

RESEARCH ARTICLE SUMMARY

HUMAN CELL ATLAS

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

The Tabula Sapiens Consortium*

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 dis-

tribution 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, 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.

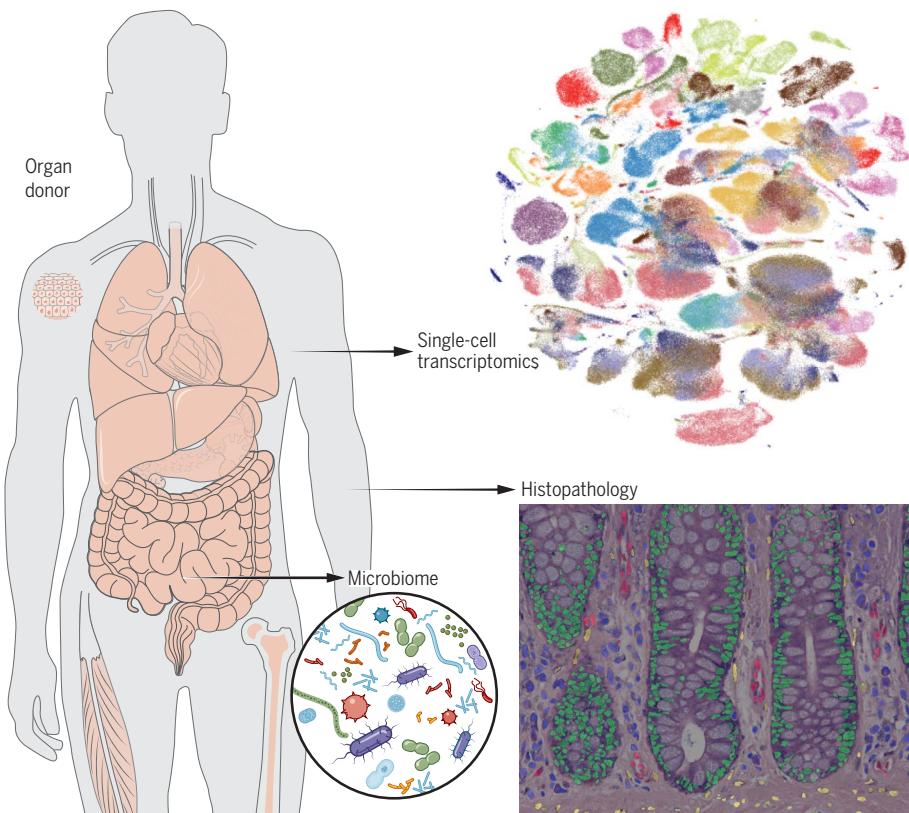
The Tabula Sapiens also provided an opportunity to densely and directly sample the human microbiome throughout the gastrointestinal tract. The intestines from two donors were sectioned into five regions: the duodenum, jejunum, ileum, and ascending and sigmoid colon. Each section was transected, and three 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. ■

The complete list of authors and their affiliations is available in the full article online.

*Corresponding author: Stephen R. Quake (steve@quake-lab.org)
Cite this article as The Tabula Sapiens Consortium, *Science* 376, eabl4896 (2022). DOI: 10.1126/science.abl4896

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abl4896>



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.

RESEARCH ARTICLE

HUMAN CELL ATLAS

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

The Tabula Sapiens Consortium*

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.

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 10 \times 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 mucosal-associated 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

All authors and their affiliations appear at the end of this paper.

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



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.

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 path-

ogens and commensals at the 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).

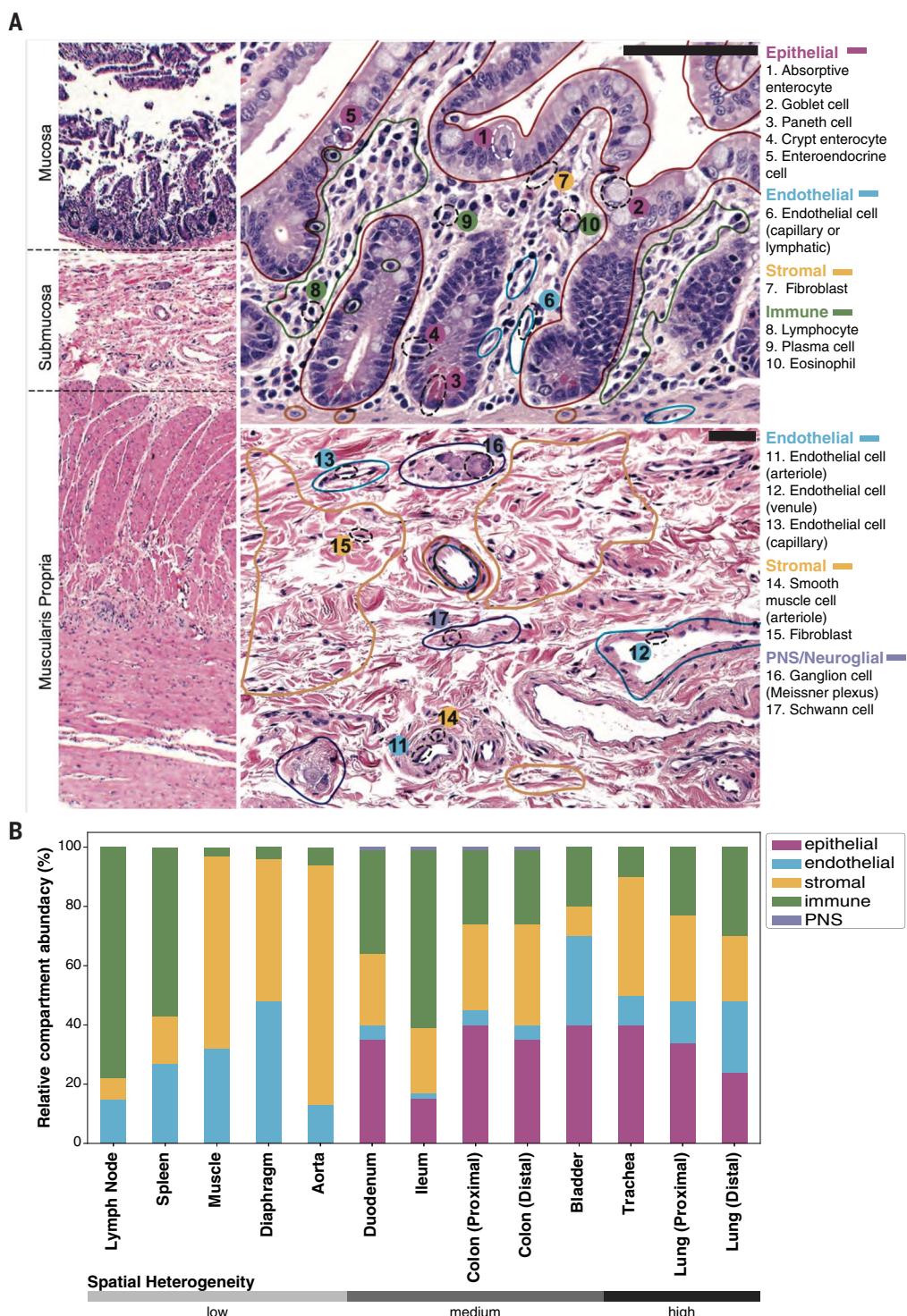
Lung ECs formed two distinct populations, which is in line with the aCap (aCap - EDNRB+) and general capillary (gCap - PLVAP+) cells described in the mouse and human lung (19) (fig. S12, C and D). The transcriptional profile of gCaps were also more similar to ECs from other tissues, indicative of their 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 CYP1B1+ 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 \times 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

Fig. 2. Comparison of single-cell transcriptomics with conventional histology

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).

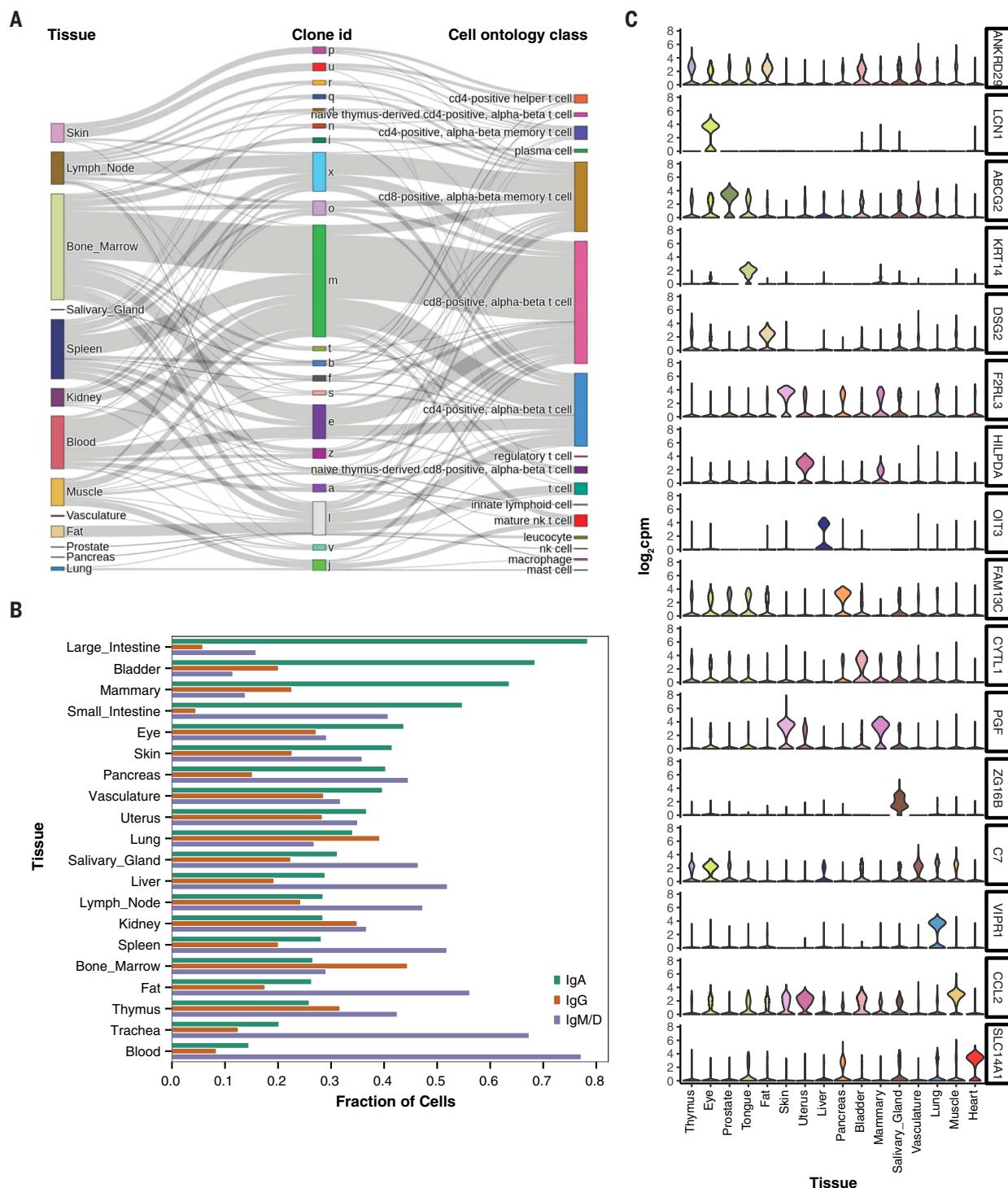


unidentified junctions. We additionally found 34,624 junctions between previously annotated 3' and 5' splice sites (3.6%). We identified 119,276 junctions between a previously annotated site and a previously unannotated site in the gene (12.4%). This leaves 584,030 putative junctions for which both splice 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 (27) 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 variants revealed by the Tabula Sapiens.

Hundreds of splice variants are used in a highly cell type-specific fashion; these can be explored in the cellxgene browser (14), which

**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.

uses a statistic called SpliceZ (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, par-

ticularly 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

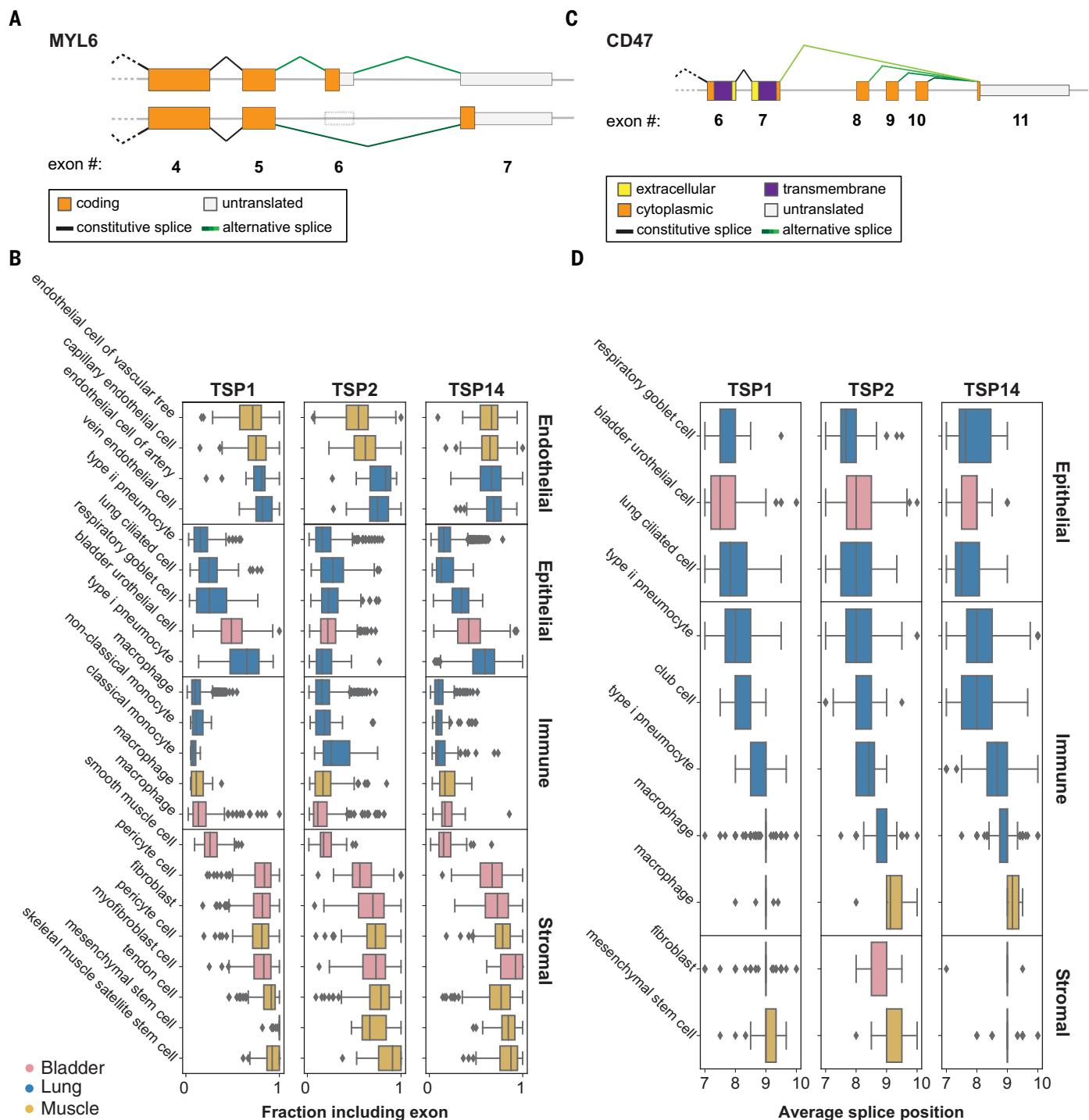


Fig. 4. Alternative splicing analysis. (A and B) The sixth exon in MYL6 is skipped at different proportions in different compartments. Cells in the immune and epithelial compartments tend to skip the exon, whereas cells in the endothelial and stromal compartments tend to include the exon. Boxes are grouped 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.

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 multisplicing 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 cytoskeleton-membrane interactions (29)—increases across the mesenchymal-to-myofibroblast transdifferentiation trajectory (fig. S16H). Another gene with a similar trajectory is MYLK (myosin light-chain kinase), which also rises as myofibroblasts attain more muscle-like properties (30). Finally, a random sampling of the most dynamic genes shared across TSP1 and TSP2 demonstrated that they share concordant trajectories and revealed some of the core genes in the transcriptional program underlying this 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 bloom 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 host-microbe 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

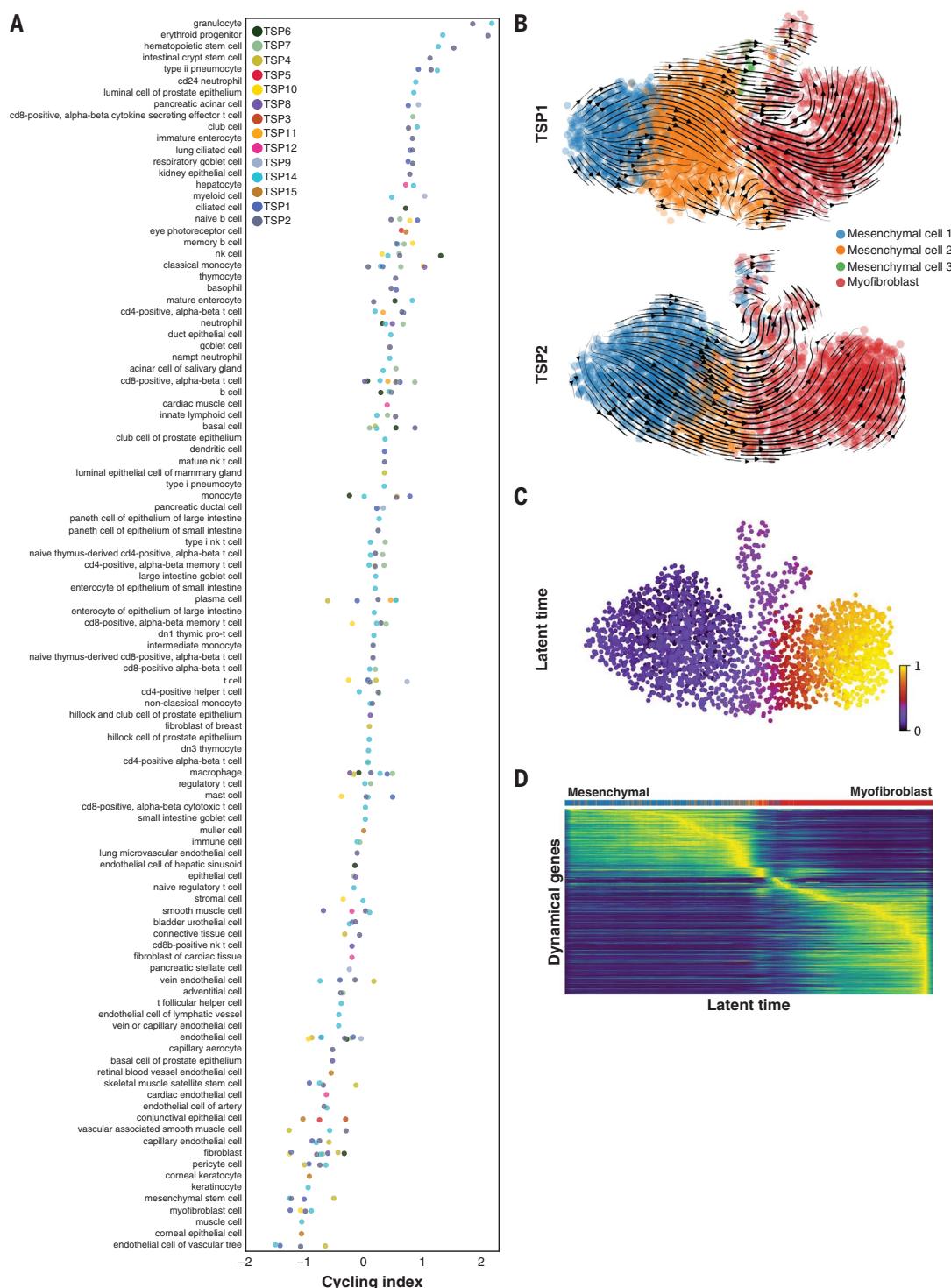


Fig. 5. Dynamic changes in cell state. (A) Cell types ordered by magnitude of cell cycling index per donor (each a separate color), with the most highly proliferative at the top and quiescent cells at the bottom of the list. (B) RNA velocity analysis demonstrating mesenchymal-to-myofibroblast transition 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.

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-cm³ 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 \times Genomics 3' V3.1 droplet-based sequencing and for FACS-sorted 384-well plate smart-seq2. Preparation began with

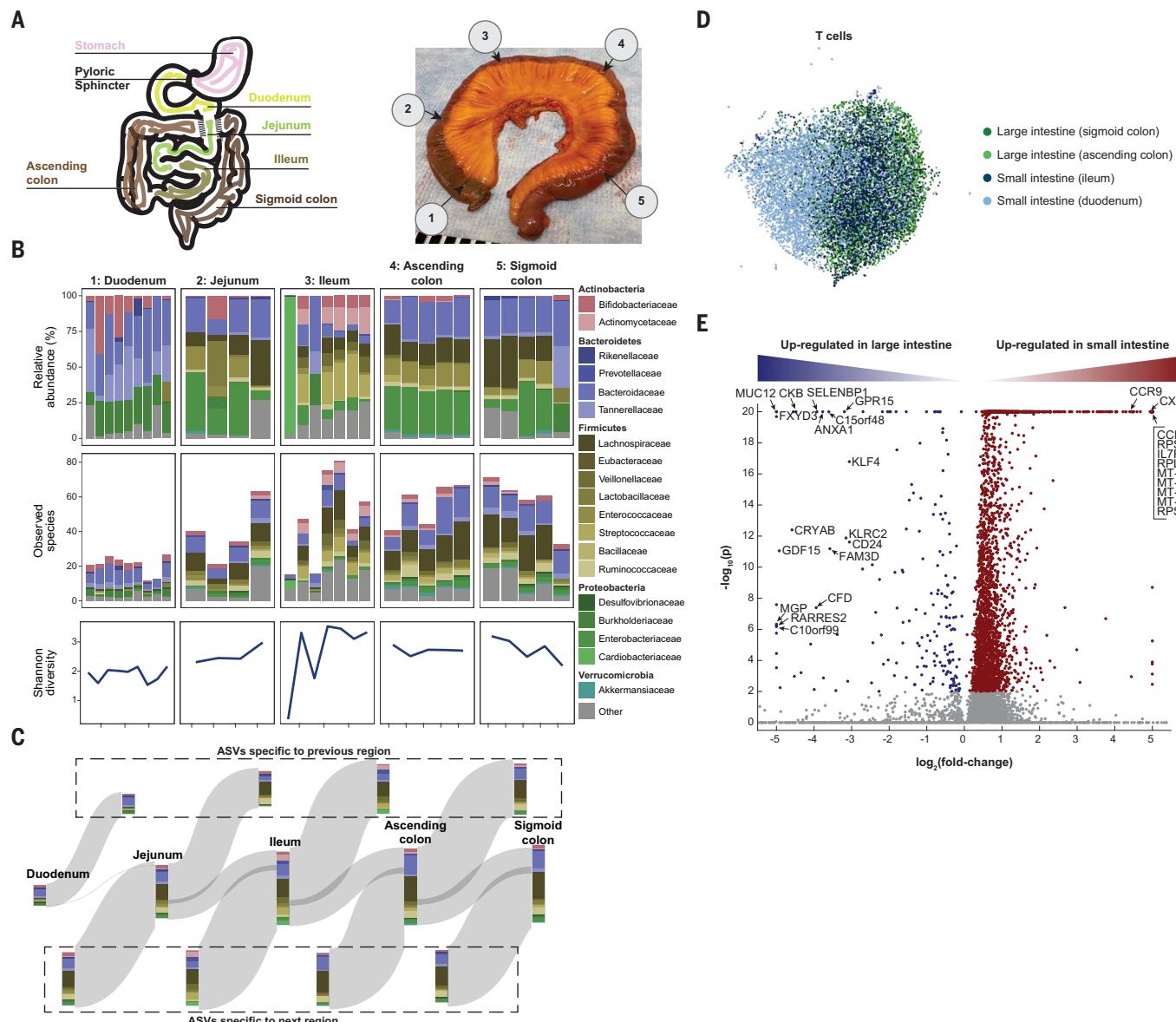


Fig. 6. High-resolution view highlights patchiness of the gut microbiome.

(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 16S 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.

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

Patricia Nguyen^{19,39,46}, Stephen R. Quake^{12,5}, Thomas A. Rando^{21,24}, Kristy Red-Horse²⁶, Jeremy Reiter⁵⁰, David A. Relman^{16,43,46}, Justin L. Sonnenburg¹⁶, Bruce Wang²⁰, Albert Wu³¹, Sean M. Wu^{19,39}, Tony Wyss-Coray^{21,24}

¹Department of Bioengineering, Stanford University, Stanford, CA, USA. ²Chan Zuckerberg Biohub, San Francisco, CA, USA.

³Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA. ⁴Howard Hughes Medical Institute, USA. ⁵Department of Applied Physics, Stanford University, Stanford, CA, USA. ⁶Department of Biomedical Data Science, Stanford University, Stanford, CA, USA. ⁷Center for Computational Biology, University of California Berkeley, Berkeley, CA, USA.

⁸Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA, USA. ⁹Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA. ¹⁰Donor Network West, San Ramon, CA, USA. ¹¹Department of Otolaryngology-Head and Neck Surgery, Stanford University School of Medicine, Stanford, California, USA. ¹²Department of Surgery, University of California San Francisco, San Francisco, CA, USA.

¹³Diabetes Center, University of California San Francisco, San Francisco, CA, USA. ¹⁴DCI Donor Services, Sacramento, CA, USA.

¹⁵Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. ¹⁶Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA.

¹⁷Department of Pediatrics, Division of Pulmonary Medicine, Stanford University, Stanford, CA, USA. ¹⁸Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA. ¹⁹Department of Medicine, Division of

Cardiovascular Medicine, Stanford University, Stanford, CA, USA.

²⁰Department of Medicine and Liver Center, University of California San Francisco, San Francisco, CA, USA. ²¹Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA. ²²Department of Surgery – Plastic and Reconstructive Surgery, Stanford University School of Medicine, Stanford, CA, USA. ²³Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA.

²⁴Paul F. Glenn Center for the Biology of Aging, Stanford University School of Medicine, Stanford, CA, USA. ²⁵Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA.

²⁶Department of Biology, Stanford University, Stanford, CA, USA.

²⁷Department of Pediatrics, Division of Cardiology, Stanford University School of Medicine, Stanford, CA, USA. ²⁸Stanford Diabetes Research Center, Stanford University School of Medicine, Stanford, CA, USA. ²⁹Center for Gynecology and Reproductive Sciences, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, San Francisco, CA, USA. ³⁰Department of Critical Care Medicine, The First Affiliated Hospital of Shantou University Medical College, Shantou, China.

³¹Department of Ophthalmology, Stanford University School of Medicine, Stanford, CA, USA. ³²Division of Nephrology, Department of Medicine, University of California San Francisco, San Francisco, CA, USA. ³³Stanford University School of Medicine, Stanford, CA, USA. ³⁴Mass Spectrometry Platform, Chan Zuckerberg Biohub, Stanford, CA, USA. ³⁵Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA. ³⁶Department of Medicine, Division of Hematology, Stanford University School of Medicine,

Stanford, CA, USA. ³⁷Department of Biochemistry and Biophysics, Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, USA. ³⁸Division 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. ³⁹Stanford Cardiovascular Institute, Stanford CA, USA. ⁴⁰Department of Chemical Engineering, Stanford University, Stanford, CA, USA.

⁴¹Department of Cell & Tissue Biology, University of California San Francisco, San Francisco, CA, USA. ⁴²Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA, USA. ⁴³Division of Infectious Diseases & Geographic Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. ⁴⁴Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA, USA.

⁴⁵Division of Nephrology, Stanford University School of Medicine, Stanford, CA, USA. ⁴⁶Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA. ⁴⁷Vera Moulton Wall Center for Pulmonary and Vascular Disease, Stanford University School of Medicine, Stanford, CA, USA. ⁴⁸Department of Urology, Stanford University School of Medicine, Stanford, CA, USA. ⁴⁹Division of Hematology/Oncology, Department of Medicine, San Francisco Veterans Affairs Health Care System, San Francisco, CA, USA.

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

Submitted 19 July 2021; accepted 31 March 2022
10.1126/science.abl4896

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

The Tabula Sapiens Consortium*, Robert C. Jones, Jim Karkanias, Mark A. Krasnow, Angela Oliveira Pisco, Stephen R. Quake, Julia Salzman, Nir Yosef, Bryan Bulthaup, Phillip Brown, William Harper, Marisa Hemenez, Ravikumar Ponnusamy, Ahmad Salehi, Bhavani A. Sanagavarapu, Eileen Spallino, Ksenia A. Aaron, Waldo Concepcion, James M. Gardner, Burnett Kelly, Nikole Neidlinger, Zifa Wang, Sheela Crasta, Saroja Kolluru, Maurizio Morri, Angela Oliveira Pisco, Serena Y. Tan, Kyle J. Travaglini, Chenling Xu, Marcela Alcántara-Hernández, Nicole Almanzar, Jane Antony, Benjamin Beyersdorf, Deviana Burhan, Kruti Calcuttawala, Matthew M. Carter, Charles K. F. Chan, Charles A. Chang, Stephen Chang, Alex Colville, Sheela Crasta, Rebecca N. Culver, Ivana Cvijović, Gaetano D'Amato, Camille Ezran, Francisco X. Galdos, Astrid Gillich, William R. Goodyer, Yan Hang, Alyssa Hayashi, Sahar Houshdaran, Xianxi Huang, Juan C. Irwin, SoRi Jang, Julia Vallve Juanico, Aaron M. Kershner, Soochi Kim, Bernhard Kiss, Saroja Kolluru, William Kong, Maya E. Kumar, Angera H. Kuo, Rebecca Leylek, Baoxiang Li, Gabriel B. Loeb, Wan-Jin Lu, Sruthi Mantri, Maxim Markovic, Patrick L. McAlpine, Antoine de Morree, Maurizio Morri, Karim Mrouj, Shravani Mukherjee, Tyler Muser, Patrick Neuhöfer, Thi D. Nguyen, Kimberly Perez, Ragini Phansalkar, Angela Oliveira Pisco, Nazan Puluka, Zhen Qi, Poorvi Rao, Hayley Raquer-McKay, Nicholas Schaum, Bronwyn Scott, Bobak Seddighzadeh, Joe Segal, Sushmita Sen, Shaheen Sikandar, Sean P. Spencer, Lea C. Steffes, Varun R. Subramaniam, Aditi Swarup, Michael Swift, Kyle J. Travaglini, Will Van Treuren, Emily Trimm, Stefan Veizades, Sivakamasundari Vijayakumar, Kim Chi Vo, Sevahn K. Vorperian, Wanxin Wang, Hannah N. W. Weinstein, Julianne Winkler, Timothy T. H. Wu, Jamie Xie, Andrea R. Yung, Yue Zhang, Angela M. Detweiler, Honey Mekonen, Norma F. Neff, Rene V. Sit, Michelle Tan, Jia Yan, Gregory R. Bean, Vivek Charu, Erna Forgó, Brock A. Martin, Michael G. Ozawa, Oscar Silva, Serena Y. Tan, Angus Toland, Venkata N. P. Vemuri, Shaked Afik, Kyle Awayan, Olga Borisovna Botvinnik, Ashley Byrne, Michelle Chen, Roozbeh Dehghannasiri, Angela M. Detweiler, Adam Gayoso, Alejandro A. Granados, Qiqing Li, Gita Mahmoudabadi, Aaron McGeever, Antoine de Morree, Julia Eve Olivieri, Madeline Park, Angela Oliveira Pisco, Neha Ravikumar, Julia Salzman, Geoff Stanley, Michael Swift, Michelle Tan, Weilun Tan, Alexander J. Tarashansky, Rohan Vanheusden, Sevahn K. Vorperian, Peter Wang, Sheng Wang, Galen Xing, Chenling Xu, Nir Yosef, Marcela Alcántara-Hernández, Jane Antony, Charles K. F. Chan, Charles A. Chang, Alex Colville, Sheela Crasta, Rebecca Culver, Les Dethlefsen, Camille Ezran, Astrid Gillich, Yan Hang, Po-Yi Ho, Juan C. Irwin, SoRi Jang, Aaron M. Kershner, William Kong, Maya E. Kumar, Angera H. Kuo, Rebecca Leylek, Shixuan Liu, Gabriel B. Loeb, Wan-Jin Lu, Jonathan S. Maltzman, Ross J. Metzger, Antoine de Morree, Patrick Neuhöfer, Kimberly Perez, Ragini Phansalkar, Zhen Qi, Poorvi Rao, Hayley Raquer-McKay, Koki Sasagawa, Bronwyn Scott, Rahul Sinha, Hanbing Song, Sean P. Spencer, Aditi Swarup, Michael Swift, Kyle J. Travaglini, Emily Trimm, Stefan Veizades, Sivakamasundari Vijayakumar, Bruce Wang, Wanxin Wang, Julianne Winkler, Jamie Xie, Andrea R. Yung, Steven E. Artandi, Philip A. Beachy, Michael F. Clarke, Linda C. Giudice, Franklin W. Huang, Kerwyn Casey Huang, Juliana Idoyaga, Seung K. Kim, Mark Krasnow, Christin S. Kuo, Patricia Nguyen, Stephen R. Quake, Thomas A. Rando, Kristy Red-Horse, Jeremy Reiter, David A. Relman, Justin L. Sonnenburg, Bruce Wang, Albert Wu, Sean M. Wu, and Tony Wyss-Coray

Science 376 (6594), eabl4896. DOI: 10.1126/science.abl4896

Mapping cell types in the human body

The Tabula Sapiens is a molecular reference atlas for more than 400 cell types of the human body. The Tabula Sapiens Consortium used single-cell transcriptomics to measure the messenger RNA molecules in each of nearly 500,000 cells from 24 tissues and organs (see the Perspective by Liu and Zhang). These data enable new insights into how the human genome parts list is used to create distinct cell types within the human organism. In addition to creating a detailed molecular definition of these cell types, the atlas reveals many other aspects of human biology, including how the same gene can be spliced differently in different cell types, how shared cell types in different tissues can have subtle differences in their identities, and how clones of the immune system can be shared across tissues. —LZ and DJ

View the article online

Use of this article is subject to the [Terms of service](#)

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

<https://www.science.org/doi/10.1126/science.abl4896>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works