of how the Tabula Sapiens represents a broadly useful reference to deeply understand and explore human biology at cellular resolution.

Materials and methods summary

Fresh, whole, and nontransplantable organs, or 1- to 2-cm^3 organ samples, were obtained from surgery and then transported on ice by courier to tissue expert laboratories, where they were immediately prepared for transcriptome sequencing. Single-cell suspensions were prepared for 10× Genomics 3′ V3.1 droplet-based sequencing and for FACS-sorted 384-well plate smart-seq2. Preparation began with dissection, digestion with enzymes, and physical manipulation; tissue-specific details are available in the complete materials and methods (12). Cell suspensions from some organs were normalized by major cell compartment (epithelial, endothelial, immune, and stromal) using antibody-labeled 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 demultiplexed 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 unique molecular identifier (UMI) counts or 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 two-dimensional (2D) 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 that uses seven different automated annotation methods. For complete materials and methods, see the supplementary materials (12).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Appendix

Appendix

The Tabula Sapiens Consortium Overall project direction and coordination: Robert C. Jones[1], Jim Karkanias[2], Mark A. 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 Cvijovi [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 C. Steffes[17], Varun R. Subramaniam[31], Aditi Swarup[31], Michael Swift[1], Kyle J. Travaglini[3], Will Van Treuren[16], Emily Trimm[26], Stefan Veizades[19,39], Sivakamasundari Vijayakumar[18], Kim Chi Vo[29], Sevahn K. Vorperian[1],[40], Wanxin Wang[29], Hannah N. W. Weinstein[38], Juliane Winkler[41], Timothy T. H. Wu[3], Jamie Xie[38], Andrea R. Yung[3], Yue Zhang[3]

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

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

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

Expert cell type annotation: Marcela Alcántara-Hernández[16], Jane Antony[18], Charles K. F. Chan[18,22], Charles A. Chang[23], Alex Colville[21,24], Sheela Crasta[1,2], Rebecca Culver[25], Les Dethlefsen[43], Camille Ezran[3], Astrid Gillich[3], Yan Hang[23,28], Po-Yi Ho[16], Juan C. Irwin[29], SoRi Jang[3], Aaron M. Kershner[18], William Kong[18], Maya E. Kumar[17], Angera H. Kuo[18], Rebecca Leylek[16], Shixuan Liu[344], 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], Hay ley 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. Reiman[16,43,46], Justin L. Sonnenburg[16], Bruce Wang[20] Albert Wu[31], Sean M. Wu[19,39], Tony Wyss-Coray[21,24]

[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, CA, USA. [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.

Data and materials availability:

The entire dataset can be explored interactively at https://tabula-sapiens-portal.ds.czbiohub.org/ (14). The code used for the analysis is available from Zenodo (36). Gene counts and metadata are available from figshare (37) and have been deposited in the Gene Expression Omnibus (GSE201333); 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 (38). SpliZ scores are available from figshare (39). To preserve the donors' genetic privacy. we require a data transfer agreement to receive the raw sequence reads. The data transfer agreement is available upon request.

REFERENCES AND NOTES

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

2. Regev A et al. , The human cell atlas. eLife 6, e27041 (2017). doi: 10.7554/eLife.27041 [PubMed: 29206104]

3. Uhlén M et al. , Tissue-based map of the human proteome. Science 347, 1260419 (2015). doi: 10.1126/science.1260419 [PubMed: 25613900]

4. Thul PJ et al. , A subcellular map of the human proteome. Science 356, eaal3321 (2017). doi: 10.1126/science.aal3321 [PubMed: 28495876]

5. Pisco AO, Tojo B, McGeever A, Single-Cell Analysis for Whole-Organism Datasets. Annu. Rev. Biomed. Data Sci. 4, 207–226 (2021). doi: 10.1146/annurev-biodatasci-092820-031008 [PubMed: 34465173]

6. Aldridge S, Teichmann SA, Single cell transcriptomics comes of age. Nat. Commun. 11, 4307 (2020). doi: 10.1038/s41467-020-18158-5 [PubMed: 32855414]

7. Wu AR, Wang J, Streets AM, Huang Y, Single-Cell Transcriptional Analysis. Annu. Rev. Anal. Chem. 10, 439–462 (2017). doi: 10.1146/annurev-anchem-061516-045228

8. Han X et al. , Construction of a human cell landscape at single-cell level. Nature 581, 303–309 (2020). doi: 10.1038/s41586-020-2157-4 [PubMed: 32214235]

9. He S et al. , Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. Genome Biol. 21, 294 (2020). doi: 10.1186/s13059-020-02210-0 [PubMed: 33287869]

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). doi: 10.1038/s41586-018-0590-4 [PubMed: 30283141]

11. The Tabula Muris Consortium, A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. Nature 583, 590–595 (2020). doi: 10.1038/s41586-020-2496-1 [PubMed: 32669714]

12. Materials and methods are available as supplementary materials.

13. Wang S et al. , Leveraging the Cell Ontology to classify unseen cell types. Nat. Commun. 12, 5556 (2021). doi: 10.1038/s41467-021-25725-x [PubMed: 34548483]

14. The Tabula Sapiens Consortium, "Tabula Sapiens Data Portal"; https://tabula-sapiens-portal.ds.czbiohub.org.

15. Steinert EM et al. , Quantifying Memory CD8 T Cells Reveals Regionalization of Immunosurveillance. Cell 161, 737–749 (2015). doi: 10.1016/j.cell.2015.03.031 [PubMed: 25957682]

16. Davies LC, Taylor PR, Tissue-resident macrophages: Then and now. Immunology 144, 541–548 (2015). doi: 10.1111/imm.12451 [PubMed: 25684236]

17. Godfrey DI, Koay H-F, McCluskey J, Gherardin NA, The biology and functional importance of MAIT cells. Nat. Immunol. 20, 1110–1128 (2019). doi: 10.1038/s41590-019-0444-8 [PubMed: 31406380]

18. Karlsson M et al. , A single-cell type transcriptomics map of human tissues. Sci. Adv. 7, eabh2169 (2021). doi: 10.1126/sciadv.abh2169 [PubMed: 34321199]

19. Gillich A et al. , Capillary cell-type specialization in the alveolus. Nature 586,785–789 (2020). doi: 10.1038/s41586-020-2822-7 [PubMed: 33057196]

20. Dehghannasiri R, Olivieri JE, Damljanovic A, Salzman J, Specific splice junction detection in single cells with SICILIAN. Genome Biol. 22, 219 (2021). doi: 10.1186/s13059-021-02434-8 [PubMed: 34353340]

21. Lonsdale J et al. , The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585 (2013). doi: 10.1038/ng.2653 [PubMed: 23715323]

22. Olivieri JE et al. , RNA splicing programs define tissue compartments and cell types at single-cell resolution. eLife 10, e70692 (2021). doi: 10.7554/eLife.70692 [PubMed: 34515025]

23. Fisher SA, Vascular smooth muscle phenotypic diversity and function. Physiol. Genomics 42A, 169–187 (2010). doi: 10.1152/physiolgenomics.00111.2010 [PubMed: 20736412]

24. Barkal AA et al. , CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. Nature 572, 392–396 (2019). doi: 10.1038/s41586-019-1456-0 [PubMed: 31367043]

25. Hayat SMG et al. , CD47: Role in the immune system and application to cancer therapy. Cell. Oncol. 43,19–30 (2020). doi: 10.1007/s13402-019-00469-5

26. Gehart H, Clevers H, Tales from the crypt: New insights into intestinal stem cells. Nat. Rev. Gastroenterol. Hepatol. 16, 19–34 (2019). doi: 10.1038/s41575-018-0081-y [PubMed: 30429586]

27. La Manno G et al. , RNA velocity of single cells. Nature 560, 494–498 (2018). doi: 10.1038/s41586-018-0414-6 [PubMed: 30089906]

28. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ, Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. 38, 1408–1414 (2020). doi: 10.1038/s41587-020-0591-3 [PubMed: 32747759]

29. Murphy ACH, Young PW, The actinin family of actin cross-linking proteins – A genetic perspective. Cell Biosci. 5, 49 (2015). doi:10.1186/s13578-015-0029-7 [PubMed: 26312134]

30. Li R, Li X, Hagood J, Zhu M-S, Sun X, Myofibroblast contraction is essential for generating and regenerating the gas-exchange surface. J. Clin. Invest. 130, 2859–2871 (2020). doi: 10.1172/JCI132189 [PubMed: 32338642]

31. Uehara S, Grinberg A, Farber JM, Love PE, A role for CCR9 in T lymphocyte development and migration. J. Immunol. 168, 2811–2819 (2002). doi: 10.4049/jimmunol.168.6.2811 [PubMed: 11884450]

32. Buechler MB et al. , Cross-tissue organization of the fibroblast lineage. Nature 593, 575–579 (2021). doi: 10.1038/s41586-021-03549-5 [PubMed: 33981032]

33. Cao J et al. , A human cell atlas of fetal gene expression. Science 370, eaba7721 (2020). doi: 10.1126/science.aba7721 [PubMed: 33184181]

34. Eraslan G et al. , Single-nucleus cross-tissue molecular reference maps to decipher disease gene function. bioRxiv2021.07.19.452954 [Preprint] (2021). doi: 10.1101/2021.07.19.452954

35. Domínguez Conde C et al. , Cross-tissue immune cell analysis reveals tissue-specific adaptations and clonal architecture in humans. bioRxiv2021.04.28.441762 [Preprint] (2021). doi: 10.1101/2021.04.28.441762

36. The Tabula Sapiens Consortium, czbiohub/tabula-sapiens: v1.0, version manuscript1, Zenodo (2022); doi: 10.5281/zenodo.6069683

37. The Tabula Sapiens Consortium, Tabula Sapiens Single-Cell Dataset, version 4, figshare (2021); doi: 10.6084/m9.figshare.14267219.v4

38. The Tabula Sapiens Consortium, Tabula Sapiens H&E Image Collection, version 2, figshare (2021); doi: 10.6084/m9.figshare.14962947.v2

39. The Tabula Sapiens Consortium, Tabula Sapiens Splicing, figshare (2021); doi: 10.6084/ m9.figshare.14977281.v1
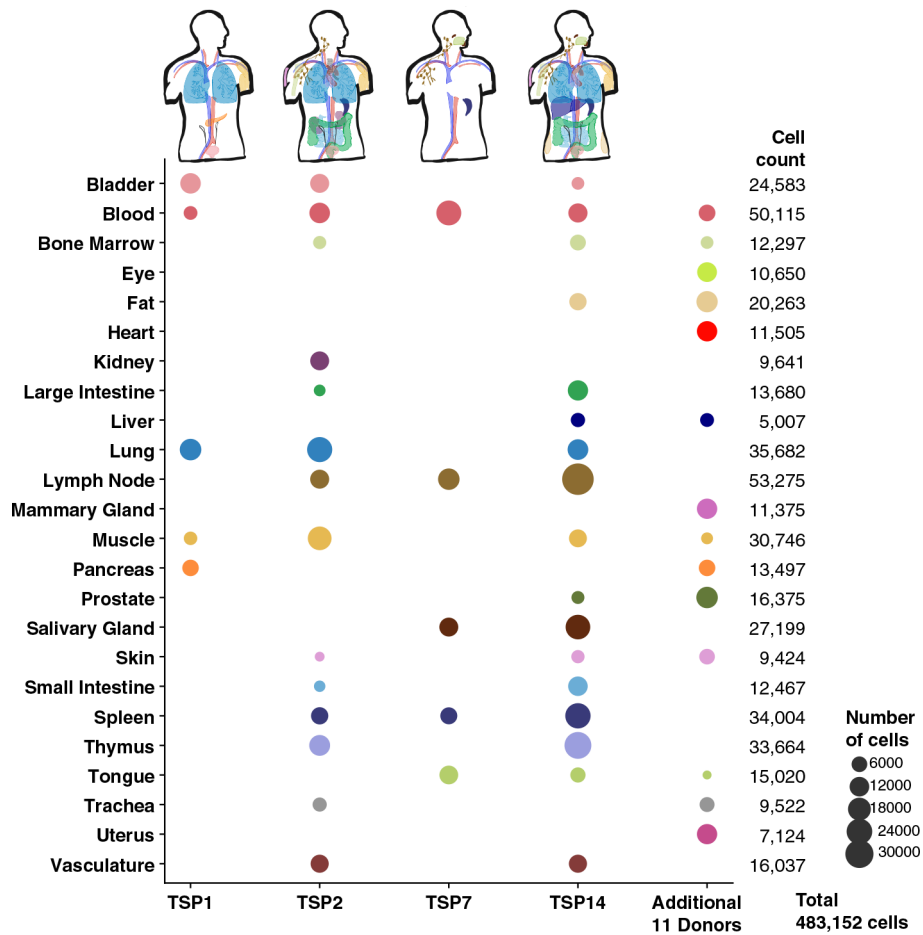
Fig. 1. Overview of Tabula Sapiens.

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 supplementary materials and methods 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 11 donors column, and the total number of cells for each organ are shown in the final column on the right.
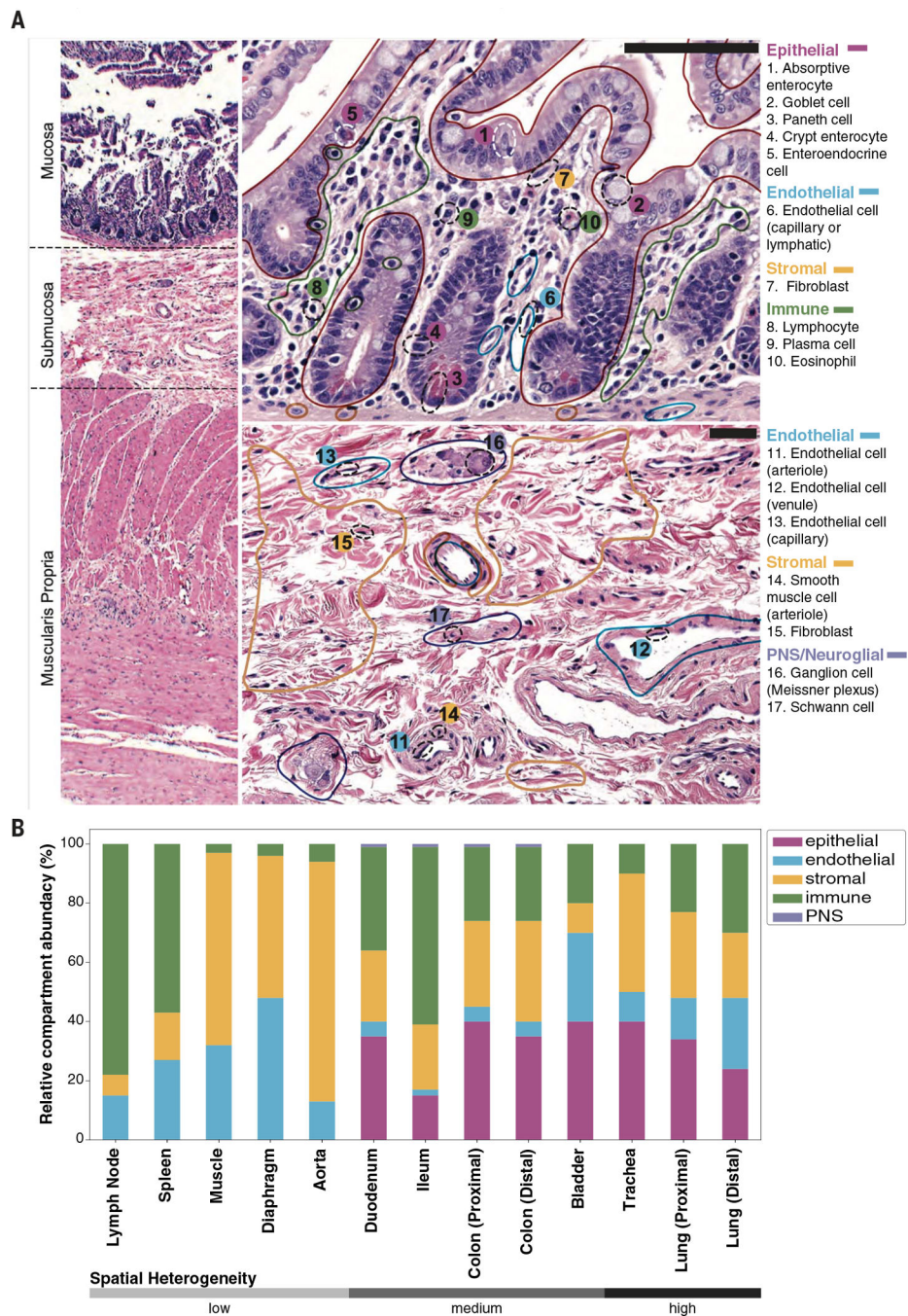
Fig. 2. Comparison of single-cell transcriptomics with conventional histology.
Clinical pathology was performed on nine tissues from donors TSP2 and TSP14. (**A**) H&E–stained image used for histology of the colon from TSP2, with compartments (solid colored lines) and individual cell types (dashed black 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 (A) and (B).
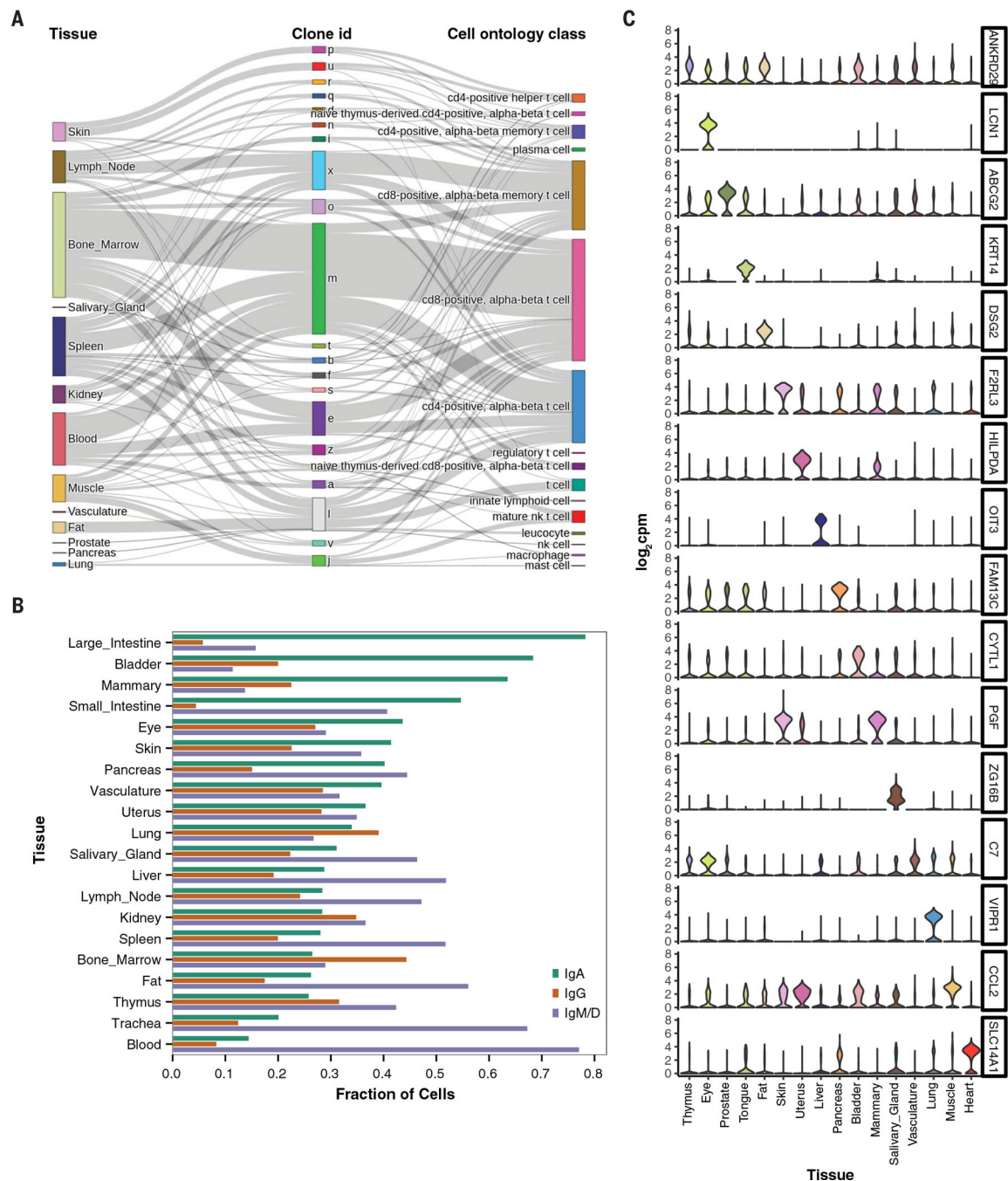
Fig. 3. Analysis of immune and endothelial cell types shared across tissues.
(**A**) Illustration of clonal distribution of T cells across multiple tissues. The majority of T cell clones are found in multiple tissues and represent a variety of T cell subtypes. nk cell, natural killer cell. (**B**) Prevalence of B cell isotypes across tissues, ordered by decreasing abundance of IgA. (**C**) Expression levels of tissue-specific endothelial markers, shown as violin plots, in the entire dataset. Many of the markers are highly tissue specific and typically were 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.
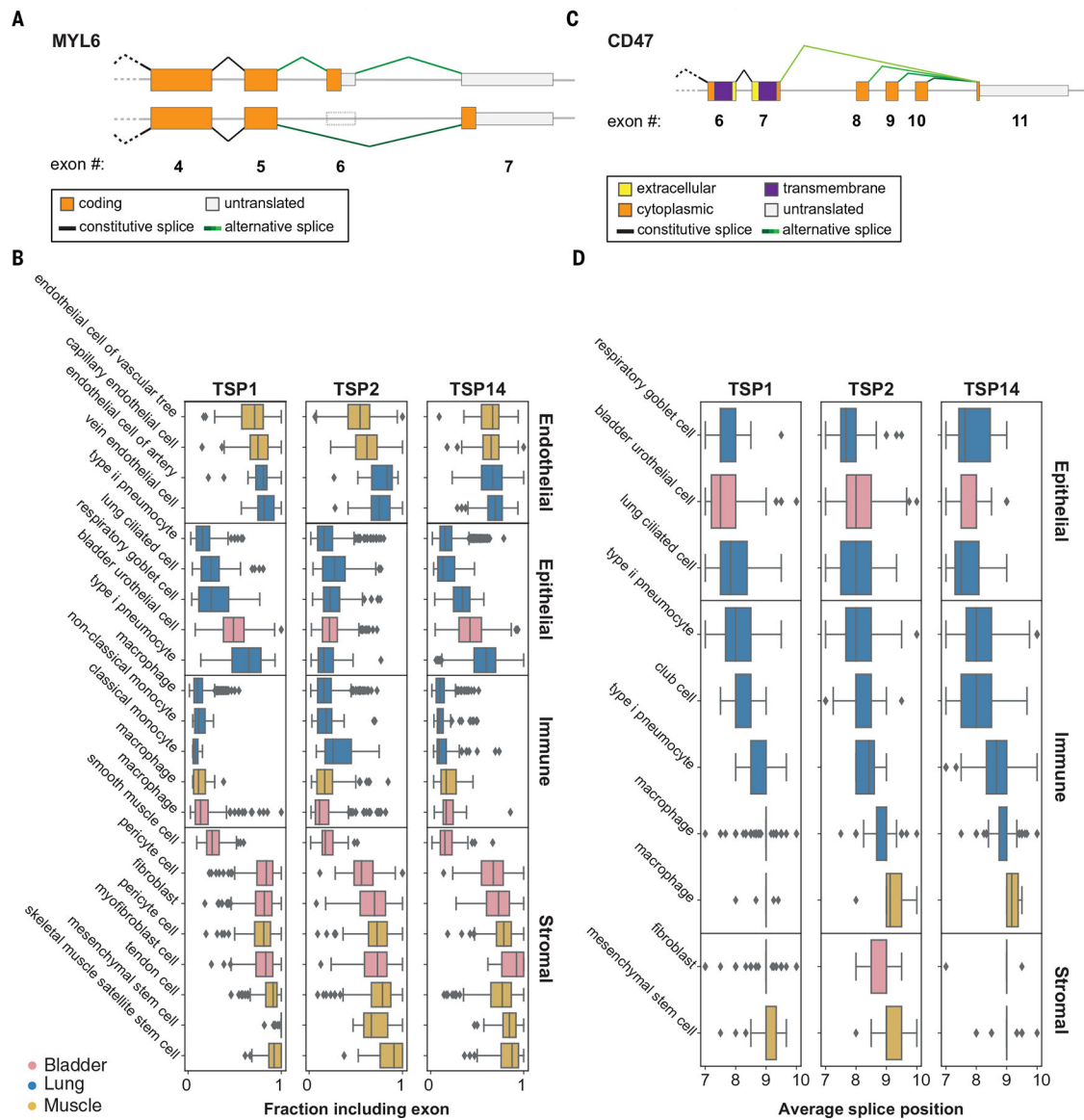
Fig. 4. Alternative splicing analysis.

(**A** and **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 >10 reads mapping to the exon-skipping event. Horizontal box plots in (B) show the distribution of exon inclusion in each cell type. (**C** and **D**) Alternative splicing in CD47 involves one 5′ splice site (exon 11; 108,047,292) and four 3′ splice sites. Horizontal box plots in (D) show the distribution of weighted averages of alternative 3′ splice sites in each cell type. Epithelial cells tend to use exons closer to the 5′ splice site compared with immune and stromal cells. Boxes are grouped by compartment and colored by tissue.
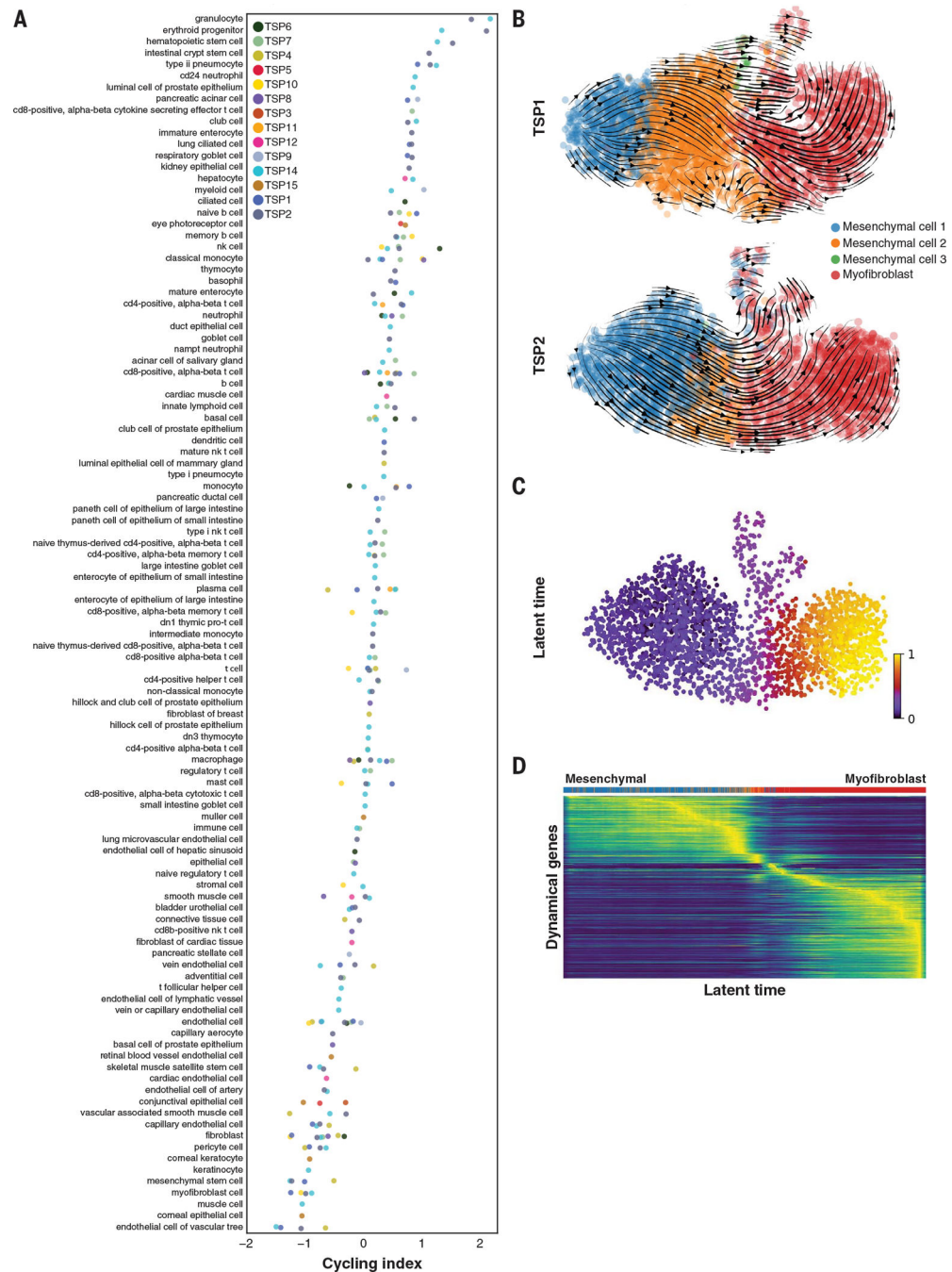
Fig. 5. Dynamic changes in cell state.
(**A**) Cell types ordered by magnitude of cell cycling index per donor (each a separate color), with the most highly proliferative at the top and quiescent cells at the bottom of the list. (**B**) RNA velocity analysis demonstrating mesenchymal-to-myofibroblast 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** and **D**) Latent time analysis of the mesenchymal-to-myofibroblast transition in the bladder, demonstrating stereotyped changes in gene expression trajectory.
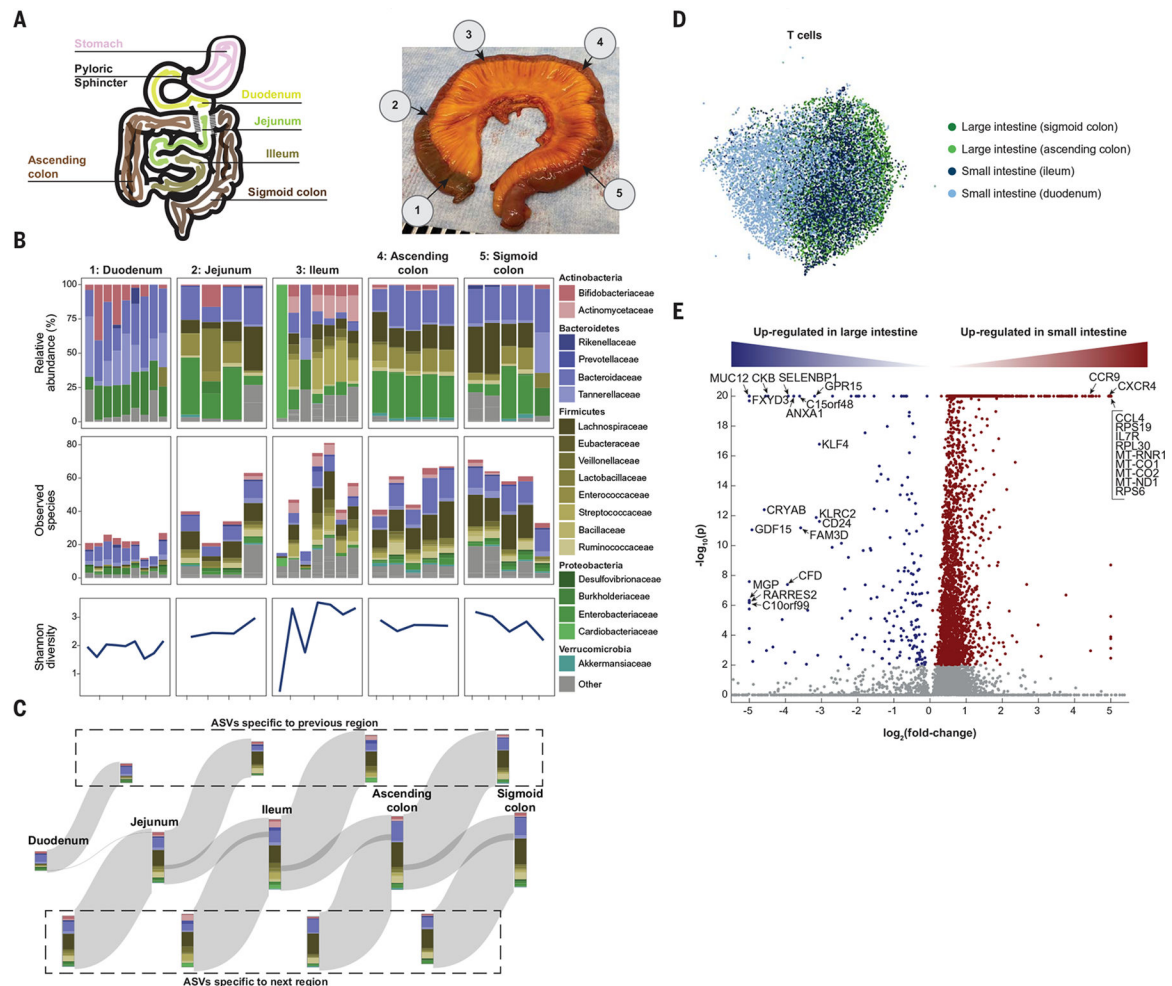
Fig. 6. High-resolution view highlights patchiness of the gut microbiome.
(**A**) Schematic (left) and photo (right) of the colon from donor TSP2, with numbers 1
to 5 representing microbiota sampling locations. (**B**) Relative abundances and richness
(number of observed species) at the family level in each sampling location, as determined
by 16*S* rRNA sequencing. The Shannon diversity, a metric of evenness, mimics richness.
Variability in relative abundance and/or richness or Shannon diversity was higher in the
duodenum, jejunum, and ileum compared with the ascending and sigmoid colon. (**C**) A
Sankey diagram showing the inflow and outflow of microbial species from each section
of the gastrointestinal tract. The stacked bar for each gastrointestinal section represents the
number of observed species in each family as the union of all sampling locations for that
section. The stacked bar flowing out represents gastrointestinal species not found in the
subsequent section, and the stacked bar flowing into each gastrointestinal section represents
the species not found in the previous section. ASVs, amplicon sequence variants. (**D**) UMAP
clustering of T cells reveals distinct transcriptome profiles in the distal and proximal small
and large intestines. (**E**) Dots in volcano plot highlight genes up-regulated in the large (left)
and small (right) intestines. Labeled dots include genes with known roles in trafficking,
survival, and activation.