Tabula Sapiens reveals transcription factor expression, senescence effects, and sex-specific features in cell types from 28 human organs and tissues

Authors

The Tabula Sapiens Consortium

Abstract

The Tabula Sapiens is a reference human cell atlas containing single cell transcriptomic data from more than two dozen organs and tissues. Here we report Tabula Sapiens 2.0 which includes data from nine new donors, doubles the number of cells in Tabula Sapiens, and adds four new tissues. This new data includes four donors with multiple organs contributed, thus providing a unique data set in which genetic background, age, and epigenetic effects are controlled for. We analyzed the combined Tabula Sapiens data for expression of transcription factors, thereby providing putative cell type specificity for nearly every human transcription factor and as well as new insights into their regulatory roles. We analyzed the molecular phenotypes of senescent cells across the entire data set, providing new insight into both the universal attributes of senescence as well as those aspects of human senescence that are specific to particular organs and cell-types. Similarly, we analyzed sex-specific gene expression across all of the identified cell types and discovered which cell types and genes have the most distinct sex based gene expression profiles. Finally, to enable accessible analysis of the voluminous medical records of Tabula Sapiens donors, we created a web application powered by a large language model that allows users to ask general questions about the health history of the donors.

Introduction

Single cell transcriptomic atlases are providing substantial new insights into human biology, including the molecular definitions of cell types, the cell type specific expression of disease-related genes, and the relationships of shared cell types across tissues, particularly those of the immune system[1–6]. These atlases are powerful reference tools whose applications in biology are just beginning to be explored. Many questions which could previously only be addressed at the bulk tissue level can now be answered with cell type specificity, thus yielding crucial new insights[6–8].

Our previous efforts focused on creating a reference human cell atlas, called the Tabula Sapiens 1.0, which consisted of data from 24 different organs and tissues from a diverse set of donors[1]. In Tabula Sapiens 2.0, we have now doubled both the number of cells and the number of donors with multi-organ contributions, thus increasing the number of tissues and donors for which genetic background, age and epigenetic effects are controlled for. Specifically, we report data from nine new donors comprising one donor with 20 tissues, one donor with 18 tissues, one donor with 15 tissues, one donor with 2 tissues, and five donors with 1 tissue each. The updated combined Tabula Sapiens dataset contains more than 1.1 million cells from bladder, blood, bone marrow, eye, ear, fat, heart, kidney, large intestine, liver, lung, lymph node, mammary, muscle, ovary, pancreas, prostate, salivary gland, skin, small intestine, spleen, stomach, testis, thymus, tongue, trachea, uterus, vasculature. Single cell transcriptomes were obtained from live cells using both droplet microfluidic emulsions (1,093,048 cells) as well as FACS into well plates (43,287 cells).

To illustrate the utility of this reference dataset, we conducted several genome-wide expression studies. Our investigation mapped which transcription factors are active in specific cell populations, revealing cell-type associations for over a thousand regulatory proteins and uncovering previously unknown aspects of their functional contributions. Additionally, we examined aging-related cellular characteristics throughout our complete collection, which revealed which common gene expression modules used by all senescent cells alongside organ- and cell-specific aging patterns. We further explored how gene activity differs between males and females across every cell population in our study. This analysis revealed that biological sex influences gene expression primarily through cell-level regulatory mechanisms rather than through tissue-level differences. The data set is publicly available and we have also created an AI based tool to explore the clinical histories of the donors.

Results

Overview of Tabula Sapiens 2.0

The Tabula Sapiens 2.0 is an integrated map of 28 tissues collected across 24 donors. Nine new donors and four new tissues were collected and analyzed together with the original Tabula Sapiens 1.0 dataset (**Fig. 1, Supp. Fig. 1 -Supp. Fig. 6**). All tissues and organs, with the exception of the respective reproductive organs, were profiled from both male (n=11) and female (n=13) donors. The donor's age ranges from 22 to 74 years old, offering one of the most comprehensive molecular profiles of human tissues across the adult lifespan (7 donors under the age of 40, 11 donors between 40 and 60, and 6 donors over 60 years of age). A detailed table with metadata can be found in **Supp. Table 1**.

For each organ and tissue, a group of experts used a defined cell ontology terminology to annotate cell types consistently across the different tissues, leading to a total of 701 distinct cell populations representing cell type combined with tissue of origin designation. The full dataset can be explored online using CELLxGENE[9,10].

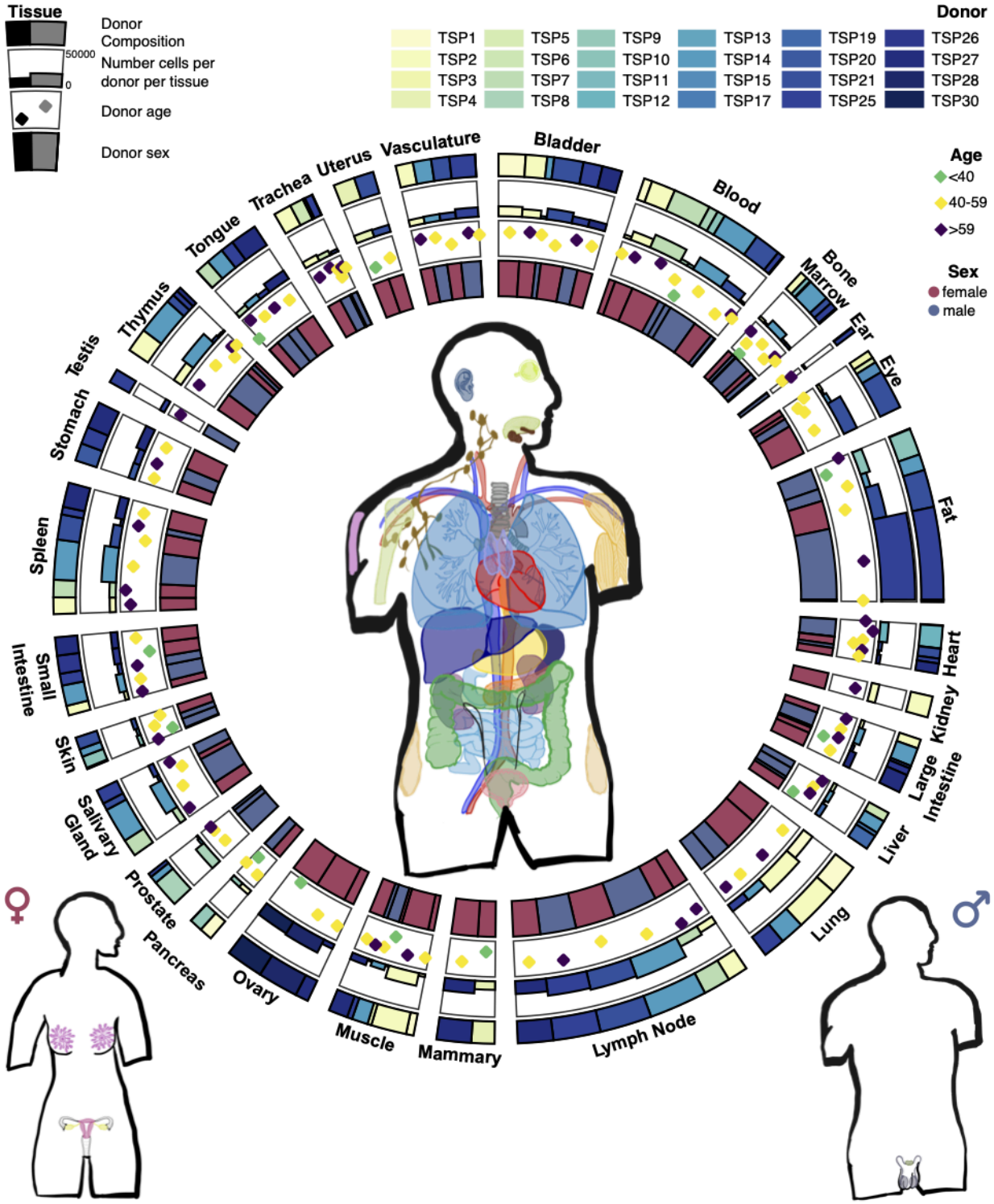


Figure 1. Overview of tissue donor demographics and tissue composition in Tabula Sapiens 2.0. The central anatomical illustration highlights the location of various tissues sampled from male and female donors, organized into distinct anatomical systems. Surrounding the anatomical figures are circular arrangements of charts showing the donor demographics of samples across various tissues. The inner stacked bar chart color-coded by donor sex for samples across every tissue (blue: male, pink: female). The diamond plot shows the donor age for samples across every tissue (green diamond: under 40, yellow diamond: 40-59, purple diamond: over 59). The bar plot shows the number of cells per donor for samples across every tissue with tissue color coded by donor ID. The outer stacked bar plot shows the donor composition for samples across every tissue. Tissues are labeled around the circular chart, providing an accessible visual summary of the tissue distribution and donor demographics.

Transcription factor analysis across cell types

A fundamental aspect of cell state and cell identity is the regulation of gene expression by transcription factors. Several large-scale consortia have systematically mapped human transcription factor (TF) activity at the tissue level. The ENCODE Project[11] established standardized pipelines for genome-wide TF binding site identification using ChIP-seq and DNase-seq footprinting, generating high-resolution maps of regulatory DNA occupancy. The Roadmap Epigenomics Consortium[12] expanded this work by creating reference epigenomic maps for 127 cell lines and tissues, integrating histone modifications, DNA methylation, and chromatin accessibility to contextualize TF function. The GTEx Consortium[13] developed statistical approaches to infer individual-specific TF activities from RNA-seq data across 49 tissues, identifying genetic variants (QTLs) that modulate TF regulatory effects. Complementing these efforts, the FANTOM Consortium used deepCAGE technology to profile transcription initiation events and reconstruct TF interaction networks governing cellular differentiation. While these foundational studies focused on bulk tissue methods, more recent work has used TF overexpression in single cells to reveal insights into TF-driven human developmental trajectories[14].

Of the 1639 putative transcription factors (TF) in The Human Transcription Factors database[15] (HTF) we investigated the 1637 TFs which appear in GRCh38 gene annotations. We computed the mean of the log-normalized expression of each transcription factor across the 175 cell types detected in our droplet microfluidic dataset. All but two TF's are expressed in at least one cell type in this data set. SHOX and ZBED1, both from the pseudoautosomal region 1 of the X and Y chromosomes, have zero raw droplet counts in every donor. The Human Protein Atlas[16–18] (HPA) reports RNA and protein detection of SHOX in a broad range of tissues with enrichment in basophils, as well as expression in fibroblast cell lines. ZBED1 is found in a small number of cells across all cell types in our plate data and the HPA also reports ubiquitous cell type expression. We therefore have been able to establish the cell type specificity of all known human transcription factors.

In order to test whether gene expression of these transcription factors was also accompanied by activity, we used the SCENIC software package to infer TF activity in single cells of each cell type separately. SCENIC identifies TF's target genes through co-expression-based GRN inference algorithms and presence of binding motifs in proximity to target genes. Activity of these TF regulons is then scored by enrichment of the identified target gene set in each cell. Of 1390 TFs that met SCENIC's entry criteria we found 839 regulons in at least one cell type. The remaining 551 did not survive SCENIC's filters at various stages; this does not mean they didn't have significant co-expression with others genes but that

their co-expression was below an internal threshold. Across the 38 broad cell classes, SCENIC identified varying numbers of active TF regulons (mean=80.3, sd=59.6). Individual regulons were found in up to 31 cell types with most regulons, however, showing high cell type specificity, being found on average in 3.6 cell types (sd=4.0) **(Supp. Fig. 7)**. Overall this provides evidence of activity by downstream gene expression for roughly half of the human TF's and does not rule out activity for the remainder.

To establish transcription factor cell type and tissue specificity, we computed the specificity statistic τ [19,20] across all donors **(Methods, Supp. Table 2)**. When computed on 5 individual samples we found r = 0.76 to 0.81 sample-to-sample correlation (**Supp. Fig. 8**). This statistic was developed for microarray and bulk tissue RNA-seq studies, and has recently been used to determine cell-type specificity in single-cell mRNA-seq data studies such as the *Drosophila* cell atlas[21] and a mouse brain atlas[22]. As seen in bulk mRNA-Seq studies[13], we found the τ distribution dips at $\tau \cong 0.85$. This same 0.85 dip occurs in the *Tabula Muris Senis* and *Drosophila* single cell atlas datasets; and also when τ is computed on a tissue basis or on a broad cell type basis in our data. Thus we defined the 890 TF's above this as cell type specific **(Fig. 2A)**. We defined the 745 TF's with $\tau < 0.85$ as non-specific and used the 100 lowest τ TF's to explore the space of ubiquitous TF's. To visualize TF expression we placed the TFs in τ order and grouped expression values by 38 broad cell class categories to create an expression heatmap (**Fig. 2B)**.

Among the 100 TF's labeled as ubiquitous, we found many well studied factors known to be central to universal cell function such as stress response, proliferation and metabolism. Examples of these include *NFAT5*, *NCOA1*, *FOXJ3*, *FOXK2*, *ATF4*, *JUN*, *FOS* and *STAT1*. To validate the existence and roles of the ubiquitous TF's, we searched the HPA for evidence of protein expression. Of the 100 ubiquitous TF's, 72 were tested in the HPA; every single one of these was detected in at least one tissue and the majority (71%) were detected in 19 or 20 of the 20 tissues tested. (**Fig 2C, Supp Fig. 9A)**. In the 32 cell types, all 72 low τ TF's tested were detected in many cell types. The fraction detected in more than 80% of cell types was 0.69, vs 0.40 for all 699 TF's. **(Supp Fig. 9C-D).** To further confirm the likelihood these TF's are active, we queried the HPA for the subcellular location of expression of the 72 ubiquitous TF's; the HPA detected expression in the nucleus for 69 of these across a broad set of tissues Altogether there is evidence of protein expression for 100% of the ubiquitous TFs found in the HPA and evidence of activity based on nuclear localization for 96% of them (**Fig 2D**).

We explored the biological function of the full set of 745 non-specific TF's ($\tau < 0.85$) by performing Gene Set Enrichment Analysis (**Fig. 2E, Methods, Supp. Table 3)** and grouping the GO terms into four broad cellular function categories: Gene Expression and Regulation, Tissue Maintenance, Cell Response to Stimulus, and Cell Metabolism. Within the Gene Expression and Regulation category, we found 25 Gene Ontology (GO) terms and sub-grouped these into gene transcription, gene regulation, gene expression of miRNA and chromatin remodeling as needed for the cell to synthesize mRNA and miRNA. Gene regulation comprises 11 GO terms and dominates the number genes driving these terms, 672, by an order of magnitude.

Transcription factors for maintenance of tissues are enriched in our data in 18 GO terms. We grouped these into cell cycle regulation, cell differentiation, cell proliferation and anatomical structure maintenance. Cell cycle regulation enriched for 33 genes in three GO terms. These genes include *TFDP1* and *TFDP2* which form complexes with E2F transcription factors to control cell cycle[23,24] and *E2F6* which specifically blocks cell cycle entry[25]. The cell differentiation group includes the two broad terms

negative regulation of cell differentiation (GO:0045596) and positive regulation of cell differentiation (GO:0045597), as well as terms identified as specific for a variety of cell types. There are 61 transcription factors enriched in the cell differentiation category. Of these, *ARNT*, *ARNTL*, *CEBPB*, *CREB1*, *CREBL2*, *ETS1*, *FOXO1*, *FOXO3*, *HIF1A*, *PPARD*, *SMAD3*, *STAT1*, *STAT3* ,*STAT5B*, *XBP1*, *ZBTB16*, *ZFPM1* and *ZNF16* are enriched in three or more terms. Cell proliferation has four GO term a): regulation of transforming growth factor beta2 production (GO:0032909), b) hemopoiesis (GO:0030097), c) regulation of cell population proliferation (GO:0042127), and d) positive regulation of myoblast proliferation (GO:2000288). These terms have enriched genes, including the 3 ATP-1 binding factors *JUN*, *JUNB*, and *JUND*; 6 STAT family transcription factors (which respond to growth factor receptors) and *KLF4*, *KLF10*, *KLF11* (which can negatively regulate cell growth). The anatomical structure maintenance group contains two GO terms with 17 enriched transcription factors. These include all 3 PBX TF's of the TALE family that interact with the HOX domain[26]; and 3 of the 4 TEAD factors known in humans to regulate tissue size and function by coupling external signals to control cell growth and specification[27].

Response to a cell's environmental stimuli is found in five groups of GO terms: chemical signaling, cell signaling, immune signaling, hormone response, and general stress response. Cell response to stress encompasses responses to fluid stress, laminar flow stress, oxidative stress, hypoxia, and chemical stress. Note, some transcription factors may be induced beyond natural levels due to processing tissue into live single-cells for analysis. Among 16 stress response genes, *ATF3*, *ATF4*, *ATF6*, *CEBPB*, *DDIT3*, *HIF1A*, *KLF2*, *KLF4*, *NFE2L2*, *RBPJ*, and *TP53* are enriched in association with two or more GO terms. *KLF2* and *KLF4* are suggested to be part of the Golgi stress response[28]. Also enriched is *SREBF2*, which regulates cholesterol synthesis in response to a broad set of stresses including inflammation, heat shock and hypoxia[29]. Oxidative stress response is thought to be regulated by enriched *NFE2L2* (NRF2)[30] through the KEAP1-NRF2-ARE pathways[31], as well as by *HIF1A*[32]. GO terms for cell signaling fall into three groups, cell signaling, IL6 and IL9 immune signaling, and peptide hormone signaling. They share the common STAT signaling pathway whereas cell signaling selectively utilizes NFAT TF's and hormone signaling utilizes NCOA TF's[33].

Within the 10 GO terms enriched for cell metabolism four involve regulation of macromolecule synthesis, five terms regulate lipid synthesis, and one maintains glucose homeostasis. The four macromolecule synthesis terms include 217 TF's, more TF's than any set of terms other than gene regulation.

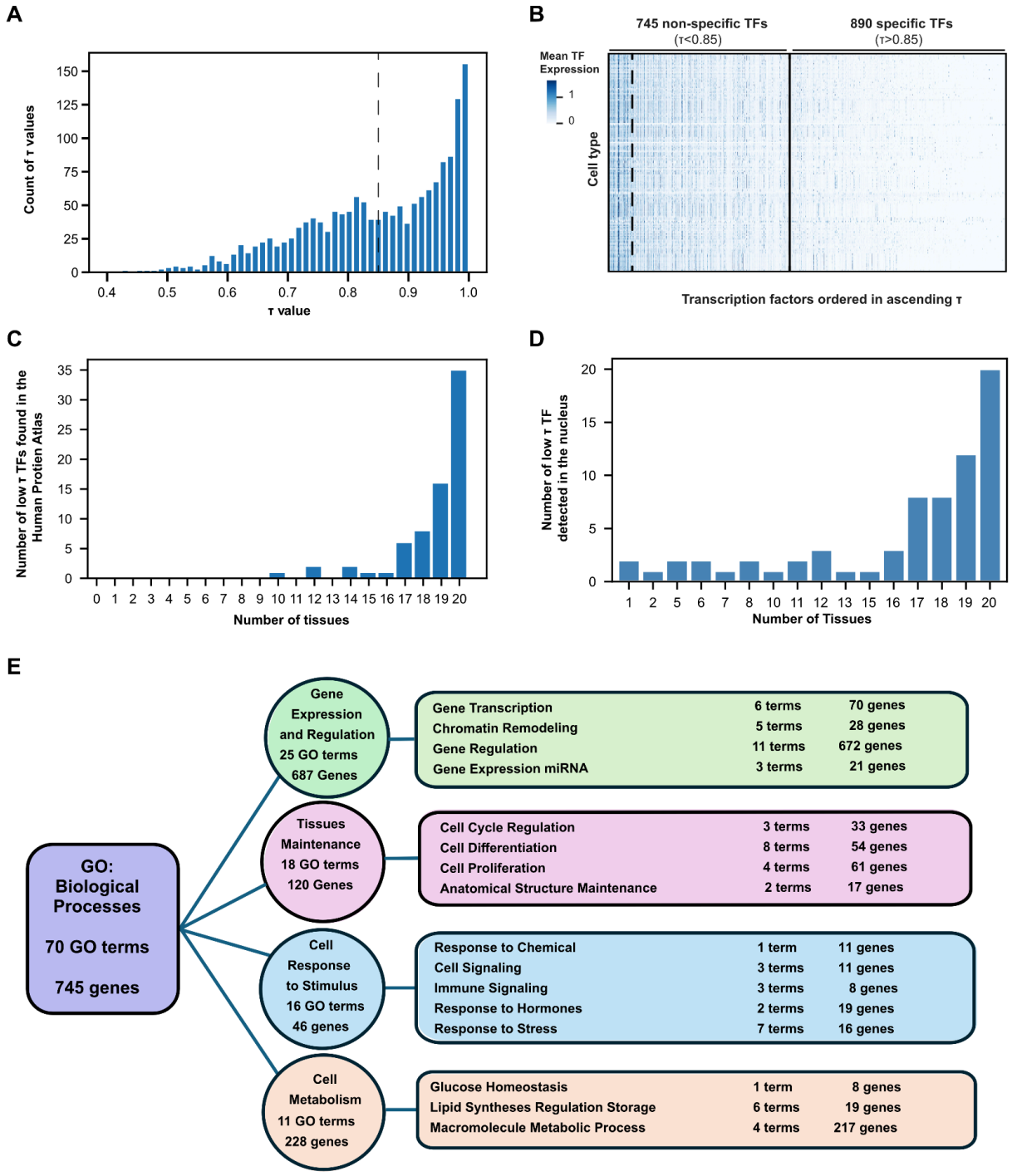


Figure 2: Transcription factor expression common across cell types. A) Histogram of the distribution of τ values for 1635 transcription factor genes computed across 175 distinct cell types in 28 Tabula Sapiens 2.0 tissues, marked with specificity threshold at τ > 0.85. B) Heatmap of the mean of log normalized expression in each cell type of 1635 TF's in the 175 cell types. The mean expression values range from 0.0 to 4.9 but for better visualization in the heatmap they have been clipped to the 99.0 percentile resulting in the heatmap scale from 0 to about 1.0. Cell types are in alphabetical order by broad

cell type. Transcription factors are arranged in τ value order from from 0.4 to 1.0 τ. The vertical bars on the heatmap mark the lowest 100 τ ubiquitous TF's, 745 τ < 0.85 non-specific TF's and the 890 τ > 0.85 cell type specific TF's. C) Distribution of the 72 ubiquitous sc-mRNA expressed TF's detected as expressed proteins by immuno-staining in 20 common tissues in the Human Protein Atlas. D) Distribution of nuclear expression of 69 ubiquitous TF's across 20 tissues. E) Summary of the Enricher GO analysis of the 745 non-cell type specific transcription factors outlined in black in A. Enrichr returned 70 GO terms that were organized by broad cell function categories shown in the circles with the number of GO terms and genes in each category. Sub-categories are shown in the rectangles with the number of GO terms and genes in each.

Within the non-specific TFs we observed many widely expressed zinc finger DNA binding family TF's. Of these, the Human Transcription Factor Database shows many are only computationally predicted, or have no defined motif, and lack experimental validation. Often the literature only mentions differential expression in the context of disease, with no known function. It is striking that so many broadly expressed transcription factors exist and are not well characterized, including many that are ubiquitous across cell types.

We then examined transcription factor expression levels for the 890 cell type specific TF's (τ>0.85). We discovered both previously known as well as novel relationships between specific TF's and the cell types in which they are expressed (**Fig. 3A)**. Each TF was grouped with the cell type of its highest mean expression and plotted in broad cell class alphabetical order. Vertical lines show the ubiquitous TF's are redistributed throughout cell types, while the horizontal bars reveal highly expressed TFs in each cell type or in a broad cell class. This provides putative cell type association for the specific TF's, many of which previously have only been identified within the context of bulk tissue expression.

Within the T-cell broad cell class we identify five TF's with τ > 0.85 *FOXP3*, *EOMES*, and three Ikaros TF's (*IKZF1*, *IKZF2*, and *IKZF3*), suggesting they are specific to T-cells (**Fig. 3B**). *FOXP3* expression is the defining factor for *FOXP3* Treg cells[34]. In these cells it coordinates with *ETS1*, which we see expressed across all T-cell types[35]. *EOMES* highest expression is in Mature NK T cells and CD8 positive alpha beta T cells[36]. The three Ikaros family TF's are well known to regulate all types of lymphocyte cell development[37,38]. We find them expressed throughout T-cell types.

Gene expression profiles in fibroblast cells are heterogeneous and often more organ or environment specific than cell type specific[39–42]. Thus, over 80% of cells in this broad category have been typed only generally as 'fibroblast cell' (**Fig. 3C**). We do find three fibroblast/organ specific TF's where we have specific fibroblast cell types. In the lung, *TCF21* has highest expression in alveolar type-2 fibroblast cells where it regulates lipofibroblast differentiation and serves as a marker for these cells[43], *TBX20* plays a key role in development and maintenance of healthy cardiac tissue[44] and *SREBF1* expression in thymic fibroblasts is known to be required for activated T-cell expansion[45].

Our data show nine stem cell type specific TF's involved in the generation of fat, bone, blood, and gut cells (**Fig. 3D**). *ZIC1* and *ZIC4* have their highest expression in mesenchymal stem cells of adipose tissue, where *ZIC1* is known to direct stem cell fate between osteogenesis and adipogenesis[46]. We also see *ZIC1* and *MYF5* expressed in skeletal muscle satellite stem cells where they regulate myogenesis[47]. *ZIC1* is suggested to be involved in brown fat precursor differentiation or in white-to-brown transdifferentiation[48].

Gut stem cells are known to be maintained by *CDX2* and *CDX1*[49], and these two regulatory TF's have $\tau >$ 0.85 in our data, with their highest expression found in transit amplifying cells and intestinal crypt stem cells in both the small intestine and colon. Specific to small intestine crypt stem cells we find expression of *ASCL2*, a master controller of intestinal crypt stem cell fate[50]. *ONECUT2*, a downstream target of *ASCL2*, has its highest expression in transit amplifying cells and crypt stem cells. However, *ONECUT2* expression is not restricted to the broad stem cell type in our data as it has similar expression levels in all the small intestine enterocyte cell types. *ONECUT2* is also suggested to regulate differentiation of M cells of Peyer's patches[51]. Three TF's with $\tau > 0.85$ have their highest expression in hematopoietic stem cell types. Consistent with the literature, *MYB* is found in decreasing amounts in hematopoietic stem cells, then myeloid progenitors and then erythroid progenitors. It also shows some expression in the intestines[52]. It is worth noting that genes most famous for their roles in early development, such as the Homeobox family and the Yamanaka factors, are also known to be expressed in adult cell types and are present in our data. For example *HOXA9* and *CBX2*, necessary for embryonic development and hematopoiesis[53–55], show the decreasing expression from hematopoietic stem cell type to progenitor cell types. We also see TF's *KLF1* and *GATA1* increasing expression in sequence towards development of erythroid progenitor cells[56,57].

The broad endothelial cell class heatmap (**Fig. 3E)** reveals 18 TF's with $> 0.85\ \tau$. Of these, *HEY1*, shows higher expression in the arterial vascular cell types where it may mediate arterial fate decisions[58]. Similarly, *LHX6* has been extensively studied in neuron development[59], but in our data is expressed in vein and venules endothelial cells suggesting a regulatory role in venous fate. *FOXC2* is expressed highest in lymphatic cells where it is known to regulate formation of the lymphatic system[60]. Together these TF's appear to regulate the fate of all the three types of vascular structures, arterial, venous and lymphatic. Finally, we see *ERG* expression known to stabilize vascular growth through regulation of VE-caderhin[61] and known to maintain vascular structure by repressing proinflammatory genes[62]. We see several less cell type specific TF's with high expression which are also known to regulate stability of the vascular tree. *FLI1* and *ERG* prevent endothelial to mesenchymal transition[63]. *ELK3* suppresses further angiogenesis[64]. *EPAS1* (HIF2A) regulates physiological response to oxygen levels. Interestingly, some *EPAS1* alleles contribute to high altitude athletic performance[65,66].

Finally, we highlight male germ cells which have 204 TF's with their highest expression in the four cell types in this broad cell class (**Fig. 3F**). In addition, 50% of these are $> 0.85\ \tau$ testis specific, consistent with findings in *Drosophila*[21]. Over 50% of these specific TF's begin with the letter 'Z' indicating a zinc finger motif. This appears to be a rich area for further study as most of these TF's are computationally derived and/or understudied, especially in humans, as most work has been done in mice. Spermatogenesis is driven by specific TF's through four stages, going from male germ cell to spermatogonia via mitosis, then to spermatocyte and spermatid through meiosis, to become sperm[67]. As is known, our data shows specific expression of *SALL4* in male germ cells where it maintains this pool of cells[68,69]. *DMRT1* is known to specifically coordinate male germ cell differentiation and our data shows its expression growing from male germ cells to the spermatogonia phase[70,71]. Spermatogonia proliferation is then driven by expression of *TCF3*[72]. *SOHLH1* is regulated by *DMRT1* and known to be active in the later stages of spermatogonia mitotic division[71]. Although in our data, *SOHLH1* has its highest expression in oocytes, we see it decreasing as male germ cells progress to spermatogonia (**Supp. Table 4).** Spermatogonia then execute major transcriptional changes to become primary spermatocytes where they progress through two rounds of meiotic division to become spermatids. This requires a complex set of transcriptional changes,

which are known in mice to be coordinated by *TCFL5* and *MYBL1*[73]. In our data *TCFL5* has its highest expression in spermatocytes. *MYBL1* has a reasonably high τ value of 0.95 but it has higher mean expression in several immune cell types. Outside of immune cells, its highest expression is in spermatocytes (**Supp. Table 4**). Finally, transition from spermatocyte to early spermatid is known to be initiated by expression of specific TF *SOX30*[73,74]. In our data *SOX30* expression is highest in spermatocytes and also high in spermatids.

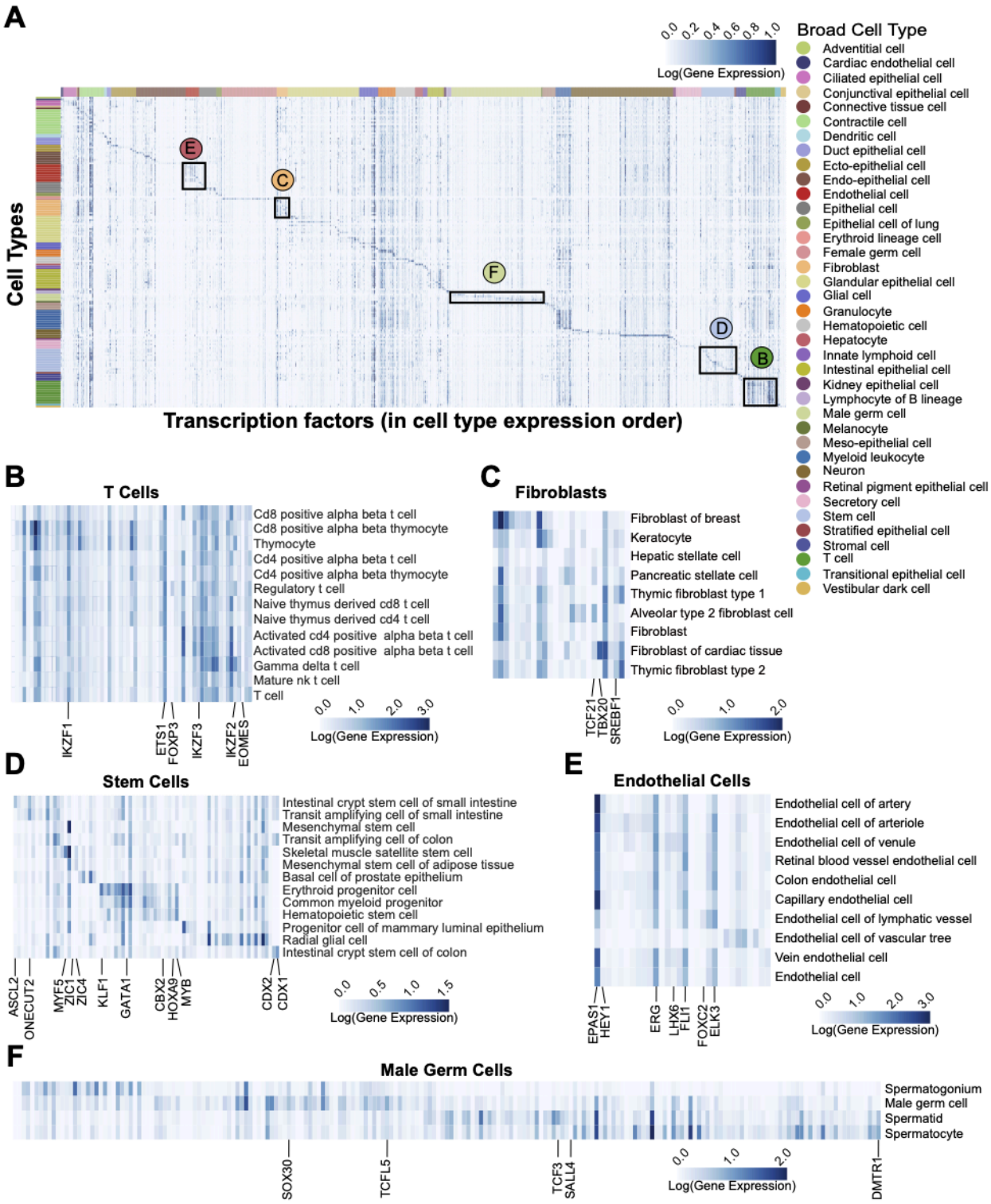


Figure 3: Transcription Factors Specific to Cell Types. A) Expression heatmap showing the mean of the log normalized expression values of 1635 human transcription factors in 175 cell types using all 28 Tabula Sapiens 2.0 tissues. For better visualization in the full heatmap (A) the max values are clipped at the 99.5 percentile. The cell type rows are organized alphabetically by broad cell type and denoted by

color on the annotation bar. The TF columns are arranged according to the cell type for which they show the highest expression. Black outlines highlight examples of common broad cell types and refer to sub-figures B-F, which are not clipped at the 99.5 percentile in order to show the full range of expression. B) T-Cells. C) Fibroblasts. D) Stem Cells. E) Endothelial Cells. F) Male Germ Cells.

Molecular profiles of senescent cells across human tissues

Cellular senescence is a state of irreversible cell cycle arrest, which can be triggered by various cellular and environmental stresses[75,76]. At a molecular level, cyclin-dependent kinase inhibitors, particularly p16-INK4a (encoded by the gene *CDKN2A*), play a crucial role in controlling the initiation and maintenance of cellular senescence[77,78]. At a cellular level, senescence has been characterized by its unique features such as enlarged cell morphology, increased senescence-associated beta-galactosidase activity (SA-β-Gal), formation of senescence-associated heterochromatin foci (SAHF), and production and secretion of inflammatory cytokines, growth factors, and proteases, known as senescence-associated secretory phenotype (SASP)[79,80]. Nonetheless, the phenotype of cellular senescence is highly variable and heterogeneous, with mechanisms not universally conserved across all senescence programs[81–87]. The accumulation of senescent cells has been shown to affect various aspects of aging[88] and their clearance has been shown to delay various age-related degenerative pathologies[89,90].

One of the key challenges in studying senescence is its heterogeneity. Studies using senescent cell culture models have demonstrated this variability, showing that cells can exhibit diverse phenotypes depending on the cell type and the specific inducer of senescence[84,85]. For instance, SASP has been shown to vary based on cellular context[86,87]. This *in vitro* heterogeneity points to the possibility that senescent cells in living organisms are similarly diverse, with distinct cell phenotypes. Compounding the issue, many classical markers of senescence, which were established using cell culture models, have limited utility *in vivo*. To tackle this issue, researchers have developed guidelines such as the Minimum Information for Cellular Senescence Experimentation in Vivo (MICSE) and the SenNet guidelines for identifying senescent cells in human tissues. These provide recommendations for evaluating senescence markers directly within tissues and in live animal models[91].

Given the heterogeneity of senescence markers and the methodological challenges in detecting them, we defined senescent cells based on the expression of *CDKN2A* and the absence of *MKI67* (a cell proliferation marker). To deploy a multi-pronged approach for identification of senescent cells as suggested in SenNet guidelines, we performed a differential expression analysis for genes representing canonical hallmarks of senescence in senescent cells as compared to non-senescent cells across all broad cell types in Tabula Sapiens 2.0. Although our operational definition of senescence drew exclusively on the cell-cycle‑arrest hallmark (*CDKN2A*+ and *MKI67*-), the resulting differential-expression profile revealed that at least one hallmark-specific gene exhibited the expected direction of change for at least two additional hallmarks in most cell types (**Supp. Fig. 10A**). Thus, the *CDKN2A*+ *MKI67*- population not only satisfies the cell-cycle arrest criterion but also displays molecular signatures consistent with multiple complementary hallmarks, aligning with the framework proposed by SenNet.

To systematically characterize the phenotype of senescence cells of different cell types across various human tissues, *CDKN2A*+ *MKI67*- cells were identified across 25 tissues from 21 donors in the Tabula

Sapiens 2.0 dataset (**Methods**). We detected 48,114 senescent cells (~4.4% of all cells) spanning 145 cell types grouped into 34 broad cell type categories[92] (**Supp. Fig. 11A & 11B**). This represents the largest and most comprehensive study of senescent cells in human tissues to date, providing an unprecedented view of their distribution and diversity across multiple tissue types and cellular contexts. We detected a small increase in senescent cells from medium (40-60y) and old donors (>=60y) as compared to young donors (<40y) (**Fig. 4A**). As expected, we observed a heterogeneity in the burden of senescent cells across different tissues, with eye, bladder, and tongue containing the highest, and the heart, muscle, and ovary containing the lowest proportion of senescent cells (**Fig. 4B**). The low burden of senescent cells in the heart, quadriceps muscle, and gastrocnemius has been observed previously[93]. Comparing the proportion of senescent cells in non-reproductive organs between male and female donors, we observed a slightly higher senescence cell burden in the spleen and lungs from female donors as compared to male donors (**Fig. 4B**). Next, we looked at canonical markers associated with the cellular senescence phenotype in our dataset. In addition to *CDKN2A*, previously known senescence-associated DNA damage marker *H2AX*, SASP regulators such as *TGFB1* and *NFKB1*, and SASP markers such as *HMGB1* and *TIMP2* were also enriched in senescent cells as compared to non-senescent cells (**Supp. Fig. 11C**).

To understand the universality of senescence phenotype, we first identified genes that were upregulated in senescent cells across individual tissues and cell types, and used these genes to quantify the cell type and tissue prevalence of senescence-associated genes (SAGs) (**Methods, Supp. Fig. 11B**). To find SAGs for individual cell types, we selected genes significantly upregulated in senescence cells of that cell type from any tissue in at least 50% of the donors ($\log_2$ fold change > 0.5, adjusted p-value < 0.001, **Supp. Table 5**). We ranked the 3792 SAGs, which were upregulated across 30 broad cell types in 25 tissues based on their cell type prevalence (**Methods, Supp. Fig. 10B**). Interestingly, after *CDKN2A*, *CDKN2B* was the most universal SAG, which was upregulated in ~60% of the cell types (i.e. 18 out of 30) (**Fig. 4C**). Surprisingly, SASP markers such as *IL6* and *IL1B*, as well as regulators such as *NFKB* and *TGFBI*, did not meet our selection criteria and were therefore not included as SAGs. Among the known SASP factors, the neutrophil-attracting chemokine *CXCL8*, as well as the pro-inflammatory cytokine *MIF*, were enriched in senescent cells for ~16% of the cell types (i.e., 5 out of 30), and *SERPINE1* was enriched in senescent cells of only two cell types (**Fig. 4C**). Gene set enrichment analysis of SAGs revealed an enrichment in ontology terms associated with cellular respiration, mitochondrial translation, protein transport, and regulation of apoptotic processes **(Supp. Fig. 11D**).

The most universal SAGs detected in this study include the gene *IL32*, encoding for the proinflammatory cytokine Interleukin-32. *IL32* h was found to be enriched in senescent cells across 43% of the cell types is present in higher mammals but not in rodents, and has recently been shown to trigger cellular senescence in cancer cells[94]. This suggests that Interleukin-32 could be an important mediator of paracrine-induced senescence across multiple human tissues (**Fig. 4C**). Interferon stimulated gene *IFI27* was also enriched in senescent cells across 40% of the cell types (**Fig. 4C**). *IFI27* also known as *ISG12a* is known to possess potent anti-proliferative and tumor suppressive properties and has been reported to be enriched in senescent cells in models of down syndrome and Alzheimer's disease[95]. Additionally, *IFI27* has also been shown to show a strong and sustained expression during TNFα-induced senescence[96]. Other genes that were enriched across multiple tissues include *BIRC3* gene, which not only regulates apoptosis, but also modulates inflammatory signaling, mitogenic kinase signaling, and cell proliferation (**Fig. 4C**). *BIRC3* gene expression has previously been described as a potential survival factor for senescent glioma cells

(GBM), where targeting the product of *BIRC3* gene was shown to act as a senolytic strategy that triggered apoptosis of GBM cells[97].

Collectively, these results point towards the absence of a fully universal senescence program, which has been suggested recently by others[81,83]. Nonetheless, we sought other broadly expressed markers that might reflect more common senescence features. To gain a more comprehensive view of senescence phenotypes across various cell types and tissues, we clustered the transcriptomes of senescent cells from various donors, tissues, and cell types using only the 3792 SAGs. We observed that this set of SAGs resulted in the senescent transcriptomes clustering together largely by related cell types, and independently of the tissue of origin or the donor (**Fig. 4, D and E**). This indicates that while there is no universal senescence phenotype, there are shared effects which are broadly shared across certain cell types. To systematically characterize the heterogeneity of senescence phenotypes across all cell types, we identified senescence-associated gene modules co-expressed within the same cell types, tissues, and donors, and summarized enrichment scores for these phenotypes for all cell types (**Methods, Supp. Fig. 11B**). Seventeen gene expression modules, each representing distinct biological processes, were variably enriched across different cell types (FDR q-value < 0.05, **Fig. 4F**), and it appears that the various different forms of human cellular senescence can be explained by combinations of these programs, along with some degree of cell-type specific changes.

Gene set enrichment analysis of senescence-associated pathways revealed that senescence programs are notably modulated by cell types and their lineage context. We found that physical cohesion pathways, such as strong homophilic cell-adhesion driven by protocadherin genes such as *PCDH17* and *DSCAML1*, were enriched in senescent barrier and structural cells such as epithelial cells and fibroblasts (**Fig. 4F**). Similarly, genes associated with epidermal development such as *SCEL*, *KLK7*, and *SPINT1* were enriched in senescent epithelial cells and fibroblasts (**Fig. 4F**). Cell-matrix adhesion-related genes, such as *ITGB1* and *ITGB8*, which encode β-integrins, were significantly enriched in senescent epithelial cells, fibroblasts, and endothelial cells. (**Fig. 4F**). Changes in production of cell adhesion molecules including β-integrins have been described in senescent cells before[98,99]. Interestingly, genes related to transcription initiation at RNA polymerase II promoter pathway, such as *PPARGC1A*, *GTF2H5*, and *ERCC3,* were found to be upregulated in senescent innate lymphoid cells and granulocytes, likely representing a compensatory response to meet the enhanced transcriptional demands of energy-intensive SASP production, metabolic reprogramming, and the maintenance of cellular viability despite growth arrest[100] (**Fig. 4F**). Secretory-pathway pressure was further underscored by the selective enrichment of ER-to-Golgi vesicle transport genes, such as *TRAPPC3* and *VCP*, in senescent intestinal epithelial cells and signal-peptide processing genes, such as *IMMP2L* and *HM13*, in senescent neurons and granulocytes. Together with the strong proteasomal and general protein-catabolism scores seen in granulocytes, lung epithelium, and stromal cells, these data reaffirm the centrality of proteostatic stress in senescence[101,102].

We found that metabolic rewiring pathways were variably enriched in senescent cells across a wide variety of cell types. Genes associated with proton motive force-driven ATP synthesis as well as broader cellular respiration, such as *ATP6V1A*, *ATP5F1B, MT-CO2*, and *COX7A2* were enriched in senescent cells of intestinal epithelium, myeloid leukocytes, and stem cells, echoing mitochondrial dysfunction that couples bioenergetic decline to SASP induction[103,104] (**Fig. 4F**). These results align with the growing appreciation that senescent cells can up-regulate oxidative phosphorylation to sustain the energy-intensive

senescence-associated secretory phenotype (SASP)[105,106]. Mitochondrial maintenance thus appears not to be a bystander but an active participant across immune, epithelial, and progenitor compartments. Genes involved in hypoxia responses such as *PTGS2*, *PTGIS*, and *INHBA* were predictably enriched in endothelial and fibroblast populations that dwell in oxygen-gradient niches.[107] (**Fig. 4F)**. Vacuolar acidification associated genes such as *ATP6V0D1* and *ATP6V1D* were enriched concurrently in endothelial, epithelial, stromal and stem cells, reinforcing that enlarged, hyper-acidic lysosomes are a hallmark of senescence, and their prominence across both differentiated and progenitor cell types further cements the universality of lysosomal remodeling in senescence[108]. Lymphoid cells showed an enrichment for the G1/S cell cycle arrest, with increased *CDKN2A* and *CDKN2B* enrichment paralleling the accumulation of $p16^{+}$ immune cells that evade clearance during ageing. In contrast, myeloid cells showed high MiDAS and proteasome activity related genes, implying a hyper-metabolic, SASP-potent phenotype distinct from the quiescent lymphoid state (**Fig. 4F**).

Collectively, we discovered distinct and cell type-specific fates for these cells that undergo cell cycle arrest. This constellation of pathways underscores that cellular senescence is a multifaceted state combining cell-cycle arrest with strategic rewiring of adhesion, metabolism, proteostasis, and organelle function, tailored to each cell types's role in tissue homeostasis and age-related pathology. Our results show that while there is not a single type of senescence, there are broad programs shared by several classes of cell types across tissues.

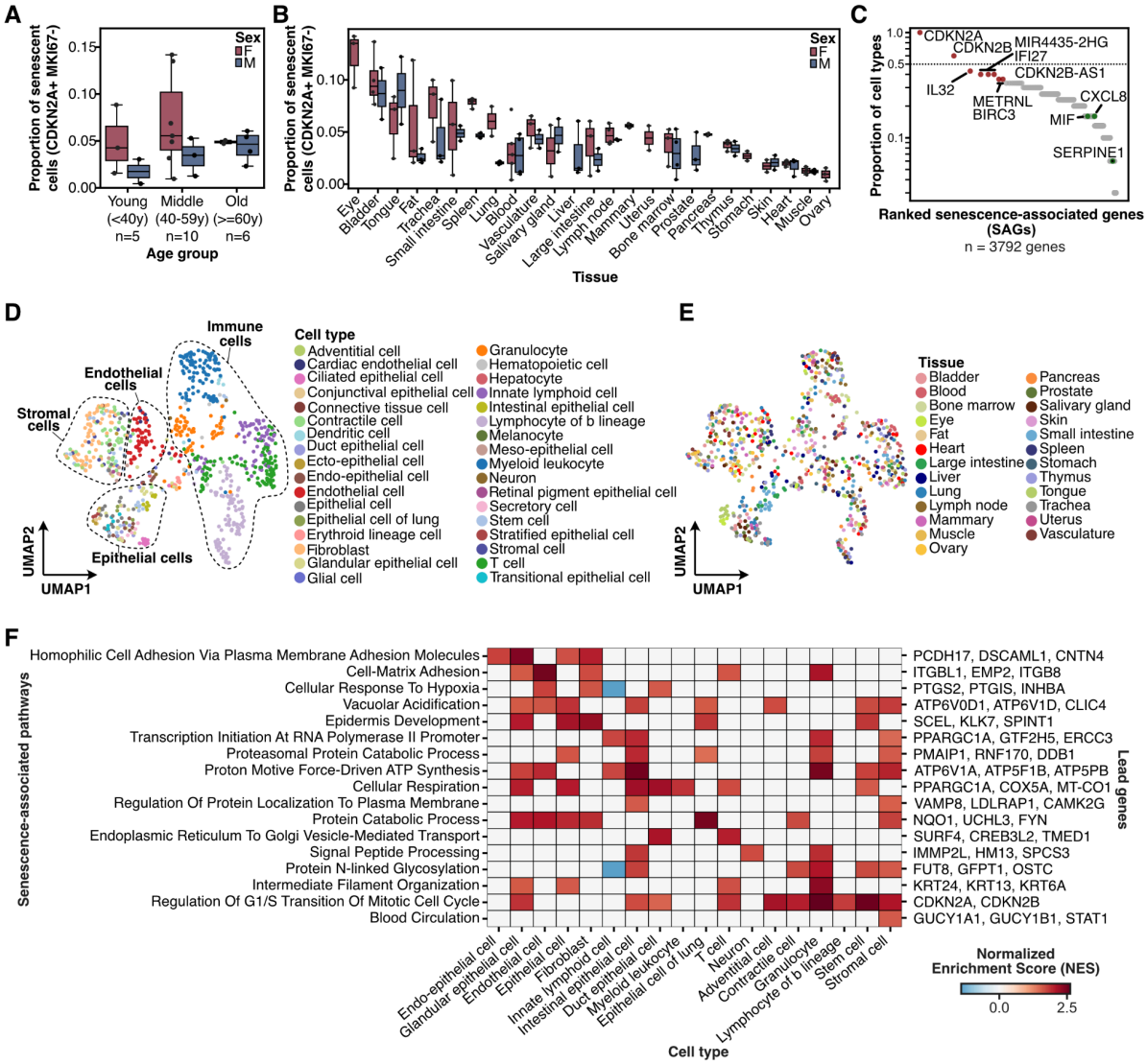


Figure 4: **Molecular profiles of senescent cells across human tissues. A)** Box plot showing the proportion of senescent cells, defined as *CDKN2A+ MKI67-* cells, across donors from three age groups. The donors within each age group were split by sex. The error bars represent the 95% confidence interval. **B)** Box plot showing the proportion of senescent cells, defined as *CDKN2A+ MKI67-* cells, across 25 tissues from 21 donors. The donors within each age group were split by sex. The error bars represent the 95% confidence interval. **C)** Dot plot showing the cell type prevalence for 3792 senescence-associated genes (SAGs). Cell type prevalence represents the number of cell types where a gene was upregulated in senescence cells as compared to non-senescent cells of that cell type from any tissue. A gene is considered to be upregulated in a cell type if the mean $\log_2$ fold-change is greater than 0.5. The top 15 most universal SAGs are highlighted in red, while known senescence-associated secretory phenotype (SASP) genes are highlighted in green. **D-E)** A Uniform Manifold Approximation and Projection (UMAP) plot showing pseudo-bulked transcriptomes of senescent cells, averaged for donors, tissues, and cell type combinations, and clustered based on the expression of senescence-associated genes alone. **D)** The pseudo-bulk transcriptomes are colored according to the cell type they represent. **E)** The pseudo-bulk transcriptomes are colored according to the tissue of origin. **F)** Heat map showing normalized enrichment scores (NES) for senescence-associated pathways across broad cell types. Enrichment scores were derived from gene set enrichment analysis using ranked enrichment between senescent and non-senescent cells for each broad cell type. Senescent

associated pathways were filtered for FDR < 0.05 to retain pathways significantly enriched in at least one broad cell type.

The impact of sex on gene expression across tissues and cell types

With the comparative sex information from male and female donors in the Tabula Sapiens 2.0 dataset, we systematically investigated differences in gene expression across a wide range of cell types. Sexual dimorphism in gene expression has been well-documented at the tissue and species level[109–113], with implications for differential disease susceptibility, drug responses and biological processes between males and females[114–118]. These differences are influenced by variations in sex chromosomes, hormonal signaling, transcription factor activity, and environmental influences[119]. One classic example of sex-dependent gene expression is the *XIST* gene[120], which plays a key role in X chromosome inactivation in females, ensuring dosage compensation between sexes. Other X-linked genes, such as *DAX1*[121], *KDM6A*[122] exhibit sex-specific expression patterns that impact processes like histone modification, gene regulation, and reproductive tissue function. Y-linked genes such as *SRY*[121] and *TSPY*[123], are crucial for male sex determination and spermatogenesis respectively. These genes are exclusively expressed in males. Recent studies also underscore tissue-specific gene expression contributing to sex differences. For example, genes like *CYP3A4* and *STAT3* are more highly expressed in females in the liver, affecting drug metabolism and pharmacokinetics[124]. Additionally, *LEP*, with higher expression in females, regulates energy balance, appetite, and fat distribution[125], while differences in *PPARG* expression between sexes influence immune responses, including the regulation of inflammation and T cell differentiation, in addition to its roles in adipogenesis, insulin sensitivity, and lipid metabolism[126,127].

Despite progress in understanding sex effects on gene expression across tissues, gaps in understanding remain at the cellular level, particularly regarding variability within cell types. We examined the prevalence of sex-biased gene transcripts across 23 donors, 12 tissues and 20 broad cell type groups from Tabula Sapiens 2.0 datasets (**Methods, Supp. Fig. 12**). Out of a total of 61,852 gene transcripts evaluated, we identified a diverse set of sex-biased genes varying across different cell types and tissues. Sex-biased genes that are specific to tissue-cell type resolution partially overlap with those identified at the tissue resolution and are almost equal in number in male and female (**Fig. 5A)**. These sex-biased genes showed a skewed pattern of tissue-cell type sharing (**Fig. 5B**), which chromosome X and Y-related genes such as *UTY*, *RPS4Y1*, *XIST*, *LINC00278, EIF1AY* and *DDX3Y* were among the most shared across tissue-cell type pairs, consistent with the pattern found in Genotype-Tissue Expression (GTEx) project at tissue resolution[117].

Studies on sex differences using bulk tissue GTEx data have revealed small but statistically significant effects of sex on gene expression, driven by sex chromosomes and hormones across various tissues, and suggest that sex correlates with tissue cellular composition[117]. Analysis of sex differences within Tabula Sapiens also reaches similar conclusions at both the pseudobulk tissue and cell type level (**Supp. Fig. 13**). Notably, we observe marked differences in the cell-type specificity of these mechanistic categories. Y-linked genes (e.g. DDX3Y, UTY) exhibit robust and consistent male-biased expression across many cell types, reflecting their strict male specificity and the lack of dosage compensation on the Y chromosome (**Supp. Fig. 13A**). Interestingly, TBL1Y, which has an X-linked homolog (TBL1X), shows variable male-biased expression across cell types. This likely indicated cell-type-specific regulation and

partial redundancy with TBL1X, which escapes X-inactivation in some contexts[128]. This identification highlights the complexity of Y-linked gene expression beyond binary sex specificity. Some additional Y-linked genes such as OFD1P12Y, XGY1, and GYG2P1 are annotated as pseudogenes, and showed inconsistent male-biased expression across cell types, probably due to their non-coding nature. X-inactivation escapees (e.g. XIST) show consistent female-biased expression patterns across a wide range of cell types. However, consistent with previous reports[128], we also observe variability in escape status across cell types, indicating the dynamic nature of XCI escape, and its contribution to both consistent and cell-type-specific sex-biased expression (**Supp. Fig. 13B-C**). Hormone-responsive genes exhibit more cell-type specific sex-biased expression. In particular, we observed that androgen-responsive gene sets are enriched for male-biased expression in specific cell types, including bladder transitional epithelial cells, bone marrow hematopoietic cells, and bladder myeloid leukocytes. In contrast, estrogen-responsive gene sets were enriched for female-biased expression in cell types such as heart fibroblasts, muscle myeloid leukocytes, and others. This suggests that hormone signaling drives localized sex differences, in contrast to the broader, more consistent effects of Y-linked genes and X-inactivation escapees (**Supp. Fig. 13D-E**). We also identified the tissue and cell type specificity of expression of sex-biased transcription factors (**Supp. Fig. 13F**).

To further investigate X chromosome inactivation (XCI), we collected sex-differentially expressed genes at an FDR 1<% across tissues and cell types. We then curated a list of XCI-related genes from past study[128], which classified X-linked genes into three categories: escape, variable, and inactive with respect to XCI status. By intersecting our significant sex-differential genes with these annotated XCI categories, we observed that genes known to escape XCI are significantly enriched among female-biased genes (**Supp. Fig. 14**). This trend was consistent at both the tissue and cell-type levels. These results support the role of XCI escapees in shaping female-biased expression and align well with previous findings from bulk RNA-seq datasets[117,128].

The bulk nature of GTEx data poses challenges to identify cell type-specific effects. A significant limitation is that bulk gene expression averages out cell type-specific differences, masking the heterogeneity within tissue samples and resulting in small effect sizes that may obscure important cell type-specific variations[129]. Therefore, we compared the sex-biased genes identified from Tabula Sapiens with those from the GTEx project to identify both congruences and disparities. Tabula Sapiens has a greater number of gene transcripts (61,852 transcripts) compared to GTEx (35,431 transcripts). Our comparison revealed that 38.2% of differentially expressed (DE) genes identified in GTEx were also detected in the Tabula Sapiens dataset across 8 tissues common to both projects, although the overlap varied by tissue (**Supp. Fig. 15A-C**). Notably, 72.3% of chromosome X-linked DE genes identified in GTEx were also found in Tabula Sapiens, along with an additional 254 chromosome X-linked genes **(Supp. Fig. 15D)**. Interestingly, effect sizes for X chromosome genes identified in Tabula Sapiens were generally significantly larger than those reported by GTEx, with the notable exception of XIST (**Supp. Fig. 15D)**. Additionally, our single cell analysis revealed more ubiquitous sex dependent effects than seen in GTEx; for example, *TSIX* was identified in the bulk GTEx data as sex-dependent in only a few tissues, but was sex dependent in nearly all Tabula Sapiens tissues (**Supp. Fig. 15D**). Interestingly, sex dependent gene expression differences are clearly shared across related cell types independent of the tissue of origin. This can be seen by measuring the correlation of differential gene expression across cell types, highlighting the role of cell type in driving gene expression patterns (**Fig. 5C**). In comparing this single cell analysis with the bulk GTEx data, there is agreement in the correlation heatmap of sex-biased gene

expression when focusing on tissue-specific differentially expressed (DE) genes, but the tissue-cell-type-specific analysis from Tabula Sapiens revealed a greater overlap of DE genes across tissues. This overlap further indicates that cell type-specific DE genes are shared across tissues, suggesting that sex bias is primarily driven by cell type specificity rather than tissue-level differences (**Supp. Fig. 15E).**

In addition to the recognition of sex-biased genes residing on the X and Y chromosome shared across more than 80 tissue-cell types, we also identified sex-biased protein-coding genes on autosomes. Notably, several genes, including *NABP1, HSPA1B* and *HSPA1A*, emerged as the most prevalent sex-biased genes across tissues such as vasculature, bladder, fat, bone marrow, as well as in immune cells such as myeloid leukocyte and T cells, and stromal cells such as contractile cells and fibroblast (**Supp. Fig. 16A**). Though the effect size for these genes were relatively small (**Fig. 5B**), they play ubiquitous roles in cellular signaling, hormonal regulation, immune response, and cellular stress responses. For example, *HSPA1A* is one of the heat shock proteins (HSPs), identifying as a lower concentration in female vasculature and potentially indicating higher risk in atherosclerotic disease[130]. *CTBP1-AS*, a corepressor of androgen receptor, is generally upregulated in prostate cancer in males[131], while *BCL6* in male has been associated with maintaining sex-dependent hepatic chromatin acetylation, with a trend of overt fatty liver and glucose intolerance in males[132]. In contrast, genes such as *CIRBP*, *ABCA5* and *HLA-DPB1* were the most prevalent genes enriched in females. Other genes with large effect sizes, such as *LCN2, PTGDS, RNASE1* were also enriched in females (**Fig. 5B)**. HLA genes play a crucial role in immune responses, and recent studies suggest they contribute disproportionately to sex biased vulnerability in immune-mediated diseases, such as autoimmune disorders[133–135]. Additionally, sex differences have been linked to the extent to which HLA molecules influence the selection and expansion of T cells, as characterized by their T cell receptor variable beta chain[136]. Adipose LCN2 is frequently enriched in females, where it acts in an autocrine and paracrine manner to promote metabolic disturbances, inflammation, and fibrosis in female adipose tissue[137]. Conversely, several genes, including *SELENOH*, *USP53*, and *ZNF618*, were upregulated in males but not females, showing high effect sizes and abundance. This distinct expression pattern between sexes suggests that males and females may employ different molecular pathways to regulate key biological processes. The identification of these shared and significantly affected genes underscores their potential impact on sex-specific traits and disease susceptibilities. Furthermore, a density plot of concordance score across different gene types showed that chrY and chrX genes exhibit the most consistent sex effects, while mitochondrial genes demonstrated the least concordance (**Fig. 5D, Supp. Fig. 16B**).

Given the cell type specificity of sex differential expression, we further investigated which cell types exhibited the greatest variability in sex-based gene expression. Cell types with the largest number of sex-biased genes include myeloid leukocyte, innate lymphoid cell, epithelial cells, and granulocytes in tissues such as bone marrow, vasculature, and blood (**Supp. Fig. 16C-D**). Scatter plots of effect sizes in female and male donors indicated the highest mean effect size in fat, bone marrow tissues, particularly in stem cells, fibroblast and immune cells, such as lymphocytes and myeloid leukocyte (**Fig. 5E**). These findings align with previous studies that emphasize the importance of cell-type-specific analyses in uncovering sex differences in gene expression. For example, fibroblasts, which show the highest mean effect size, are known for their crucial roles in tissue homeostasis, processes that can be differentially regulated by sex hormones[138,139]. Immune cells such as lymphocytes, especially in the fat and bone marrow, reflect known sex differences in immune responses[140]. To link bulk and single-cell findings, we

examined autosomal genes from GTEx that were most strongly co-expressed with X-linked genes and compared them to top-ranking tissue-cell types identified in Tabula Sapiens. These GTEx-derived autosomal gene modules were highly correlated with the TS-inferred sex-biased cell types (**Supp. Fig. 17**).

We summarized the representative functional enrichment for each cell type between male and female donors (**Fig. 5F**). In general, pathways are shared by both sexes but can be upregulated in a sex-dependent fashion in a highly cell type specific manner. In females, pathways related to cellular respiration and fatty acid metabolism are enriched in female fibroblast, while immune activity is enriched in granulocyte and T cells. In males, pathways associated with inflammatory response were predominantly enriched in endothelial cells and myeloid leukocytes, while androgen responses were predominantly enriched in epithelial, myeloid leukocyte and hematopoietic cells. These findings suggest that males and females leverage different molecular pathways to regulate key biological processes, which may contribute to sex-specific disease susceptibilities and physiological functions.

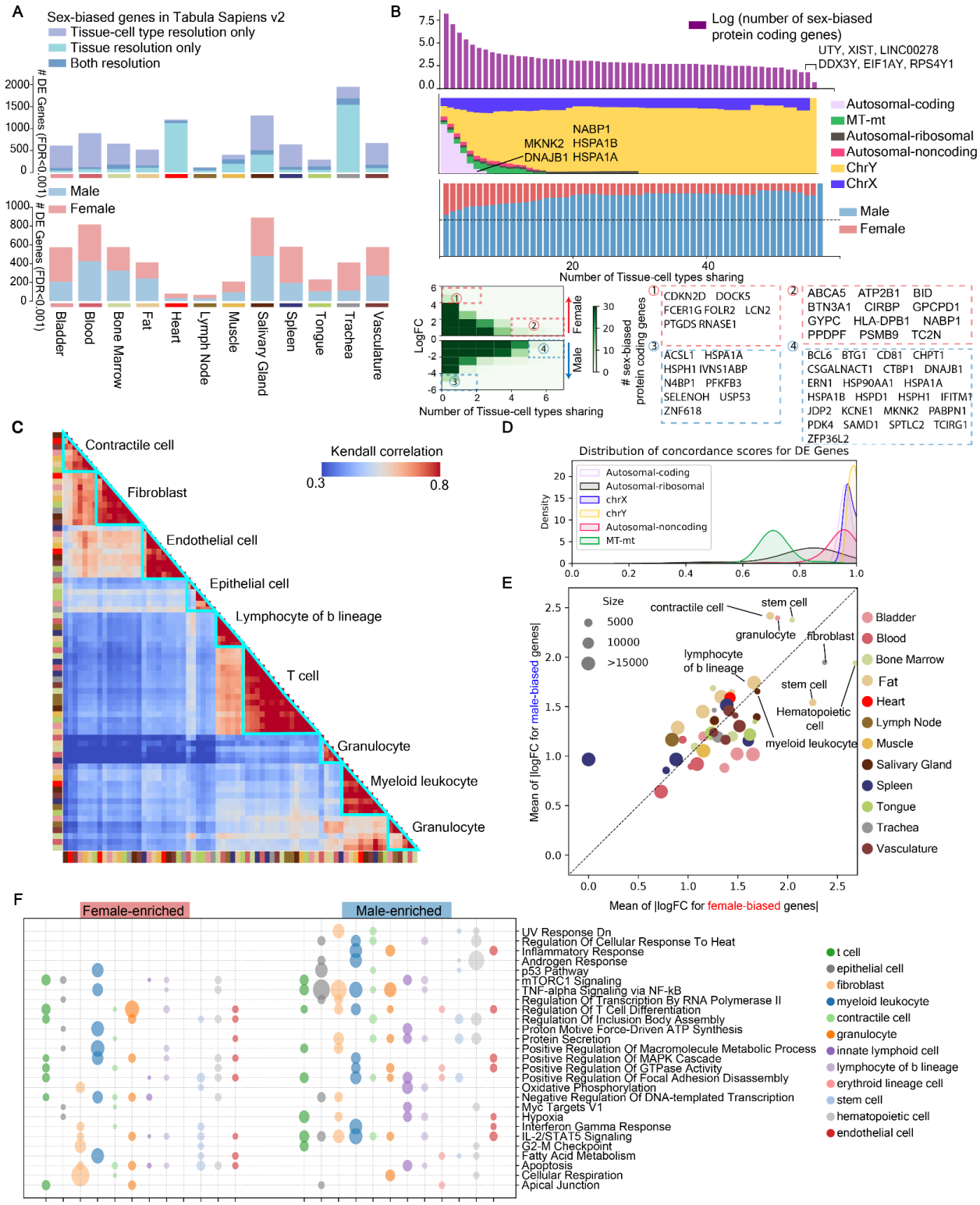


Figure 5: **Sex-biased gene expression across tissue-cell types. A**) Bar graphs displaying the number of sex-biased genes identified in the Tabula Sapiens 2.0, at both the tissue resolution and tissue-cell type resolution. Top: the total number of sex-biased genes across tissue-cell types in Tabula Sapiens 2.0, and overlap with sex-biased genes found at the tissue resolution. Bottom: The proportion of male- and female-enriched genes identified at the tissue-cell type level, and displayed for each tissue. **B**) Top: A bar plot showing the log-transformed number of sex-biased

protein-coding genes and the extent to which they are shared across tissue-cell types. The number represents the tissue-cell types sharing that gene, with values greater than the indicated threshold. Middle: Heatmaps showing the distribution of gene types (e.g., autosomal-coding, X/Y-linked) alongside the number of tissue-cell types sharing them. Bottom left: The distribution of log-transformed fold changes (logFC) for male- and female-enriched genes, grouped by tissue-cell type sharing. Bottom right: A detailed gene list for the numbered regions from the heatmap. **C**) Clustering heatmap of Kendall correlations for sex-biased gene expression across major cell types. The inner bar denotes the specific cell type, while the outer bar represents the tissue of origin. The clustering highlights patterns of similarity in sex-biased gene expression between cell types, revealing which tissues and cells share similar sex-specific expression profiles. **D**) Distribution of concordance scores for differentially expressed genes, categorized by autosomal-coding, X/Y chromosome-linked, and mitochondrial genes. **E**) Scatter plot showing the mean logFC for female- and male-enriched genes across cell types. The size of the dots represents the number of sex-biased genes identified in each tissue-cell type. **F**) Bubble plot of representative ontology terms for major cell types. The bubble size corresponds to the adjusted p-value of the enriched terms. The color coding for cell types and tissues is consistent throughout the figure.

Searchable Extended Metadata

The Tabula Sapiens 2.0 also includes de-identified medical records for all donors. These medical records are voluminous (an average of 13 pages per donor, 291 total pages) and the overhead associated with their interpretation is often prohibitive. Therefore, we developed ChatTS (https://singlecellgpt.com/chatTSP?password=chatTS; source code available at https://github.com/Harper-Hua/ChatTS)—a web application powered by a large language model—to facilitate use of Tabula Sapiens for the broader research community.

On the front end, a user asks a question about donors' medical records. Then, on the back end, an external LLM is prompted to answer this question on a per-donor basis, using free-form text that was manually extracted from each donor's medical record (including physical, imaging, and laboratory studies, as well as social and medical histories). Results are concatenated and returned as a table which can be downloaded and manually verified (**Fig. 6A**). Interestingly, we observed that the underlying LLM can interpret charts and offer conclusions that are not explicitly stated in the records (**Fig. 6B**). For example, when asked which donors have heart disease, the response did not require the explicit words "heart disease" in the medical records, but rather inferred it from symptoms and treatments.

We benchmarked ChatTS with a GPT-4o backend against 330 manually curated question-answer pairs regarding donor metadata. ChatTS showed satisfactory performance (83.2% agreement with manually curated data), and in many cases, the ChatTS-generated answers were more informative than the manually curated data, **Supp. Table 7**. Differences between manual and ChatTS answers tended to be related to ambiguities in the questions, their definitions, and information present in the medical records, and not to hallucination.

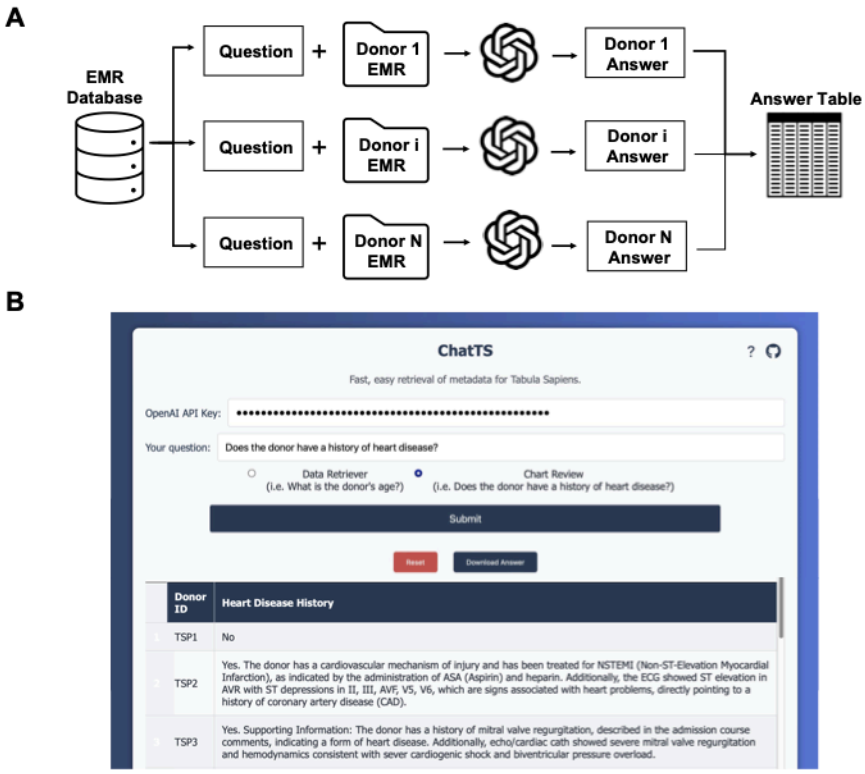


Figure 6. ChatTS web application overview. A) Schematic representation of the ChatTS web application architecture. **B)** Screenshot of the application, demonstrating inference based on information in charts.

Limitations of the Study

RNA abundance, while informative, does not fully capture protein levels or post-translational modifications, limiting to some extent inference on functional protein states and short time scale behavior of the cells. Cell type abundances measured by dissociation do not necessarily represent true biological abundance due to differential dissociation effects and because enrichment was performed by compartment (immune, epithelial, endothelial and stromal) to maximize the number of cell types captured. Conversely, very rare cell populations may be missing or due to the requirements of live single-cell RNA-seq, as samples were required to be processed quickly and with consistent methods overnight for each donor. As might be expected of any adult who has lived a full life, none of our donors were perfectly healthy. All donors are missing some tissues due to organ transplantation, which always had a higher priority than our study.

Conclusions

The Tabula Sapiens 2.0, featuring over 1.1 million cells across 28 tissues from healthy donors aged 22 to 74 years, offers a comprehensive multiorgan dataset enabling systematic analysis of rare cell populations, while accounting for technical artifacts and donor-specific variations. Global analysis of transcription factor expression enabled a comprehensive analysis to identify which transcription factors are cell type specific and which are ubiquitously expressed. Collectively, we observe heterogeneous and cell

type-specific senescence phenotypes and identify several novel and sometimes contradictory fates for cells that undergo cell cycle arrest. Understanding the diverse senescence phenotypes across different cell types offers promising avenues for developing targeted senolytics, tailored to address age-related diseases affecting specific tissues and cell types. Finally, sex-dependent gene expression analyses provide the beginnings of an understanding which genes and cell types are most strongly sexually dimorphic and the consequent implications for cellular and tissue functions. By comparing male and female gene expression profiles, we identified significant patterns and quantified the impact of sex on gene regulation, contributing to our understanding of biological differences and their consequences for health and disease.

Tabula Sapiens 2.0 is the most comprehensive human cell atlas reference to date, as measured by the breadth of tissues and cell types. The vignettes included in this publication illustrate the kind of exploratory analysis and hypothesis building that such a dataset enables. Following on the previous Tabula projects[1,21,141,142], all data is available for the community to continue learning and improving as a reference atlas until we have a deeper understanding of the molecular representation of cell types and cell states. The community found many uses of Tabula Sapiens 1.0, ranging from being a tool to predict potential on target toxicity of drug candidates to being a reference standard used to evaluate bioinformatic tools and train AI models, and we expect that the expanded version described here will only increase such applications. For example, the discovery of cell type specific transcription factor usage will point the way for future biochemical studies to characterize the function of these transcription factors, as will the identification of previously undetected ubiquitous transcription factors which are found in all cell types. We have also already seen the use of Tabula Sapiens 2.0 as a hold out data set to test the zero shot capability of AI models which are trained on Tabula Sapiens 1.0.

Resource availability

Lead contact

Further information and requests about reagents and other resources used in this work should be directed to and will be fulfilled by the lead contact, Stephen R Quake (steve@quake-lab.org).

Data and code availability

- The entire dataset can be explored interactively at the Tabula Sapiens Data Portal[143] (https://tabula-sapiens-portal.ds.czbiohub.org/).
- The code used for the analysis is available on the github repository for Tabula Sapiens (https://github.com/czbiohub-sf/tabula-sapiens/).
- Gene counts and metadata are publicly available from figshare[144] and cellxgene[143].
- 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.
- To preserve the donors' genetic privacy, we require a data transfer agreement to receive the raw sequence reads. The data transfer agreement is available in the data portal.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Acknowledgements

This project has been made possible in part by grant nos. 2019-203354, 2020-224249, 2021-237288, 2021-006486, 2022-316725. from the Chan Zuckerberg Initiative DAF; an advised fund of Silicon Valley Community Foundation; and by support from the Chan Zuckerberg Biohub San Francisco.

We thank Donor Network West for procuring the organs and tissues from the donors for this project. We are also grateful to the UCSF Liver Center (funded by NIH P30DK026743) for assistance with the liver cell isolations, and B. Tojo for the original artwork in Fig. 1. We would like to express our gratitude and thanks to donor WEM and his family at Donor Network West, as well as to all the anonymous organ and tissue donors and their families for giving both the gift of life and the gift of knowledge through their generous donations. We would also like to thank Dr. Konstantin Kahnert in Dr. Emma Lundberg's lab for extracting subcellular location information from the Human Protein Atlas website.

Declaration of interests

The authors declare no competing interests.

STAR methods

Key resources table

[will be moved here once manuscript is ready, for now please check the file STARmethods_KeyResourcesTable_TabulaSapiensv2.docx]

Organ and tissue procurement

To maintain consistency in the overall Tabula Sapiens dataset, organ and tissue procurement followed the same procedure used in the first phase of the project. Donated organs and tissues were procured at various hospital locations in the Northern California Region through collaboration with a not-for-profit organization, Donor Network West (DNW, SanRamon, CA, USA). DNW is a federally mandated organ procurement organization (OPO) for Northern California. Recovery of non-transplantable organ and tissue was considered for research studies only after obtaining records of first-person authorization (i.e., donor's consent during his/her DMV registrations) and/or consent from the family members of the donor. However, the pancreas from donor TSP9 was provided by Stanford University Hospital under appropriate regulatory procedures. The research protocol was approved by the DNW's internal ethics committee (Research project STAN-19-104) and the medical advisory board, as well as by the Institutional Review Board at Stanford University which determined that this project does not meet the definition of human subject research as defined in federal regulations 45 CFR46.102 or 21 CFR 50.3.

Tissues were processed consistently across all donors. Each tissue was collected, and transported on ice, as quickly as possible to preserve cell viability. A private courier service was used to keep the time between organ procurement and initial tissue preparation to less than one hour. Single cell suspensions from each organ were prepared in tissue expert laboratories at Stanford and UCSF. For some tissues the dissociated cells were purified into compartment-level batches (immune, stromal, epithelial and endothelial) and then recombined into balanced cell suspensions in order to enhance sensitivity for rare cell types as described further below.

Tissue preparation protocols and cell suspension preparation protocols

The methods for tissue dissociation and staining of cell suspensions for bladder, blood, bone marrow, eye, fat, heart, intestine, kidney, liver, lung, lymph node, mammary gland, pancreas, prostate, salivary gland, skeletal muscle, skin, spleen, tongue, trachea, thymus, uterus, and vasculature are described in the methods section of the Tabula Sapiens[1].

Inner Ear - Initial tissue preparation protocol

Whole organ utricles were collected from organ donors. Bilateral utricles were harvested from organ donors as previously described[145–147]. Briefly, a post-auricular incision was made followed by a modified transcanal approach to expose the middle ear. The tympanic membrane, malleus and incus were removed while keeping the stapes in situ. To expose the vestibular organs, the bony covering of the vestibule was thinned using a diamond burr on low speed, the stapes footplate removed, and the oval window widened.

The utricle was harvested from the elliptical recess and placed in PBS on ice for single-cell RNA sequencing analysis.

Inner Ear – 10x Genomics sample preparation

Utricles from organ donors were placed in DMEM/F12 (Thermo Fisher Scientific/Gibco, 11-039-021) with 5% FBS during transport to the laboratory. Tissues were then washed twice in DMEM/F12 and any debris and bone microdissected away. The whole utricle was then incubated with thermolysin (0.5 mg/mL; Sigma Aldrich, T7902) for 45 min at 37ºC. Next, the whole utricles were digested using Accutase (Thermo Fisher Scientific, 00-4555-56) for 50 min at 37ºC and single cell suspension was obtained by trituration using a 1 mL pipette. Single cell suspension was achieved using a 40 µm filter. DMEM/F12 media with 5% FBS was used for this step. Cells were then centrifuged at 300g for 5 min at 4ºC. The supernatant was removed, and cells were resuspended in media. Number of cells per microliter was quantified using a hemocytometer

Large and Small Intestines – Initial Tissue Preparation Protocol

Segments of the small intestine (duodenum and ileum) and large intestine (ascending and sigmoid colon) were transported on ice to Stanford University. Upon arrival, the tissues were sectioned into approximately 5 cm segments and rinsed multiple times with 35 mL of ice-cold PBS (Thermo Fisher Scientific 10010023) to remove residual lumen contents. The tissues were then opened with the mucosal side facing up, and 10 mL of ice-cold PBS was injected using a 21G needle. Mucosal blebs were then dissected, minced into approximately 2 mm pieces, and digested in 10 mL of digestion medium (DMEM/F12: Thermo Fisher Scientific 12634010, 1x Penicillin-Streptomycin-Neomycin: Sigma P4083, 1x Normocin: Invivogen ant-nr-2, 1x HEPES buffer: Corning 25-060-CI, 1x Glutamax: Thermo Fisher Scientific 35050061, and 10 µM ROCK inhibitor Y-27632: MedChem Express HY-10583) containing Collagenase III (300 U/mL: Worthington Biochemical Corporation LS004182) and DNase I (200 U/mL: Worthington Biochemical Corporation LS002139) at 37°C for 90 minutes. During digestion, the mixture was pipetted every 15 minutes with a 10 mL pipette to facilitate tissue breakdown. After digestion, 30 mL of intestinal wash buffer (HBSS: Corning 21-022-CV, 2% FBS, 1x Penicillin-Streptomycin-Neomycin , 1x Normocin, and 10 µM ROCK inhibitor Y-27632) was added to the suspension, which was then centrifuged at 1500 rpm for 5 minutes at 4°C. The supernatant was discarded, and the pellet was resuspended in 5 mL of ACK lysis buffer (Thermo Fisher Scientific A1049201) to lyse red blood cells and incubated for 5 minutes at room temperature. The suspension was centrifuged again, and the pellet was resuspended in intestinal wash buffer with DNase I, then filtered through a 100 µm cell strainer (Miltenyi Biotec 130-098-463). The final cell suspension was counted using trypan blue (Thermo Fisher Scientific 15250061) and adjusted to a concentration of 10^6 cells/mL for direct single-cell analysis using the 10x Genomics platform or flow sorting.

Ovary - Initial tissue preparation protocol

Ovaries were finely dissected to remove surrounding tissues. Ovaries were gently chopped using a razor blade and then transferred to a 50 mL Falcon™ conical tube with a final volume of 30 mL 2 mg/mL collagenase type IV in DMEM/F12. Tubes were incubated at 37 ºC shaking at 250 rpm for 20 min then centrifuged at 250 x g for 5 min. The supernatant was aspirated and 30 mL 0.25 % trypsin-EDTA was added. Tubes were incubated at 37 ºC shaking at 250 rpm for 20 min then centrifuged at 250 x g for 5

min. The supernatant was aspirated, and the tissue was quenched with 30 mL 10 % FBS in DMEM/F12. Using a P1000, tissue was gently triturated to generate a single cell suspension. The cell suspension was filtered using a reversible 37 um sieve. The sieve was rinsed with 1 mL quench buffer (<37 um filtered flow through) and then flipped and placed in a new tube for rinsing with 1 mL quench buffer (>37 um unfiltered mature oocyte pool).

Single oocytes and/or follicles were handpicked from the cell suspension under magnification (2 to 6X) on a dissection microscope using an EZ-Grip pipette with a 125 um EZ-Tip and washed in 0.04% BSA-PBS. Single oocytes and/or follicles were then transferred in 0.6uL of 0.04% BSA-PBS to individual wells in a 96-well plate containing a lysis buffer master mix. Lysed oocytes and/or follicles were stored at -80 ºC to enhance lysis prior to cDNA and sequencing library construction. Nine follicles, which were not dissociated into single-cell suspensions, represent the bulk transcriptomes of entire ovarian follicles.

Ovary - FACS/SS2 sample preparation

Filtered cell suspensions were centrifuged at 250 x g for 5 min, aspirated, and washed in FACS Buffer (2 % FBS-PBS). The following antibodies were used at a 1:200 dilution in FACS Buffer: EPCAM-PE (Biolegend, 324206); CD45-FITC (Biolegend, 304038); CD31-APC (Biolegend, 303116). The cell suspensions were incubated 1 hr on ice for staining, then washed 3 times. Prior to sorting, Sytox Blue was added at a 1:1000 dilution (ThermoFisher, S34857) and live, single cells were sorted for EPCAM+ (oocytes), CD45+ (endothelial), CD45+ (immune), EPCAM-/CD45-/CD31- (stromal/granulosa) populations into 384-well lysis buffer plates following color compensation. Plates were stored at -80 ºC until processing for sequencing.

Stomach – Initial Tissue Preparation Protocol

Ligated whole stomachs were transported on ice to Stanford University. Upon arrival, a 6x3cm longitudinal strip was cut from the anterior aspect of the stomach antrum, 6 cm proximal to the pylorus. Stomach tissue was first rinsed in 1X PBS to remove clots and debris, and blotted dry to remove excess mucus. The tissue was then dissected, separating the mucosa/submucosa layer from the muscularis/serosa layer, and the layers were then weighed. Each layer was digested separately to maximize viability. The mucosa was first incubated with 5mM EDTA in HBSS buffer without Ca^{2+} for 10 minutes on a shaker at 37 °C (x3), with washes and collection of epithelial cells. After this, both mucosa and muscularis layers were minced with fine sharp scissors in digestion media with 0.8 mg/ml collagenase Type IV Worthington (Sigma) and 0.05 mg/mL DNase I (Roche) (10mL per 2.5 gram of tissue). Two serial mucosa digestion were done in an orbital shaker (300 rpm) at 37 °C, 20 min each, and for muscularis 40 min each (vortexing halfway). Digestions were stopped with cold R10-EDTA, and cells were filtered from undigested tissue, washed and then stained with CD45-PE (Biolegend, #304007) for cell counting and sorting. The Mucosa digestion and EDTA epithelial fraction were combined and DAPI stained (Thermofisher Cat# D1306). Live DAPI-, CD45+ and CD45- cells were then FAC-sorted on a BD Aria. Sorted cells from mucosa and muscularis were mixed back together at a 70:30 CD45+ to CD45- ratio, and used for 10X single-cell RNA sequencing.

Testis- Initial Tissue preparation protocol

Testis tissues were dissected out from the tunica and lightly dissociated and minced using sanitized surgical forceps and razor blade. Two to three pieces of 0.5g testis tissue was placed in a 50mL Falcon tube with 10mL prewarmed collagenase solution at 32 °C (1 mg/mL collagenase I (Worthington Biochemicals #LS004196), 1 mM EDTA, 0.5ul/mL DNaseI (50mg/mL stock, Worthington Biochemicals #LS002139) in PBS). Testis tissue was first dissociated by vigorous pipetting and then incubated at 32 °C for 8 minutes with 2-3 times pipetting during the incubation to resuspend cells. After incubation, cells were dissociated again at room temperature by vigorous pipetting for 2-4 minutes. Cells were spinned down at 250xg for 5min at room temperature and then the supernatant was removed. The dissociation steps were repeated one more time before 5mL of TripLE (Thermo Fisher Scientific #12604013) plus 1uL/mL DNaseI (50mg/mL stock) was added to the dissociated cells and incubated at 32 °C for 15 min with pipetting every 5 min. Cells were then washed with 30mL PBS and filtered through a 70 µm cell strainer and then a 40 µm cell strainer. Cell number was recorded before cells were spin down at 250xg for 10min. Then the supernatant was removed, and cells were resuspended at 1000 cells/uL for cell capture in microfluidic droplets with the 10x Genomics platform.

Testis – FACS/SS2 sample preparation

After cells were spun down at 250xg for 10min and after passing the cell strainer, cells were washed once in 3mL FACS buffer (2%FBS (Thermofisher #A5670701), 1mM EDTA (Invitrogen #AM9260G) in PBS(Gibco #10010023)), spin down at 25xg for 5min, and resuspended in 400uL of FACS buffer. Antibodies cocktail (2 µL of EPCAM-PE (Biolegend, 324206), 2uL of CD45-FITC (Biolegend, 304038), 2uL of CD31-APC (Biolegend, 303116)) was added to the cell suspension, and cells were incubated on ice for 30min, with mix by tapping every 10min. After incubation, cells were washed twice with 2mL FACS buffer and spin down at 250xg for 5min, and resuspended in 1mL FACS buffer. Sytox blue was added before sorting.

Thymus– Initial Tissue Preparation Protocol

The tissue dissociation and staining of cell suspensions for thymus is described in the methods section of the Tabula Sapiens[1]. Our analysis revealed the presence of cells originating from fat and intra-thymic lymph nodes among those attributed to the thymus, likely due to difficulties in tissue collection caused by the natural involution of the thymus with age. Consequently, we excluded the thymus from our analysis of sex-biased gene expression.

10x Genomics protocol

10x Genomics kits used were Chromium Next GEM Single Cell 3′ Kit v3.1 or Chromium Next GEM Single Cell 5′ Kit v2. The protocols provided by the manufacturer were followed. In general, two 10X Chromium channels were loaded per tissue with 7000 cells, with the goal of obtaining data for 10,000 viable cells per tissue.

Organ and cell coverage

Our goals were to characterize the gene expression profile of 10,000 cells from each organ and detect as many cell types as possible. As explained in detail for each organ, about ⅔ of the organs employed a

MACS based enrichment strategy, either to balance cell types between four compartments; epithelial, endothelial, immune, and stromal. This ensured abundant cell types in one compartment did not mask rare cell types in another. Two 10x reactions per organ were loaded with 7,000 cells each with the goal to yield 10,000 QC-passed cells. Four 384-well Smartseq2 plates were run per organ. In most organs, one plate was used for each compartment (epithelial, endothelial, immune, and stromal), however, to capture rare cells, some organ experts allocated cells across the four plates differently. The use of two 10x reactions enabled some flexibility to distinguish in the data the anatomical position of the sample or allowed enrichments other than epithelial, endothelial, immune, and stromal. It also served as insurance against losing an entire organ due to a clog of the 10X chip.

Flow Sorting

Details of the sorting can be found in Tabula Sapiens[1]. Briefly, after dissociation, the single cells from each organ and tissue that were destined for plates were isolated into 384 plates via FACS. On some cell suspensions destined for 10x tissue compartments were enriched by FACS to balance the cell types as described in the tissue specific methods. Most sorting was done with SH800S (Sony) sorters. The last column of each 384 well plate was intentionally unsorted so that ERCC's controls could be used as a plate processing control. Immediately after sorting, plates were sealed with a pre-labelled aluminum seal, centrifuged, and flash frozen on dry ice to ensure full cell lysis. Typical sort times were 4 to 9 minutes per 384-well plate.

Smart-seq2 Protocol

Plate-based sequencing of tissues used the 384 plate modification of Smart-seq2[148] as described in Tabula Sapiens[1] and Tabula Muris Senis[142].

Smart-seq2.5 Protocol

Ovary plates in the object denoted "SS3" were processed into cDNA by a modified protocol called "Smart-seq2.5," which is a revision of the Smart-seq2 protocol that takes advantage of the improved reagents and reaction conditions used in Smart-seq3xpress[149]. Briefly, lysis plates were prepared by dispensing 0.5 µL lysis buffer (0.15% Triton X100 (Sigma-Aldrich, 93443-100ML), 10% polyethylene glycol 8000 (Sigma-Aldrich, P1458-50ML), 0.6 u/µL Recombinant RNase Inhibitor (Takara Bio, 2313B), 1 mM of each dNTP (Roche, NTMIXKB), 0.25 µM biotinylated oligo-dT30VN (Integrated DNA Technologies, 5'-biotin-AAGCAGTGGTATCAACGCAGAGTACT30VN-3′), and 1:600,000 ERCC spike-in RNA (Thermo Fisher Scientific, 4456740)) into 384-well hard-shell PCR plates (Bio-Rad HSP3901) using a Mantis liquid handler (Formulatrix). The lysis plates were then sealed with AlumaSeal CS films (Sigma-Aldrich, Z722634), spun down, snap-frozen on dry ice, and stored at -80 °C until sorting. Dissociated cells were sorted into thawed lysis plates by FACS as described above, after which the plates were spun down, snap-frozen on dry ice, and stored at -80 °C until further processing.

For cDNA synthesis, plates were first incubated at 72 °C for 10 min to ensure cell lysis and RNA denaturation. Then 0.5 µL RT mix (50 mM Tris-HCl pH 8.0 (Thermo Fisher Scientific, BP1758-100), 60 mM NaCl (Invitrogen, AM9760G), 5 mM $MgCl_2$ (Invitrogen, AM9530G), 2 mM GTP (Thermo Scientific, R1461), 16 mM mM DTT (Promega, P117A), 1.5 mM TSO (Integrated DNA Technologies, 5'-AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG-3'), 0.5 u/µL Recombinant RNase Inhibitor

(Takara Bio, 2313B), and 4 u/µL Maxima H Minus Reverse Transcriptase (Thermo Scientific Scientific, EP0751)) was dispensed into each well using a Mantis liquid handler. Reverse transcription was then carried out on a Bio-Rad C100 x384 thermal cycler using the following program: 1) 42 °C for 90 minutes, 2) 10 cycles of 50 °C for 2 min and 42 °C for 2 min, and 3) 85 °C for 5 min. The resulting cDNA was amplified by dispensing 1.5 µL pre-amplification mix (1.67x SeqAmp PCR buffer (Takara Bio, 638526), 0.083 µM ISPCR primer (Integrated DNA Technologies, 5'-AAGCAGTGGTATCAACGCAGAGT-3'), and 0.042 u/µL SeqAmp DNA polymerase (Takara Bio, 638504); concentrations were chosen to reach 1x SeqAmp PCR buffer, 0.5 µM ISPCR primer, and 0.025 u/µL SeqAmp DNA polymerase in the 2.5 µL final reaction) using a Mantis liquid handler and then performing PCR on a BioRad C1000 384-well thermal cycler (Bio-Rad) with the following program: 1) 95 °C for 1 min, 2) 13-19 cycles of 98 °C for 10 s, then 65 °C for 30 s, and 68 °C for 4 min, and 3) 72 °C for 10 min. The resulting cDNA was diluted 1:5 by adding 10 uL of buffer EB (Qiagen Cat#19086) using a Mantis liquid handler. Sequencing libraries were prepared and pooled from the diluted cDNA plates using the same protocol as for Smart-seq2.

Sequencing

All Sequencing was done by the CZBiohub San Francisco Sequencing Team. 10X libraries were loaded on Illumina NovaSeq 6000 S4 flow cells in sets of 16 libraries with the goal of generating 50,000 to 75,000 reads per cell. SmartSeq libraries were run in sets of 20 plate libraries with a target of 1M reads per cell.

Biobanking and Histology

Where possible, additional tissue samples were collected from the vicinity of sequenced specimens for frozen and fixed biobanking. Frozen samples were washed and flash frozen with liquid nitrogen. Fixed samples were fixed in 10% buffered formalin and paraffin embedded (FFPE). Hematoxylin and eosin (H&E) stained slides were generated from the FFPE samples using standard methods, digitally scanned using Leica Aperio AT2 23AT2100. Images are available on the Tabula Sapiens portal.

scRNAseq data extraction

Sequences from the NovaSeq 6000 were de-multiplexed using bcl2fastq version 2.20.0.4.22. Reads were aligned to the Gencode Reference version 41 (GRCh38) genome using STAR[150] version 2.7.11b with parameters TK. Gene counts were produced using HTSEQ[151] version 2.0.5 with default parameters, except 'stranded' was set to 'false', and 'mode' was set to 'intersection-nonempty'. Sequences from the microfluidic droplet platform were de-multiplexed and aligned using CellRanger version 7.0.1, available from 10x Genomics with default parameters.

scRNAseq data pre-processing and cell type annotations

Gene count tables were combined with the metadata variables using the Scanpy[152] Python package version 1.9.6. We first roughly filtered the dataset to remove any cells with fewer than 100 genes and 1000 counts (unique molecular identifiers or UMIs). In order to filter out reads from ambient RNA we ran DecontX[153] (implemented in R package celda v1.16.1) separately for each 10X run, using default parameters given the full background and the filtered cells. After the DecontX filtering step, we re-filtered the dataset more strictly using a minimum threshold of 200 non-mitochondrial genes and 2500 non-mitochondrial counts

for the droplet (10X) cells and 500 non-mitochondrial genes and 5000 non-mitochondrial counts for the smartseq FACS sorted cells. This filtered gene-count matrix was then used for the analysis.

Ambient RNA and barcode swapping[154] are known problems in 10x sequencing. After using DecontX to remove ambient RNA, we removed all cells sharing both the cell and transcript barcode but not the same sample barcode in each sequencing run in order to remove cells generated by barcode swapping.

In the analysis step, we first integrated the multiple batches of data from each donor to generate a unified visualization of the cells using scVI[155] from scvi-tools[156] release 0.20.0. For training the variational autoencoder neural network, we used the following hyper parameters: n_latent=50, n_layers=5, dropout_rate=0.1. We allowed each gene to have its own variance parameter by setting dispersion="gene". We trained the scVI model for 50 iterations with all available data and corrected the batch effect associated with donor and technology. scVI generated a harmonized latent space that was then projected to a 2D space using UMAP. This process was done individually for each organ, and we then shared the harmonized data along with the reduced dimensional latent space in a h5ad format data object compatible with both Scanpy and CELLxGENE. CELLxGENE is a data exploration and visualization tool that allows users to interactively explore any scRNAseq dataset[9,10]. Manual annotation was performed by tissue experts using CELLxGENE. Each data object contained three main components: gene count data, cell-wise metadata, and gene-wise metadata for their organ of interest. CELLxGENE allows the user to color cells by any cell metadata such as donor and compartment. Cells can also be colored by gene expression data. The user can also select cells based on any meta data features, or using a lasso tool. Following each organ and/or tissue manual annotation procedure, a data object containing the new annotations was generated using the same scVI parameters and the annotations were regularized to follow the cell ontology[157]. We did not correct for batch effects associated with organ even though each organ is sequenced separately because of concerns of removing biological variation by over-correction. Cell types missing in the current public version of the cell ontology were added to the provisional Tabula Sapiens cell ontology.

Since Tabula Sapiens was annotated by a large number of experts, quality control (QC) was performed on the manual annotations by using the automatic annotation tool popV[158]. PopV was applied to all organs in Tabula Sapiens donors 1 and 2 and predictability scores were generated for all cells by running a 5-fold cross validation. For donors 3 to 15 a draft automated annotation was generated using PopularVote. This was followed by manual inspection and annotation of all tissues in this set. This process was iterated in this manuscript by adding manual expert annotations for most organs from donors 17-25 before running popV on all the remaining unannotated cells and donors and cleaning up the labels manually as described above.

Quality control for contamination correction

To address potential barcode mixtures arising from sequencing experiments where different samples from each donor were loaded in the same 10X Genomics strips, we implemented a contamination correction filter. Specifically, we checked for overlapping barcodes across samples within the same donor. Cells sharing the same barcode identified in multiple samples from the same donor were considered potential contaminants and were excluded from further analysis. This process reduced the total number of cells from 1,150,192 to 1,105,354, enhancing the specificity of downstream analyses.

Transcription factor activity analysis

Transcription factor activity within each cell type was analyzed by running pySCENIC[159,160] (v0.12.1) on gene expression data from each cell type separately. The analysis was based on hg38-refseq_r80-mc_v10_clust-gene_based databases downloaded from resources.aertslab.org. Outputs from the ctx step of pySCENIC were used to infer downstream targets of transcription factors while outputs of the aucell step were used for transcription factor activity.

Transcription factor analysis across cell types using the τ statistic

τ analysis was done using only the droplet data for consistency across organs. This included 175 cell types of the 180 total cell types identified. The mean of the log normalized expression of each TF was computed for each cell type. Mean expression ranged from 0.0 to 4.9. The τ statistic (2,3) was used to generate a cell type specificity value for each of the 1635 transcription factors from these mean expression values.

$$\tau = \frac{\sum_{i=1}^{N} (1 - x_i)}{N - 1}$$

Every TF has a vector, x, of its mean expression in each cell type of the 175(N) cell types. TF vectors were normalized by their maximum expression value and then inverted by subtracting them from 1 to create a vector for each TF of values ranging from 0 to 1 with low values in cell types where it is highly expressed and values near 1 where it is lowly expressed. This vector is summed to make a single value for each TF. These sums are normalized to make τ range from 0 to 1. The Tspex[161] package, downloaded from https://apcamargo.github.io/tspex/, version-0.6.3 was used to perform this by calling "tspex.TissueSpecificity(TF_exp, 'τ', log = False)". The resulting τ values range between 0.4 and 1.0 where low τ values are ubiquitous transcription factors and high τ values are cell type specific. Heatmaps were made using the seaborn clustermap function in seaborn package (v-0.12.2). In the large heatmaps with all cell types, "vmax" was set to the 99.5 percentile, clipping the max value near 1.0 in order to improve visualization. Based on the distribution of τ values in the mouse atlas and a similar cutoff used in the fly cell atlas, we chose a τ threshold of 0.85 as the demarcation between specific (860 TF's) and non-specific (745 TF's). There was no clear transition in the τ distribution to distinguish ubiquitous from non-specific TF's. We attempted to find one by incorporating a lower bound for the mean expression and/or for the fraction of cells expressing the TF, but a principled cutoff criteria did not become clear. We then looked for protein expression of 100 lowest τ TF's in the HPA to get a different window on ubiquity.

Transcription factor Gene Ontology Analysis

Curated TF gene names were obtained from the "Collection of known and likely humanTFs (1639 proteins)" data set, downloaded from the University of Toronto transcription factor database. (https://humantfs.ccbr.utoronto.ca/allTFs.php.) 1639 of these gene names were found in our GTF file of which 1635 TF's showed expression in our dataset. SHOX and ZBED1 had zero expression. The 745 non-specific ($\tau < 0.85$) transcription factors were analyzed for gene set enrichment using the GSEApy package[162] (v-0.10.5) with GO_Biological_Process_2021 geneset against a background of all 1635 transcription factors. 69 GO terms had an adjusted P-val < 0.02. These were organized manually into broad cellular functions and related terms were placed into subgroups.

Analysis of the senescent cells phenotypes

Single-cell gene expression data from Tabula Sapiens 2.0 were used to identify senescent cells. Tissues from the sex with only a single donor were excluded from the analysis. After removing 52,149 cells from six tissues, the final dataset included 1.08 million cells from 25 tissues, collected from 21 different donors. Senescent cells were defined as CDKN2A+ MKI67- cells in the filtered dataset. A total of 48,114 senescent cells (~4.4% of all cells) were identified, spanning 145 cell types grouped into 34 broad categories. These cells were analyzed across different donors, tissues, and cell types. To identify senescence-associated genes (SAGs), differential gene expression analysis (DGEA) was performed by comparing senescent cells to non-senescent cells from the same cell type, donor, and tissue. To detect robust and statistically significant SAGs, we performed differential gene expression analysis (wilcoxon double-sided test) using the Scanpy package (v-1.10.1) with the following thresholds: log fold-change > 0.5, adjusted p-value < 0.001, and minimum percent of non-zero expression in senescent cell group = 0.5. SAGs from different donor-tissue-cell types were then grouped to select SAGs enriched in at least 50% of the donors across tissue-cell types. These enriched SAGs were combined into a final list of 3792 genes, which were enriched in at least 50% donors for at least one tissue and one broad cell type. To visualize the cell type prevalence of SAGs or senescence associated hallmark genes, we reperformed differential gene expression analysis at cell type-level for each of the 3,792 SAGs by comparing all senescent and non-senescent cells across broad cell types and calculated $\log_2$ fold changes for genes with statistically significant differences (adjusted p-value < 0.01) between senescent and non-senescent cells for a given cell type. We then used these cell type-level $\log_2$ fold changes as enrichment scores to assess the cell type prevalence for these SAGs. The number of cell types in which each SAG was enriched in senescent cells with $\log_2$ fold-change > 0.5 was considered to be the cell type prevalence for that SAG.

To identify coordinated transcriptional programs, expression data for 3792 senescence-associated genes (SAGs) were analyzed across senescent cells from multiple donors, tissues, and cell types. Mean senescence-associated transcriptomic profiles of senescent cells were generated and used for visualaization of senescent transcriptomes on a UMAP. Single-cell senescence-associated transcriptomic profiles of senescent cells (48114 cells x 3792 genes) were used for co-expression analysis using the cNMF consensus non-negative matrix factorization implemented in the cNMF package (v-1.5.4)[163]. To identify co-expressed gene modules, consensus non-negative matrix factorization (cNMF) was applied to the raw gene expression count matrix across a range of matrix ranks (K = 10 to 50), with 10 iterations per K. The optimal number of factors, K = 34, was determined by evaluating the trade-off between reconstruction error and stability across K values. All factorization outputs from the 10 iterations at K = 34 were aggregated and subjected to density filtering, yielding 34 robust gene modules. Each module comprised the top 100 genes most strongly associated with the corresponding factor.

For each cNMF factor, the top 100 genes with the highest weights were selected for gene ontology (GO) analysis using the GSEApy package[162] (v-1.1.3) and the GO_Biological_Processes_2023 database. For 100 genes in each cNMF factor, pathways with an adjusted p-value < 0.01 were retained. Among these, up to 10 pathways were selected per factor based on having at least five overlapping genes and ≤ 25% overlap with already selected pathways within the same list. A second round of filtering was performed to remove pathways with > 25% gene overlap across different gene lists to remove redundant pathways across factors. Pathway names, overlaps, GO-IDs, adjusted p-values, and associated genes per cNMF factors were then assembled. This analysis characterized the biological processes associated with each

SAG module. For the 34 cNMF factors, a total of 31 distinct pathways were significantly associated with known biological processes (adjusted p-value < 0.01). Gene set enrichment analysis (GSEA) was then used to quantify enrichment of senescent associated pathways across different broad cell types. First, cell type-level log fold changes between senescent and non-senescent cells were calculated for all broad cell types. Gene sets corresponding to the ontologies representing senescence-associated pathways were then compiled using Biomart in the GSEApy package. These gene sets were then filtered to contain genes identified within the 3792 SAGs. Pre-ranked GSEA was then performed with cell type-level log fold changes and the filtered senescence-associated gene sets to characterize the enrichment of senescence-associated gene sets across cell types. GSEA results were filtered for FDR < 0.05 to select 17 senescence-associated pathways that were significantly enriched in one of the cell types. Normalized enrichment scores and the lead genes within each pathway were visualized on a Heatmap.

Preprocessing for sex difference analysis

Smart-seq Data Exclusion. We filtered out cells associated with the Smart-seq protocol, reducing the overall cell numbers from 1,136,218 to 1,093,048..

Low statistical power exclusion. To ensure reliable comparison of sex differences across tissue and cell type reliably, we excluded tissue-cell type pairs that met either of the following criteria: (a) only one donor was represented, or (b) fewer than 100 cells were present in either the female or male group. This threshold was set to ensure sufficient statistical power for differential expression analyses.

PCA density filter. Outliers in the data were identified and removed using a Principal Component Analysis (PCA) density filter. Cells falling outside the high-density regions in the PCA space were considered outliers and excluded. This step ensured that extreme variations not representative of the underlying biological processes did not skew the results.

Anatomical Position Consideration: For each tissue and cell type, we checked for variability in anatomical position to avoid biases introduced by spatial heterogeneity across samples. If multiple anatomical positions were present for a given tissue-cell type, we checked for overlap between male and female donors. Only tissue-cell type pairs with shared anatomical positions between sexes were included in the analysis. If common anatomical positions were found, the data was further filtered to retain only samples from those positions. Additionally, a minimum of two male and two female donors was required to maintain sufficient statistical power. This step ensured that sex differences observed in gene expression were not confounded by anatomical variability across samples. The organ number for a rigor statistical power decreased from 28 to 13 organs.

Focusing on broad cell types for the sex-differential analysis, the number of tissue-cell type pairs decreased from 329 to 81 adequately represented groups as a result.

Pseudo-Bulk differential gene expression analysis for sex differences

To assess differential gene expression between male and female samples at the tissue-cell type level, we employed a pseudo-bulk approach with edgeR (v-4.0.1)[164,165], which has been recommended as an effective method to prevent false discoveries in datasets with covariates[166]. This method aggregates

single-cell data to create bulk-like profiles, enhancing statistical power while accounting for variability among donors and other covariates. For each tissue-cell type pair, we extracted the subset of cells corresponding to that specific tissue and cell type from the filtered dataset. A PCA density filter was applied to remove outliers by performing PCA on the scVI latent space representations and calculating a density estimate. Cells falling below the density cutoff were excluded to ensure that extreme variations did not affect the results. We then processed the data at the donor level. For each donor within the tissue-cell type subset, we included only those donors with at least 50 cells to ensure sufficient representation. For donors meeting this criterion, we randomly sampled 500 cells with replacement from the donor's cells to standardize the sample size across donors and reduce computational load. Clustering was performed on the donor-specific data using the Leiden algorithm with 30 nearest neighbors, 50 principal components and resolution 0.8, based on the scVI representations. This allowed us to capture intra-donor heterogeneity. For each cluster within the donor's data, we aggregated the raw UMI counts to create pseudo-bulk expression profiles. This resulted in multiple pseudo-bulk samples per donor, capturing intra-donor heterogeneity. Each pseudo-bulk sample was then annotated with metadata including donor ID, tissue type, cell type, age, sex, ethnicity, and a unique cluster ID combining sex, donor ID, and Leiden cluster label. Genes were filtered to retain those expressed in at least 10% of the cells within the dataset, ensuring that lowly expressed genes did not skew the analysis. Key genes of interest, such as sex chromosome-linked genes, were retained regardless of their expression levels. The pseudo-bulk samples from all donors for each tissue-cell type pair were concatenated to create a comprehensive dataset for differential expression analysis.

Differential expression between male and female pseudo-bulk samples was assessed using edgeR's generalized linear model (GLM) framework, which models count data using negative binomial distributions. The model included covariates such as donor age and batch effects where applicable to account for potential confounding factors. Genes with an adjusted p-value (Benjamini-Hochberg correction) less than 0.05 and an absolute $\log_2$ fold change greater than 0.25 were considered significantly differentially expressed.

Comparison between Tabula Sapiens and GTEx

To directly compare the differentially expressed genes identified from the Tabula Sapiens and the Genotype-Tissue Expression (GTEx) project, we performed a pseudo-bulk analysis of the Tabula Sapiens data at the tissue level. This approach is similar to the pseudo-bulk differential gene expression analysis described previously but focuses on aggregating gene expression data per tissue rather than per tissue-cell type pair.

For the GTEx data, we obtained publicly available results for differentially expressed (DE) genes between male and female samples across various tissues. These results were sourced directly from the GTEx project, which provides pre-processed differential expression analyses performed on bulk RNA-seq data from healthy human tissues. By using the publicly available DE gene lists from GTEx, we ensured that the comparison between Tabula Sapiens and GTEx was based on consistent and standardized analyses performed by the GTEx consortium. To compare the differentially expressed genes identified from Tabula Sapiens and GTEx, we matched genes based on their gene symbols and aligned tissues between the two datasets. Overlaps between the lists of differentially expressed genes were assessed using Venn diagrams.

Searchable extended metadata

Electronic Medical Records (EMRs) for 22 of 24 donors were obtained as PDF files from each hospital by Donor Network West, who manually de-identified each record. Our goal in developing ChatTS was to create a searchable database of extended donor metadata by feeding this information to an LLM to generate customized metadata tables for the user.

The EMRs had a mean±SD of 13±2 pages per donor, totaling 291 pages across all donors with EMRs available. Charts contained medical and social history taken at admission, and the full course of treatment following admission, until the donor was pronounced brain dead. This included positive and negative test results, values recorded for tests, other tabular and time-series information, as well as doctor notes. The initial size, scope, and format of the database was beyond current interpretability capabilities of LLMs, including the ability to understand tabular and time series data. Furthermore, their initial size was beyond context window limits, seemed to lead to sporadic information retrieval, and quickly overwhelmed API rate limits. Thus, to decrease complexity and size, we opted to create a database consisting of only free text present in the EMRs. Therefore, we manually extracted all free text from the EMRs, dropping tabular or time series data. To further decrease database size, we removed negative test results, and missing values (entered in the EMR as "NA" or "?"). The final size of this database is roughly 352KB and it contains a single .txt file per-donor.

ChatTS Application Architecture

A single .txt file is stored for each donor. A user is required to supply their own OpenAI API key for billing purposes, and then asks a question. The user can ask questions in two different modes, Data Retrieval and Chart Review. In either case, the input question is taken from the user and fed through an LLM in a first pass to transform the question into the form of asking for a single donor (for example, "Does anyone have a history of lupus" would be transformed to something like "Does this donor have a history of lupus"). Then, this question is added to an input prompt independently per-donor. Each input prompt contains 1) the user's transformed question, 2) a single donor's EMR as manually curated text, and 3) a set of instructions on how to answer the question. In Data Retrieval mode, this set of instructions is to simply return yes, no, or a single number. Meanwhile, in Chart Review mode, the instructions request the LLM to provide a detailed answer with supporting information from the EMR text. We also parallelize the API requests to decrease total processing time experienced by the user. In summary, when the user asks a question, 22 independent prompts are created (one per donor) and asked to the LLM in parallel. The results are then aggregated and added to a central metadata table that can be downloaded as a csv file. While we use OpenAI, we built the LLM integration flexibly and could use any LLM as backend.

ChatTS Application Server

The application is written in Python and HTML and hosted on an Amazon EC2 instance to allow flexibility in scaling to meet user demand. The application backend makes requests to GPT-4o through OpenAI's API. We have deposited the source code online at https://github.com/Harper-Hua/ChatTS.

References

1. The Tabula Sapiens Consortium (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science *376*, eabl4896. https://doi.org/10.1126/science.abl4896.

2. Domínguez Conde, C., Xu, C., Jarvis, L.B., Rainbow, D.B., Wells, S.B., Gomes, T., Howlett, S.K., Suchanek, O., Polanski, K., King, H.W., et al. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science *376*, eabl5197. https://doi.org/10.1126/science.abl5197.

3. Sikkema, L., Ramírez-Suástegui, C., Strobl, D.C., Gillett, T.E., Zappia, L., Madissoon, E., Markov, N.S., Zaragosi, L.-E., Ji, Y., Ansari, M., et al. (2023). An integrated cell atlas of the lung in health and disease. Nat. Med. *29*, 1563–1577. https://doi.org/10.1038/s41591-023-02327-2.

4. Kumar, T., Nee, K., Wei, R., He, S., Nguyen, Q.H., Bai, S., Blake, K., Pein, M., Gong, Y., Sei, E., et al. (2023). A spatially resolved single-cell genomic atlas of the adult human breast. Nature *620*, 181–191. https://doi.org/10.1038/s41586-023-06252-9.

5. Melms, J.C., Biermann, J., Huang, H., Wang, Y., Nair, A., Tagore, S., Katsyv, I., Rendeiro, A.F., Amin, A.D., Schapiro, D., et al. (2021). A molecular single-cell lung atlas of lethal COVID-19. Nature *595*, 114–119. https://doi.org/10.1038/s41586-021-03569-1.

6. Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*, eaah4573. https://doi.org/10.1126/science.aah4573.

7. Siletti, K., Hodge, R., Mossi Albiach, A., Lee, K.W., Ding, S.-L., Hu, L., Lönnerberg, P., Bakken, T., Casper, T., Clark, M., et al. (2023). Transcriptomic diversity of cell types across the adult human brain. Science *382*, eadd7046. https://doi.org/10.1126/science.add7046.

8. Aizarani, N., Saviano, A., Sagar, Mailly, L., Durand, S., Herman, J.S., Pessaux, P., Baumert, T.F., and Grün, D. (2019). A human liver cell atlas reveals heterogeneity and epithelial progenitors. Nature *572*, 199–204. https://doi.org/10.1038/s41586-019-1373-2.

9. Megill, C., Martin, B., Weaver, C., Bell, S., Prins, L., Badajoz, S., McCandless, B., Pisco, A.O., Kinsella, M., Griffin, F., et al. (2021). cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices. Preprint at bioRxiv, https://doi.org/10.1101/2021.04.05.438318 https://doi.org/10.1101/2021.04.05.438318.

10. Program, C.S.-C.B., Abdulla, S., Aevermann, B., Assis, P., Badajoz, S., Bell, S.M., Bezzi, E., Cakir, B., Chaffer, J., Chambers, S., et al. (2023). CZ CELL×GENE Discover: A single-cell data platform for scalable exploration, analysis and modeling of aggregated data. Preprint at bioRxiv, https://doi.org/10.1101/2023.10.30.563174 https://doi.org/10.1101/2023.10.30.563174.

11. The ENCODE (ENCyclopedia Of DNA Elements) Project (2004). Science *306*, 636–640. https://doi.org/10.1126/science.1105136.

12. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048. https://doi.org/10.1038/nbt1010-1045.

13. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.

14. Joung, J., Ma, S., Tay, T., Geiger-Schuller, K.R., Kirchgatterer, P.C., Verdine, V.K., Guo, B., Arias-Garcia, M.A., Allen, W.E., Singh, A., et al. (2023). A transcription factor atlas of directed differentiation. Cell *186*, 209-229.e26. https://doi.org/10.1016/j.cell.2022.11.026.

15. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665. https://doi.org/10.1016/j.cell.2018.01.029.

16. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419. https://doi.org/10.1126/science.1260419.

17. Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Björk, L., Breckels, L.M., et al. (2017). A subcellular map of the human proteome. Science *356*, eaal3321. https://doi.org/10.1126/science.aal3321.

18. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single-cell type transcriptomics map of human tissues. Sci. Adv. *7*, eabh2169. https://doi.org/10.1126/sciadv.abh2169.

19. Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinforma. Oxf. Engl. *21*, 650–659. https://doi.org/10.1093/bioinformatics/bti042.

20. Kryuchkova-Mostacci, N., and Robinson-Rechavi, M. (2017). A benchmark of gene expression tissue-specificity metrics. Brief. Bioinform. *18*, 205–214. https://doi.org/10.1093/bib/bbw008.

21. Li, H., Janssens, J., De Waegeneer, M., Kolluru, S.S., Davie, K., Gardeux, V., Saelens, W., David, F., Brbić, M., Spanier, K., et al. (2022). Fly Cell Atlas: a single-nucleus transcriptomic atlas of the adult fruit fly. Science *375*, eabk2432. https://doi.org/10.1126/science.abk2432.

22. Yao, Z., van Velthoven, C.T.J., Kunst, M., Zhang, M., McMillen, D., Lee, C., Jung, W., Goldy, J., Abdelhak, A., Baker, P., et al. (2023). A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. bioRxiv, 2023.03.06.531121. https://doi.org/10.1101/2023.03.06.531121.

23. Trimarchi, J.M., Fairchild, B., Wen, J., and Lees, J.A. (2001). The E2F6 transcription factor is a component of the mammalian Bmi1-containing polycomb complex. Proc. Natl. Acad. Sci. U. S. A. *98*, 1519–1524. https://doi.org/10.1073/pnas.98.4.1519.

24. Dyson, N. (1998). The regulation of E2F by pRB-family proteins. Genes Dev. *12*, 2245–2262. https://doi.org/10.1101/gad.12.15.2245.

25. Gaubatz, S., Wood, J.G., and Livingston, D.M. (1998). Unusual proliferation arrest and transcriptional control properties of a newly discovered E2F family member, E2F-6. Proc. Natl. Acad. Sci. U. S. A. *95*, 9190–9195.

26. Bobola, N., and Sagerström, C.G. (2024). TALE transcription factors: Cofactors no more. Semin. Cell Dev. Biol. *152–153*, 76–84. https://doi.org/10.1016/j.semcdb.2022.11.015.

27. Currey, L., Thor, S., and Piper, M. (2021). TEAD family transcription factors in development and disease. Development *148*, dev196675. https://doi.org/10.1242/dev.196675.

28. Sasaki, K., Sakamoto, M., Miyake, I., Tanaka, R., Tanaka, R., Tanaka, A., Terami, M., Komori, R., Taniguchi, M., Wakabayashi, S., et al. (2023). Expression of transcription factors KLF2 and KLF4 is induced by the mammalian Golgi stress response. Preprint at bioRxiv, https://doi.org/10.1101/2023.05.16.541051 https://doi.org/10.1101/2023.05.16.541051.

29. Hua, X., Yokoyama, C., Wu, J., Briggs, M.R., Brown, M.S., Goldstein, J.L., and Wang, X. (1993). SREBP-2, a second basic-helix-loop-helix-leucine zipper protein that stimulates transcription by binding to a sterol regulatory element. Proc. Natl. Acad. Sci. U. S. A. *90*, 11603–11607.

30. Itoh, K., Chiba, T., Takahashi, S., Ishii, T., Igarashi, K., Katoh, Y., Oyake, T., Hayashi, N., Satoh, K., Hatayama, I., et al. (1997). An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. Biochem. Biophys. Res. Commun. *236*, 313–322. https://doi.org/10.1006/bbrc.1997.6943.

31. Zhang, J., Ohta, T., Maruyama, A., Hosoya, T., Nishikawa, K., Maher, J.M., Shibahara, S., Itoh, K., and Yamamoto, M. (2006). BRG1 Interacts with Nrf2 To Selectively Mediate HO-1 Induction in Response to Oxidative Stress. Mol. Cell. Biol. *26*, 7942–7952. https://doi.org/10.1128/MCB.00700-06.

32. Bae, T., Hallis, S.P., and Kwak, M.-K. (2024). Hypoxia, oxidative stress, and the interplay of HIFs and NRF2 signaling in cancer. Exp. Mol. Med. *56*, 501–514. https://doi.org/10.1038/s12276-024-01180-8.

33. Piskacek, M., Havelka, M., Jendruchova, K., and Knight, A. (2019). Nuclear hormone receptors: Ancient 9aaTAD and evolutionally gained NCoA activation pathways. J. Steroid Biochem. Mol. Biol. *187*, 118–123. https://doi.org/10.1016/j.jsbmb.2018.11.008.

34. Rudensky, A.Y. (2011). Regulatory T cells and Foxp3. Immunol. Rev. *241*, 260–268. https://doi.org/10.1111/j.1600-065X.2011.01018.x.

35. Mouly, E., Chemin, K., Nguyen, H.V., Chopin, M., Mesnard, L., Leite-de-Moraes, M., Burlen-defranoux, O., Bandeira, A., and Bories, J.-C. (2010). The Ets-1 transcription factor controls the development and function of natural regulatory T cells. J. Exp. Med. *207*, 2113–2125. https://doi.org/10.1084/jem.20092153.

36. Llaó-Cid, L., Roessner, P.M., Chapaprieta, V., Öztürk, S., Roider, T., Bordas, M., Izcue, A., Colomer, D., Dietrich, S., Stilgenbauer, S., et al. (2021). EOMES is essential for antitumor activity of CD8+ T cells in chronic lymphocytic leukemia. Leukemia *35*, 3152–3162. https://doi.org/10.1038/s41375-021-01198-1.

37. Bernardi, C., Maurer, G., Ye, T., Marchal, P., Jost, B., Wissler, M., Maurer, U., Kastner, P., Chan, S., and Charvet, C. (2021). CD4+ T cells require Ikaros to inhibit their differentiation toward a pathogenic cell fate. Proc. Natl. Acad. Sci. U. S. A. *118*, e2023172118. https://doi.org/10.1073/pnas.2023172118.

38. Fan, Y., and Lu, D. (2016). The Ikaros family of zinc-finger proteins. Acta Pharm. Sin. B *6*, 513–521. https://doi.org/10.1016/j.apsb.2016.06.002.

39. Parker, J.B., Valencia, C., Akras, D., DiIorio, S.E., Griffin, M.F., Longaker, M.T., and Wan, D.C. (2023). Understanding Fibroblast Heterogeneity in Form and Function. Biomedicines *11*, 2264. https://doi.org/10.3390/biomedicines11082264.

40. LeBleu, V.S., and Neilson, E.G. (2020). Origin and functional heterogeneity of fibroblasts. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. *34*, 3519–3536. https://doi.org/10.1096/fj.201903188R.

41. Muhl, L., Genové, G., Leptidis, S., Liu, J., He, L., Mocci, G., Sun, Y., Gustafsson, S., Buyandelger, B., Chivukula, I.V., et al. (2020). Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. Nat. Commun. *11*, 3953. https://doi.org/10.1038/s41467-020-17740-1.

42. Hu, M.S., Moore, A.L., and Longaker, M.T. (2018). A Fibroblast Is Not a Fibroblast Is Not a Fibroblast. J. Invest. Dermatol. *138*, 729–730. https://doi.org/10.1016/j.jid.2017.10.012.

43. Park, J., Ivey, M.J., Deana, Y., Riggsbee, K.L., Sörensen, E., Schwabl, V., Sjöberg, C., Hjertberg, T., Park, G.Y., Swonger, J.M., et al. (2019). The Tcf21 lineage constitutes the lung lipofibroblast population. Am. J. Physiol. Lung Cell. Mol. Physiol. *316*, L872–L885. https://doi.org/10.1152/ajplung.00254.2018.

44. Shen, T., Aneas, I., Sakabe, N., Dirschinger, R.J., Wang, G., Smemo, S., Westlund, J.M., Cheng, H., Dalton, N., Gu, Y., et al. (2011). Tbx20 regulates a genetic program essential to adult mouse cardiomyocyte function. J. Clin. Invest. *121*, 4640–4654. https://doi.org/10.1172/JCI59472.

45. Kidani, Y., Elsaesser, H., Hock, M.B., Vergnes, L., Williams, K.J., Argus, J.P., Marbois, B.N., Komisopoulou, E., Wilson, E.B., Osborne, T.F., et al. (2013). Sterol regulatory element-binding proteins are essential for the metabolic programming of effector T cells and adaptive immunity. Nat. Immunol. *14*, 489–499. https://doi.org/10.1038/ni.2570.

46. Thottappillil, N., Gomez-Salazar, M.A., Xu, M., Qin, Q., Xing, X., Xu, J., Broderick, K., Yea, J.-H., Archer, M., Ching-Yun Hsu, G., et al. (2023). ZIC1 Dictates Osteogenesis Versus Adipogenesis in Human Mesenchymal Progenitor Cells Via a Hedgehog Dependent Mechanism. Stem Cells Dayt. Ohio *41*, 862–876. https://doi.org/10.1093/stmcls/sxad047.

47. Himeda, C.L., Barro, M.V., and Emerson, C.P. (2013). Pax3 synergizes with Gli2 and Zic1 in transactivating the Myf5 epaxial somite enhancer. Dev. Biol. *383*, 7–14. https://doi.org/10.1016/j.ydbio.2013.09.006.

48. Perugini, J., Bordoni, L., Venema, W., Acciarini, S., Cinti, S., Gabbianelli, R., and Giordano, A. (2019). Zic1 mRNA is transiently upregulated in subcutaneous fat of acutely cold-exposed mice. J. Cell. Physiol. *234*, 2031–2036. https://doi.org/10.1002/jcp.27301.

49. Grainger, S., Hryniuk, A., and Lohnes, D. (2013). Cdx1 and Cdx2 Exhibit Transcriptional Specificity in the Intestine. PLoS ONE *8*, e54757. https://doi.org/10.1371/journal.pone.0054757.

50. Van Der Flier, L.G., Van Gijn, M.E., Hatzis, P., Kujala, P., Haegebarth, A., Stange, D.E., Begthel, H., Van Den Born, M., Guryev, V., Oving, I., et al. (2009). Transcription Factor Achaete Scute-Like 2 Controls Intestinal Stem Cell Fate. Cell *136*, 903–912. https://doi.org/10.1016/j.cell.2009.01.031.

51. Luna Velez, M.V., Neikes, H.K., Snabel, R.R., Quint, Y., Qian, C., Martens, A., Veenstra, G.J.C., Freeman, M.R., van Heeringen, S.J., and Vermeulen, M. (2023). ONECUT2 regulates RANKL-dependent enterocyte and microfold cell differentiation in the small intestine; a multi-omics study. Nucleic Acids Res. *51*, 1277–1296. https://doi.org/10.1093/nar/gkac1236.

52. Pattabiraman, D.R., and Gonda, T.J. (2013). Role and potential for therapeutic targeting of MYB in leukemia. Leukemia *27*, 269–277. https://doi.org/10.1038/leu.2012.225.

53. Lawrence, H.J., Christensen, J., Fong, S., Hu, Y.-L., Weissman, I., Sauvageau, G., Humphries, R.K., and Largman, C. (2005). Loss of expression of the Hoxa-9 homeobox gene impairs the proliferation and repopulating ability of hematopoietic stem cells. Blood *106*, 3988–3994. https://doi.org/10.1182/blood-2005-05-2003.

54. Ramos-Mejía, V., Navarro-Montero, O., Ayllón, V., Bueno, C., Romero, T., Real, P.J., and Menendez, P. (2014). HOXA9 promotes hematopoietic commitment of human embryonic stem cells. Blood *124*, 3065–3075. https://doi.org/10.1182/blood-2014-03-558825.

55. Wang, D., Tanaka-Yano, M., Meader, E., Kinney, M.A., Morris, V., Lummertz da Rocha, E., Liu, N., Liu, T., Zhu, Q., Orkin, S.H., et al. (2022). Developmental maturation of the hematopoietic system controlled by a Lin28b-let-7-Cbx2 axis. Cell Rep. *39*, 110587. https://doi.org/10.1016/j.celrep.2022.110587.

56. Siatecka, M., and Bieker, J.J. (2011). The multifunctional role of EKLF/KLF1 during erythropoiesis. Blood *118*, 2044–2054. https://doi.org/10.1182/blood-2011-03-331371.

57. Shimizu, R., and Yamamoto, M. (2023). Recent progress in analyses of GATA1 in hematopoietic disorders: a mini-review. Front. Hematol. *2*, 1181216. https://doi.org/10.3389/frhem.2023.1181216.

58. Fischer, A., Schumacher, N., Maier, M., Sendtner, M., and Gessler, M. (2004). The Notch target genes Hey1 and Hey2 are required for embryonic vascular development. Genes Dev. *18*, 901–911. https://doi.org/10.1101/gad.291004.

59. Kim, D.W., Liu, K., Wang, Z.Q., Zhang, Y.S., Bathini, A., Brown, M.P., Lin, S.H., Washington, P.W., Sun, C., Lindtner, S., et al. (2021). Gene regulatory networks controlling differentiation, survival, and diversification of hypothalamic Lhx6-expressing GABAergic neurons. Commun. Biol. *4*, 95. https://doi.org/10.1038/s42003-020-01616-7.

60. Norrmén, C., Ivanov, K.I., Cheng, J., Zangger, N., Delorenzi, M., Jaquet, M., Miura, N., Puolakkainen, P., Horsley, V., Hu, J., et al. (2009). FOXC2 controls formation and maturation of lymphatic collecting vessels through cooperation with NFATc1. J. Cell Biol. *185*, 439–457. https://doi.org/10.1083/jcb.200901104.

61. Birdsey, G.M., Shah, A.V., Dufton, N., Reynolds, L.E., Osuna Almagro, L., Yang, Y., Aspalter, I.M., Khan, S.T., Mason, J.C., Dejana, E., et al. (2015). The Endothelial Transcription Factor ERG Promotes Vascular Stability and Growth through Wnt/β-Catenin Signaling. Dev. Cell *32*, 82–96. https://doi.org/10.1016/j.devcel.2014.11.016.

62. Sperone, A., Dryden, N.H., Birdsey, G.M., Madden, L., Johns, M., Evans, P.C., Mason, J.C., Haskard, D.O., Boyle, J.J., Paleolog, E.M., et al. (2011). The transcription factor Erg inhibits vascular inflammation by repressing NF-kappaB activation and proinflammatory gene expression in endothelial cells. Arterioscler. Thromb. Vasc. Biol. *31*, 142–150.

https://doi.org/10.1161/ATVBAHA.110.216473.

63. Nagai, N., Ohguchi, H., Nakaki, R., Matsumura, Y., Kanki, Y., Sakai, J., Aburatani, H., and Minami, T. (2018). Downregulation of ERG and FLI1 expression in endothelial cells triggers endothelial-to-mesenchymal transition. PLoS Genet. *14*, e1007826. https://doi.org/10.1371/journal.pgen.1007826.

64. Heo, S.-H., and Cho, J.-Y. (2014). ELK3 suppresses angiogenesis by inhibiting the transcriptional activity of ETS-1 on MT1-MMP. Int. J. Biol. Sci. *10*, 438–447. https://doi.org/10.7150/ijbs.8095.

65. Tian, H., McKnight, S.L., and Russell, D.W. (1997). Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. Genes Dev. *11*, 72–82. https://doi.org/10.1101/gad.11.1.72.

66. Hanaoka, M., Droma, Y., Basnyat, B., Ito, M., Kobayashi, N., Katsuyama, Y., Kubo, K., and Ota, M. (2012). Genetic variants in EPAS1 contribute to adaptation to high-altitude hypoxia in Sherpas. PloS One *7*, e50566. https://doi.org/10.1371/journal.pone.0050566.

67. Bettegowda, A., and Wilkinson, M.F. (2010). Transcription and post-transcriptional regulation of spermatogenesis. Philos. Trans. R. Soc. B Biol. Sci. *365*, 1637–1651. https://doi.org/10.1098/rstb.2009.0196.

68. Lovelace, D.L., Gao, Z., Mutoji, K., Song, Y.C., Ruan, J., and Hermann, B.P. (2016). The regulatory repertoire of PLZF and SALL4 in undifferentiated spermatogonia. Dev. Camb. Engl. *143*, 1893–1906. https://doi.org/10.1242/dev.132761.

69. Gassei, K., and Orwig, K.E. (2013). SALL4 expression in gonocytes and spermatogonial clones of postnatal mouse testes. PloS One *8*, e53976. https://doi.org/10.1371/journal.pone.0053976.

70. Zarkower, D. (2013). DMRT genes in vertebrate gametogenesis. Curr. Top. Dev. Biol. *102*, 327–356. https://doi.org/10.1016/B978-0-12-416024-8.00012-X.

71. Zhang, T., and Zarkower, D. (2017). DMRT proteins and coordination of mammalian spermatogenesis. Stem Cell Res. *24*, 195–202. https://doi.org/10.1016/j.scr.2017.07.026.

72. Zhou, D., Fan, J., Liu, Z., Tang, R., Wang, X., Bo, H., Zhu, F., Zhao, X., Huang, Z., Xing, L., et al. (2021). TCF3 Regulates the Proliferation and Apoptosis of Human Spermatogonial Stem Cells by Targeting PODXL. Front. Cell Dev. Biol. *9*, 695545. https://doi.org/10.3389/fcell.2021.695545.

73. Säflund, M., and Özata, D.M. (2023). The MYBL1/TCFL5 transcription network: two collaborative factors with central role in male meiosis. Biochem. Soc. Trans. *51*, 2163–2172. https://doi.org/10.1042/BST20231007.

74. Bai, S., Fu, K., Yin, H., Cui, Y., Yue, Q., Li, W., Cheng, L., Tan, H., Liu, X., Guo, Y., et al. (2018). Sox30 initiates transcription of haploid genes during late meiosis and spermiogenesis in mouse testes. Dev. Camb. Engl. *145*, dev164855. https://doi.org/10.1242/dev.164855.

75. Kuilman, T., Michaloglou, C., Mooi, W.J., and Peeper, D.S. (2010). The essence of senescence. Genes Dev. *24*, 2463–2479. https://doi.org/10.1101/gad.1971610.

76. Hayflick, L., and Moorhead, P.S. (1961). The serial cultivation of human diploid cell strains. Exp.

Cell Res. *25*, 585–621. https://doi.org/10.1016/0014-4827(61)90192-6.

77. Beauséjour, C.M., Krtolica, A., Galimi, F., Narita, M., Lowe, S.W., Yaswen, P., and Campisi, J. (2003). Reversal of human cellular senescence: roles of the p53 and p16 pathways. EMBO J. *22*, 4212–4222. https://doi.org/10.1093/emboj/cdg417.

78. Liu, J.-Y., Souroullas, G.P., Diekman, B.O., Krishnamurthy, J., Hall, B.M., Sorrentino, J.A., Parker, J.S., Sessions, G.A., Gudkov, A.V., and Sharpless, N.E. (2019). Cells exhibiting strong *p16* ^{INK4a} promoter activation in vivo display features of senescence. Proc. Natl. Acad. Sci. *116*, 2603–2611. https://doi.org/10.1073/pnas.1818313116.

79. Dodig, S., Čepelak, I., and Pavić, I. (2019). Hallmarks of senescence and aging. Biochem. Medica *29*, 030501. https://doi.org/10.11613/BM.2019.030501.

80. Acosta, J.C., Banito, A., Wuestefeld, T., Georgilis, A., Janich, P., Morton, J.P., Athineos, D., Kang, T.-W., Lasitschka, F., Andrulis, M., et al. (2013). A complex secretory program orchestrated by the inflammasome controls paracrine senescence. Nat. Cell Biol. *15*, 978–990. https://doi.org/10.1038/ncb2784.

81. Hernandez-Segura, A., Nehme, J., and Demaria, M. (2018). Hallmarks of Cellular Senescence. Trends Cell Biol. *28*, 436–453. https://doi.org/10.1016/j.tcb.2018.02.001.

82. Wiley, C.D., Velarde, M.C., Lecot, P., Liu, S., Sarnoski, E.A., Freund, A., Shirakawa, K., Lim, H.W., Davis, S.S., Ramanathan, A., et al. (2016). Mitochondrial Dysfunction Induces Senescence with a Distinct Secretory Phenotype. Cell Metab. *23*, 303–314. https://doi.org/10.1016/j.cmet.2015.11.011.

83. Sanborn, M.A., Wang, X., Gao, S., Dai, Y., and Rehman, J. (2023). SenePy: Unveiling the Cell-Type Specific Landscape of Cellular Senescence through Single-Cell Analysis in Living Organisms. Preprint at bioRxiv, https://doi.org/10.1101/2023.08.30.555644 https://doi.org/10.1101/2023.08.30.555644.

84. Tao, W., Yu, Z., and Han, J.-D.J. (2024). Single-cell senescence identification reveals senescence heterogeneity, trajectory, and modulators. Cell Metab. *36*, 1126-1143.e5. https://doi.org/10.1016/j.cmet.2024.03.009.

85. Neri, F., Zheng, S., Watson, M., Desprez, P.-Y., Gerencser, A.A., Campisi, J., Wirtz, D., Wu, P.-H., and Schilling, B. (2024). Senescent cell heterogeneity and responses to senolytic treatment are related to cell cycle status during cell growth arrest. Preprint at bioRxiv, https://doi.org/10.1101/2024.06.22.600200 https://doi.org/10.1101/2024.06.22.600200.

86. Coppé, J.-P., Patil, C.K., Rodier, F., Sun, Y., Muñoz, D.P., Goldstein, J., Nelson, P.S., Desprez, P.-Y., and Campisi, J. (2008). Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. PLoS Biol. *6*, 2853–2868. https://doi.org/10.1371/journal.pbio.0060301.

87. Basisty, N., Kale, A., Jeon, O.H., Kuehnemann, C., Payne, T., Rao, C., Holtz, A., Shah, S., Sharma, V., Ferrucci, L., et al. (2020). A proteomic atlas of senescence-associated secretomes for aging biomarker development. PLoS Biol. *18*, e3000599. https://doi.org/10.1371/journal.pbio.3000599.

88. van Deursen, J.M. (2014). The role of senescent cells in ageing. Nature *509*, 439–446.

https://doi.org/10.1038/nature13193.

89. Baker, D.J., Wijshake, T., Tchkonia, T., LeBrasseur, N.K., Childs, B.G., van de Sluis, B., Kirkland, J.L., and van Deursen, J.M. (2011). Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. Nature *479*, 232–236. https://doi.org/10.1038/nature10600.

90. Gasek, N.S., Kuchel, G.A., Kirkland, J.L., and Xu, M. (2021). Strategies for Targeting Senescent Cells in Human Disease. Nat. Aging *1*, 870–879. https://doi.org/10.1038/s43587-021-00121-8.

91. Ogrodnik, M., Acosta, J.C., Adams, P.D., Fagagna, F. d'Adda di, Baker, D.J., Bishop, C.L., Chandra, T., Collado, M., Gil, J., Gorgoulis, V., et al. (2024). Guidelines for minimal information on cellular senescence experimentation in vivo. Cell *187*, 4150–4175. https://doi.org/10.1016/j.cell.2024.05.059.

92. Biran, A., Zada, L., Abou Karam, P., Vadai, E., Roitman, L., Ovadya, Y., Porat, Z., and Krizhanovsky, V. (2017). Quantitative identification of senescent cells in aging and disease. Aging Cell *16*, 661–671. https://doi.org/10.1111/acel.12592.

93. Yousefzadeh, M.J., Zhao, J., Bukata, C., Wade, E.A., McGowan, S.J., Angelini, L.A., Bank, M.P., Gurkar, A.U., McGuckian, C.A., Calubag, M.F., et al. (2020). Tissue specificity of senescent cell accumulation during physiologic and accelerated aging of mice. Aging Cell *19*, e13094. https://doi.org/10.1111/acel.13094.

94. Pham, T.-H., Park, H.-M., Kim, J., Hong, J.-T., and Yoon, D.-Y. (2021). Interleukin-32θ Triggers Cellular Senescence and Reduces Sensitivity to Doxorubicin-Mediated Cytotoxicity in MDA-MB-231 Cells. Int. J. Mol. Sci. *22*, 4974. https://doi.org/10.3390/ijms22094974.

95. Jin, M., Xu, R., Wang, L., Alam, M.M., Ma, Z., Zhu, S., Martini, A.C., Jadali, A., Bernabucci, M., Xie, P., et al. (2022). Type-I-interferon signaling drives microglial dysfunction and senescence in human iPSC models of Down syndrome and Alzheimer's disease. Cell Stem Cell *29*, 1135-1153.e8. https://doi.org/10.1016/j.stem.2022.06.007.

96. Kandhaya-Pillai, R., Miro-Mur, F., Alijotas-Reig, J., Tchkonia, T., Kirkland, J.L., and Schwartz, S. (2017). TNFα-senescence initiates a STAT-dependent positive feedback loop, leading to a sustained interferon signature, DNA damage, and cytokine secretion. Aging *9*, 2411–2435. https://doi.org/10.18632/aging.101328.

97. Tomimatsu, N., Di Cristofaro, L.F.M., Kanji, S., Samentar, L., Jordan, B.R., Kittler, R., Habib, A.A., Espindola-Netto, J.M., Tchkonia, T., Kirkland, J.L., et al. (2025). Targeting cIAP2 in a novel senolytic strategy prevents glioblastoma recurrence after radiotherapy. EMBO Mol. Med. *17*, 645–678. https://doi.org/10.1038/s44321-025-00201-x.

98. Shin, E.-Y., Park, J.-H., You, S.-T., Lee, C.-S., Won, S.-Y., Park, J.-J., Kim, H.-B., Shim, J., Soung, N.-K., Lee, O.-J., et al. (2020). Integrin-mediated adhesions in regulation of cellular senescence. Sci. Adv. *6*, eaay3909. https://doi.org/10.1126/sciadv.aay3909.

99. Mun, G.I., and Boo, Y.C. (2010). Identification of CD44 as a senescence-induced cell adhesion gene responsible for the enhanced monocyte recruitment to senescent endothelial cells. Am. J. Physiol.-Heart Circ. Physiol. *298*, H2102–H2111. https://doi.org/10.1152/ajpheart.00835.2009.

100. Kumari, R., and Jat, P. (2021). Mechanisms of Cellular Senescence: Cell Cycle Arrest and

Senescence Associated Secretory Phenotype. Front. Cell Dev. Biol. *9*, 645593. https://doi.org/10.3389/fcell.2021.645593.

101. Deschênes-Simard, X., Lessard, F., Gaumont-Leclerc, M.-F., Bardeesy, N., and Ferbeyre, G. (2014). Cellular senescence and protein degradation: Breaking down cancer. Cell Cycle *13*, 1840–1858. https://doi.org/10.4161/cc.29335.

102. Hamazaki, J., and Murata, S. (2024). Relationships between protein degradation, cellular senescence, and organismal aging. J. Biochem. (Tokyo) *175*, 473–480. https://doi.org/10.1093/jb/mvae016.

103. Chapman, J., Fielder, E., and Passos, J.F. (2019). Mitochondrial dysfunction and cell senescence: deciphering a complex relationship. FEBS Lett. *593*, 1566–1579. https://doi.org/10.1002/1873-3468.13498.

104. Martini, H., and Passos, J.F. (2023). Cellular senescence: all roads lead to mitochondria. FEBS J. *290*, 1186–1202. https://doi.org/10.1111/febs.16361.

105. Kim, S.-J., Mehta, H.H., Wan, J., Kuehnemann, C., Chen, J., Hu, J.-F., Hoffman, A.R., and Cohen, P. (2018). Mitochondrial peptides modulate mitochondrial function during cellular senescence. Aging *10*, 1239–1256. https://doi.org/10.18632/aging.101463.

106. Hutter, E., Renner, K., Pfister, G., Stöckl, P., Jansen-Dürr, P., and Gnaiger, E. (2004). Senescence-associated changes in respiration and oxidative phosphorylation in primary human fibroblasts. Biochem. J. *380*, 919–928. https://doi.org/10.1042/bj20040095.

107. Xu, H., Wang, H., Ning, X., Xu, Z., and Zhang, G. (2024). Integrated bioinformatics and validation reveal PTGS2 and its related molecules to alleviate TNF-α-induced endothelial senescence. Vitro Cell. Dev. Biol. - Anim. *60*, 888–902. https://doi.org/10.1007/s11626-024-00931-1.

108. Tan, J.X., and Finkel, T. (2023). Lysosomes in senescence and aging. EMBO Rep. *24*, e57265. https://doi.org/10.15252/embr.202357265.

109. Lipinska, A., Cormier, A., Luthringer, R., Peters, A.F., Corre, E., Gachon, C.M.M., Cock, J.M., and Coelho, S.M. (2015). Sexual Dimorphism and the Evolution of Sex-Biased Gene Expression in the Brown Alga Ectocarpus. Mol. Biol. Evol. *32*, 1581–1597. https://doi.org/10.1093/molbev/msv049.

110. Rinn, J.L., and Snyder, M. (2005). Sexual dimorphism in mammalian gene expression. Trends Genet. *21*, 298–305. https://doi.org/10.1016/j.tig.2005.03.005.

111. Naurin, S., Hansson, B., Hasselquist, D., Kim, Y.-H., and Bensch, S. (2011). The sex-biased brain: sexual dimorphism in gene expression in two species of songbirds. BMC Genomics *12*, 37. https://doi.org/10.1186/1471-2164-12-37.

112. Dewing, P., Shi, T., Horvath, S., and Vilain, E. (2003). Sexually dimorphic gene expression in mouse brain precedes gonadal differentiation. Mol. Brain Res. *118*, 82–90. https://doi.org/10.1016/S0169-328X(03)00339-5.

113. Williams, T.M., and Carroll, S.B. (2009). Genetic and molecular insights into the development and evolution of sexual dimorphism. Nat. Rev. Genet. *10*, 797–804. https://doi.org/10.1038/nrg2687.

114. Ober, C., Loisel, D.A., and Gilad, Y. (2008). Sex-specific genetic architecture of human disease. Nat. Rev. Genet. *9*, 911–922. https://doi.org/10.1038/nrg2415.

115. Chi, L., Liu, C., Gribonika, I., Gschwend, J., Corral, D., Han, S.-J., Lim, A.I., Rivera, C.A., Link, V.M., Wells, A.C., et al. (2024). Sexual dimorphism in skin immunity is mediated by an androgen-ILC2-dendritic cell axis. Science *384*, eadk6200. https://doi.org/10.1126/science.adk6200.

116. Klein, S.L., and Flanagan, K.L. (2016). Sex differences in immune responses. Nat. Rev. Immunol. *16*, 626–638. https://doi.org/10.1038/nri.2016.90.

117. Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A.D.H., Cotter, D.J., Parsana, P., Kasela, S., Balliu, B., Viñuela, A., et al. (2020). The impact of sex on gene expression across human tissues. Science *369*, eaba3066. https://doi.org/10.1126/science.aba3066.

118. Puttur, F., and Lloyd, C.M. (2024). Sex differences in tissue immunity. Science *384*, 159–160. https://doi.org/10.1126/science.ado8542.

119. Arnold, A.P. (2017). A general theory of sexual differentiation. J. Neurosci. Res. *95*, 291–300. https://doi.org/10.1002/jnr.23884.

120. Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the Mechanism of X Chromosome Inactivation. Annu. Rev. Genet. *36*, 233–278. https://doi.org/10.1146/annurev.genet.36.042902.092433.

121. Swain, A., Narvaez, V., Burgoyne, P., Camerino, G., and Lovell-Badge, R. (1998). Dax1 antagonizes Sry action in mammalian sex determination. Nature *391*, 761–767. https://doi.org/10.1038/35799.

122. Kaneko, S., and Li, X. (2018). X chromosome protects against bladder cancer in females via a KDM6A-dependent epigenetic mechanism. Sci. Adv. *4*, eaar5598. https://doi.org/10.1126/sciadv.aar5598.

123. Lau, Y.-F.C. (1999). Gonadoblastoma, Testicular and Prostate Cancers, and the TSPY Gene. Am. J. Hum. Genet. *64*, 921–927. https://doi.org/10.1086/302353.

124. Waxman, D.J., and Holloway, M.G. (2009). Sex Differences in the Expression of Hepatic Drug Metabolizing Enzymes. Mol. Pharmacol. *76*, 215–228. https://doi.org/10.1124/mol.109.056705.

125. Obradovic, M., Sudar-Milovanovic, E., Soskic, S., Essack, M., Arya, S., Stewart, A.J., Gojobori, T., and Isenovic, E.R. (2021). Leptin and Obesity: Role and Clinical Implication. Front. Endocrinol. *12*. https://doi.org/10.3389/fendo.2021.585887.

126. Ferré, P. (2004). The Biology of Peroxisome Proliferator-Activated Receptors: Relationship With Lipid Metabolism and Insulin Sensitivity. Diabetes *53*, S43–S50. https://doi.org/10.2337/diabetes.53.2007.S43.

127. Zhang, M.A., Rego, D., Moshkova, M., Kebir, H., Chruscinski, A., Nguyen, H., Akkermann, R., Stanczyk, F.Z., Prat, A., Steinman, L., et al. (2012). Peroxisome proliferator-activated receptor (PPAR)α and -γ regulate IFNγ and IL-17A production by human T cells in a sex-specific way. Proc. Natl. Acad. Sci. *109*, 9505–9510. https://doi.org/10.1073/pnas.1118458109.

128. Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M.A., Marshall, J.L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., et al. (2017). Landscape of X chromosome inactivation across human tissues. Nature *550*, 244–248. https://doi.org/10.1038/nature24265.

129. Dong, M., Thennavan, A., Urrutia, E., Li, Y., Perou, C.M., Zou, F., and Jiang, Y. (2021). SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references. Brief. Bioinform. *22*, 416–427. https://doi.org/10.1093/bib/bbz166.

130. Dulin, E., García-Barreno, P., and Guisasola, M.C. (2010). Extracellular heat shock protein 70 (HSPA1A) and classical vascular risk factors in a general population. Cell Stress Chaperones *15*, 929–937. https://doi.org/10.1007/s12192-010-0201-2.

131. Takayama, K., Horie-Inoue, K., Katayama, S., Suzuki, T., Tsutsumi, S., Ikeda, K., Urano, T., Fujimura, T., Takagi, K., Takahashi, S., et al. (2013). Androgen-responsive long noncoding RNA CTBP1-AS promotes prostate cancer. EMBO J. *32*, 1665–1680. https://doi.org/10.1038/emboj.2013.99.

132. Nikkanen, J., Leong, Y.A., Krause, W.C., Dermadi, D., Maschek, J.A., Van Ry, T., Cox, J.E., Weiss, E.J., Gokcumen, O., Chawla, A., et al. (2022). An evolutionary trade-off between host immunity and metabolism drives fatty liver in male mice. Science *378*, 290–295. https://doi.org/10.1126/science.abn9886.

133. Stein, M.M., Conery, M., Magnaye, K.M., Clay, S.M., Billstrand, C., Nicolae, R., Naughton, K., Ober, C., and Thompson, E.E. (2021). Sex-specific differences in peripheral blood leukocyte transcriptional response to LPS are enriched for HLA region and X chromosome genes. Sci. Rep. *11*, 1107. https://doi.org/10.1038/s41598-020-80145-z.

134. Kamitaki, N., Sekar, A., Handsaker, R.E., de Rivera, H., Tooley, K., Morris, D.L., Taylor, K.E., Whelan, C.W., Tombleson, P., Loohuis, L.M.O., et al. (2020). Complement genes contribute sex-biased vulnerability in diverse disorders. Nature *582*, 577–581. https://doi.org/10.1038/s41586-020-2277-x.

135. Ngo, S.T., Steyn, F.J., and McCombe, P.A. (2014). Gender differences in autoimmune disease. Front. Neuroendocrinol. *35*, 347–369. https://doi.org/10.1016/j.yfrne.2014.04.004.

136. Schneider-Hohendorf, T., Görlich, D., Savola, P., Kelkka, T., Mustjoki, S., Gross, C.C., Owens, G.C., Klotz, L., Dornmair, K., Wiendl, H., et al. (2018). Sex bias in MHC I-associated shaping of the adaptive immune system. Proc. Natl. Acad. Sci. *115*, 2168–2173. https://doi.org/10.1073/pnas.1716146115.

137. Chella Krishnan, K., Sabir, S., Shum, M., Meng, Y., Acín-Pérez, R., Lang, J.M., Floyd, R.R., Vergnes, L., Seldin, M.M., Fuqua, B.K., et al. (2019). Sex-specific metabolic functions of adipose Lipocalin-2. Mol. Metab. *30*, 30–47. https://doi.org/10.1016/j.molmet.2019.09.009.

138. Ogunmoroti, O., Osibogun, O., Zhao, D., Mehta, R.C., Ouyang, P., Lutsey, P.L., Robinson-Cohen, C., and Michos, E.D. (2022). Associations between endogenous sex hormones and FGF-23 among women and men in the Multi-Ethnic Study of Atherosclerosis. PLoS ONE *17*, e0268759. https://doi.org/10.1371/journal.pone.0268759.

139. Gauthier, V., Kyriazi, M., Nefla, M., Pucino, V., Raza, K., Buckley, C.D., and Alsaleh, G. (2023). Fibroblast heterogeneity: Keystone of tissue homeostasis and pathology in inflammation and

ageing. Front. Immunol. *14*. https://doi.org/10.3389/fimmu.2023.1137659.

140. Chang, E., Varghese, M., and Singer, K. (2018). Gender and Sex Differences in Adipose Tissue. Curr. Diab. Rep. *18*, 69. https://doi.org/10.1007/s11892-018-1031-3.

141. 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, and Principal investigators (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*, 367–372. https://doi.org/10.1038/s41586-018-0590-4.

142. Tabula Muris Consortium (2020). A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. Nature *583*, 590–595. https://doi.org/10.1038/s41586-020-2496-1.

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

144. Pisco, A. (2024). Tabula Sapiens v2. (figshare). https://doi.org/10.6084/M9.FIGSHARE.27921984.V1 https://doi.org/10.6084/M9.FIGSHARE.27921984.V1.

145. Aaron, K.A., Hosseini, D.K., Vaisbuch, Y., Scheibinger, M., Grillet, N., Heller, S., Wang, T., and Cheng, A.G. (2022). Selection Criteria Optimal for Recovery of Inner Ear Tissues from Deceased Organ Donors. Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol. *43*, e507–e514. https://doi.org/10.1097/MAO.0000000000003496.

146. Vaisbuch, Y., Hosseini, D.K., Wagner, A., Hirt, B., Mueller, M., Ponnusamy, R., Heller, S., Cheng, A.G., Löwenheim, H., and Aaron, K.A. (2022). Surgical Approach for Rapid and Minimally Traumatic Recovery of Human Inner Ear Tissues From Deceased Organ Donors. Otol. Neurotol. Off. Publ. Am. Otol. Soc. Am. Neurotol. Soc. Eur. Acad. Otol. Neurotol. *43*, e519–e525. https://doi.org/10.1097/MAO.0000000000003500.

147. Wang, T., Ling, A.H., Billings, S.E., Hosseini, D.K., Vaisbuch, Y., Kim, G.S., Atkinson, P.J., Sayyid, Z.N., Aaron, K.A., Wagh, D., et al. (2024). Single-cell transcriptomic atlas reveals increased regeneration in diseased human inner ear balance organs. Nat. Commun. *15*, 4833. https://doi.org/10.1038/s41467-024-48491-y.

148. Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181. https://doi.org/10.1038/nprot.2014.006.

149. Hagemann-Jensen, M., Ziegenhain, C., and Sandberg, R. (2022). Scalable single-cell RNA sequencing from full transcripts with Smart-seq3xpress. Nat. Biotechnol. *40*, 1452–1457. https://doi.org/10.1038/s41587-022-01311-4.

150. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinforma. Oxf. Engl. *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

151. Putri, G.H., Anders, S., Pyl, P.T., Pimanda, J.E., and Zanini, F. (2022). Analysing high-throughput sequencing data in Python with HTSeq 2.0. Bioinformatics *38*, 2943–2945.

https://doi.org/10.1093/bioinformatics/btac166.

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

153. Yang, S., Corbett, S.E., Koga, Y., Wang, Z., Johnson, W.E., Yajima, M., and Campbell, J.D. (2020). Decontamination of ambient RNA in single-cell RNA-seq with DecontX. Genome Biol. *21*, 57. https://doi.org/10.1186/s13059-020-1950-6.

154. Griffiths, J.A., Richard, A.C., Bach, K., Lun, A.T.L., and Marioni, J.C. (2018). Detection and removal of barcode swapping in single-cell RNA-seq data. Nat. Commun. *9*, 2667. https://doi.org/10.1038/s41467-018-05083-x.

155. Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nat. Methods *15*, 1053–1058. https://doi.org/10.1038/s41592-018-0229-2.

156. Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Valiollah Pour Amiri, V., Hong, J., Wu, K., Jayasuriya, M., Mehlman, E., Langevin, M., et al. (2022). A Python library for probabilistic analysis of single-cell omics data. Nat. Biotechnol. *40*, 163–166. https://doi.org/10.1038/s41587-021-01206-w.

157. Diehl, A.D., Meehan, T.F., Bradford, Y.M., Brush, M.H., Dahdul, W.M., Dougall, D.S., He, Y., Osumi-Sutherland, D., Ruttenberg, A., Sarntivijai, S., et al. (2016). The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. J. Biomed. Semant. *7*, 44. https://doi.org/10.1186/s13326-016-0088-7.

158. Ergen, C., Xing, G., Xu, C., Kim, M., Jayasuriya, M., McGeever, E., Oliveira Pisco, A., Streets, A., and Yosef, N. (2024). Consensus prediction of cell type labels in single-cell data with popV. Nat. Genet., 1–8. https://doi.org/10.1038/s41588-024-01993-3.

159. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.

160. Van de Sande, B., Flerin, C., Davie, K., De Waegeneer, M., Hulselmans, G., Aibar, S., Seurinck, R., Saelens, W., Cannoodt, R., Rouchon, Q., et al. (2020). A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat. Protoc. *15*, 2247–2276. https://doi.org/10.1038/s41596-020-0336-2.

161. Camargo, A.P., Vasconcelos, A.A., Fiamenghi, M.B., Pereira, G.A.G., and Carazzolle, M.F. (2020). tspex: a tissue-specificity calculator for gene expression data. Preprint, https://doi.org/10.21203/rs.3.rs-51998/v1 https://doi.org/10.21203/rs.3.rs-51998/v1.

162. Fang, Z., Liu, X., and Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics *39*, btac757. https://doi.org/10.1093/bioinformatics/btac757.

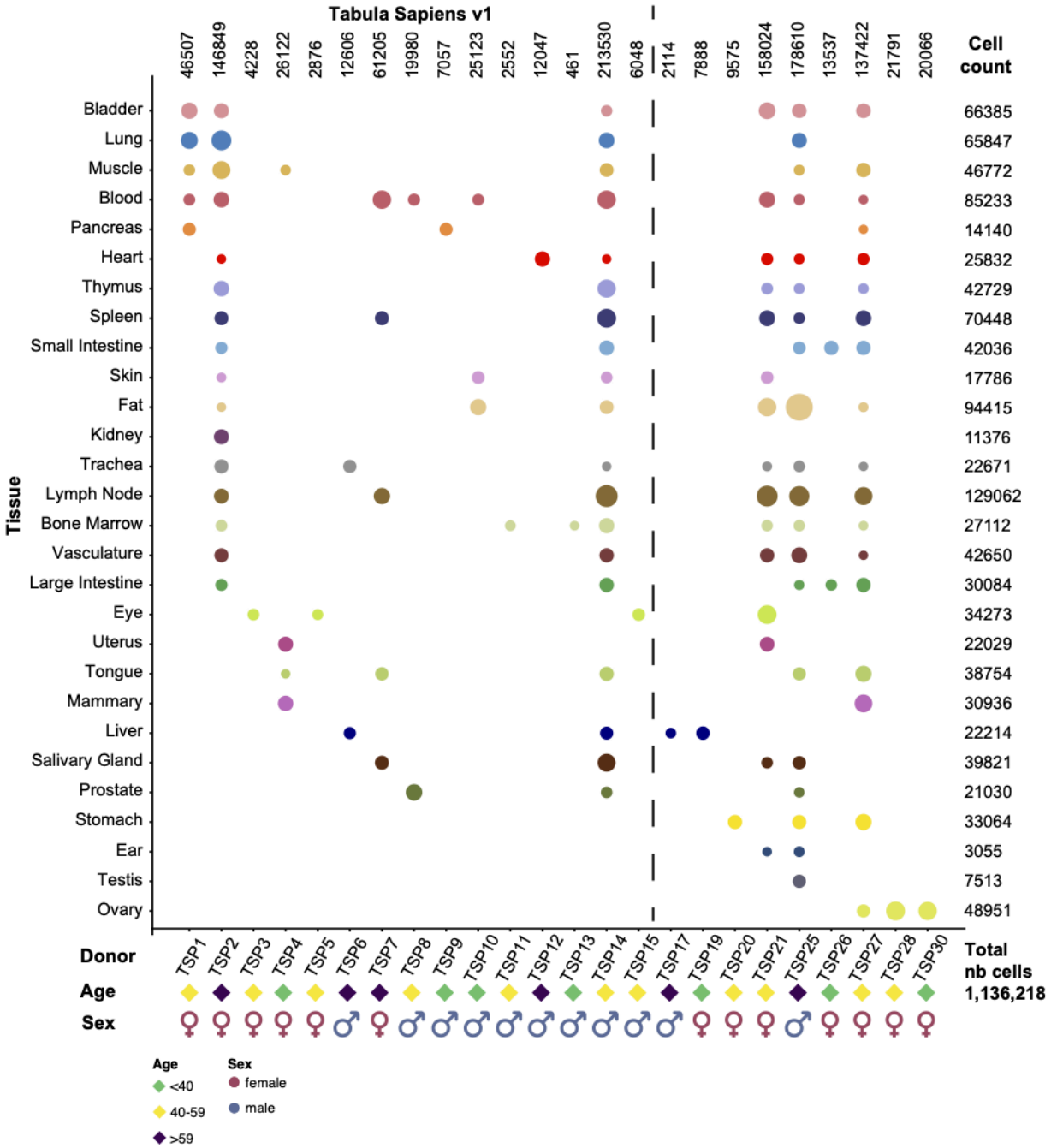
163. Kotliar, D., Veres, A., Nagy, M.A., Tabrizi, S., Hodis, E., Melton, D.A., and Sabeti, P.C. (2019). Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. eLife *8*, e43803. https://doi.org/10.7554/eLife.43803.

164. Chen, Y., Chen, L., Lun, A.T.L., Baldoni, P.L., and Smyth, G.K. (2024). edgeR 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Preprint at bioRxiv, https://doi.org/10.1101/2024.01.21.576131 https://doi.org/10.1101/2024.01.21.576131.

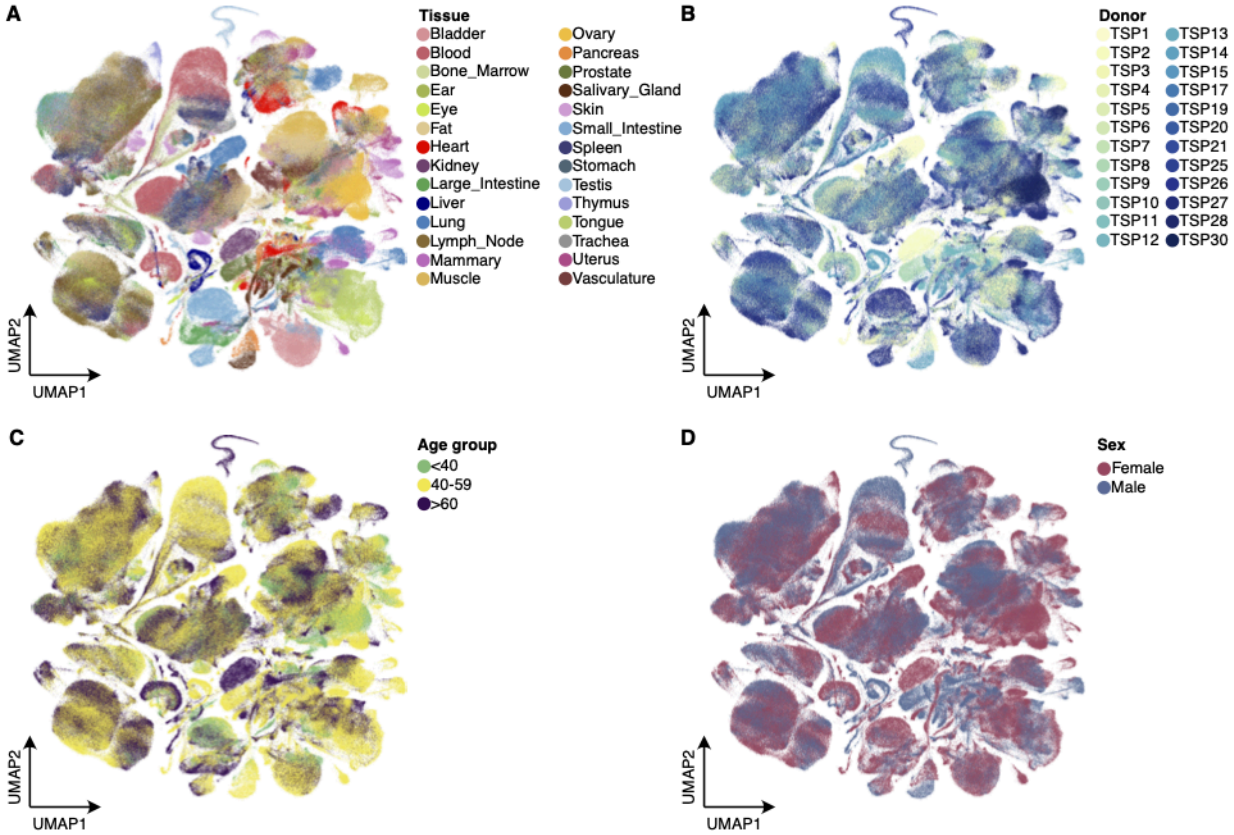
165. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

166. Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Qaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. Nat. Commun. *12*, 5692. https://doi.org/10.1038/s41467-021-25960-2.

Supplementary Figures

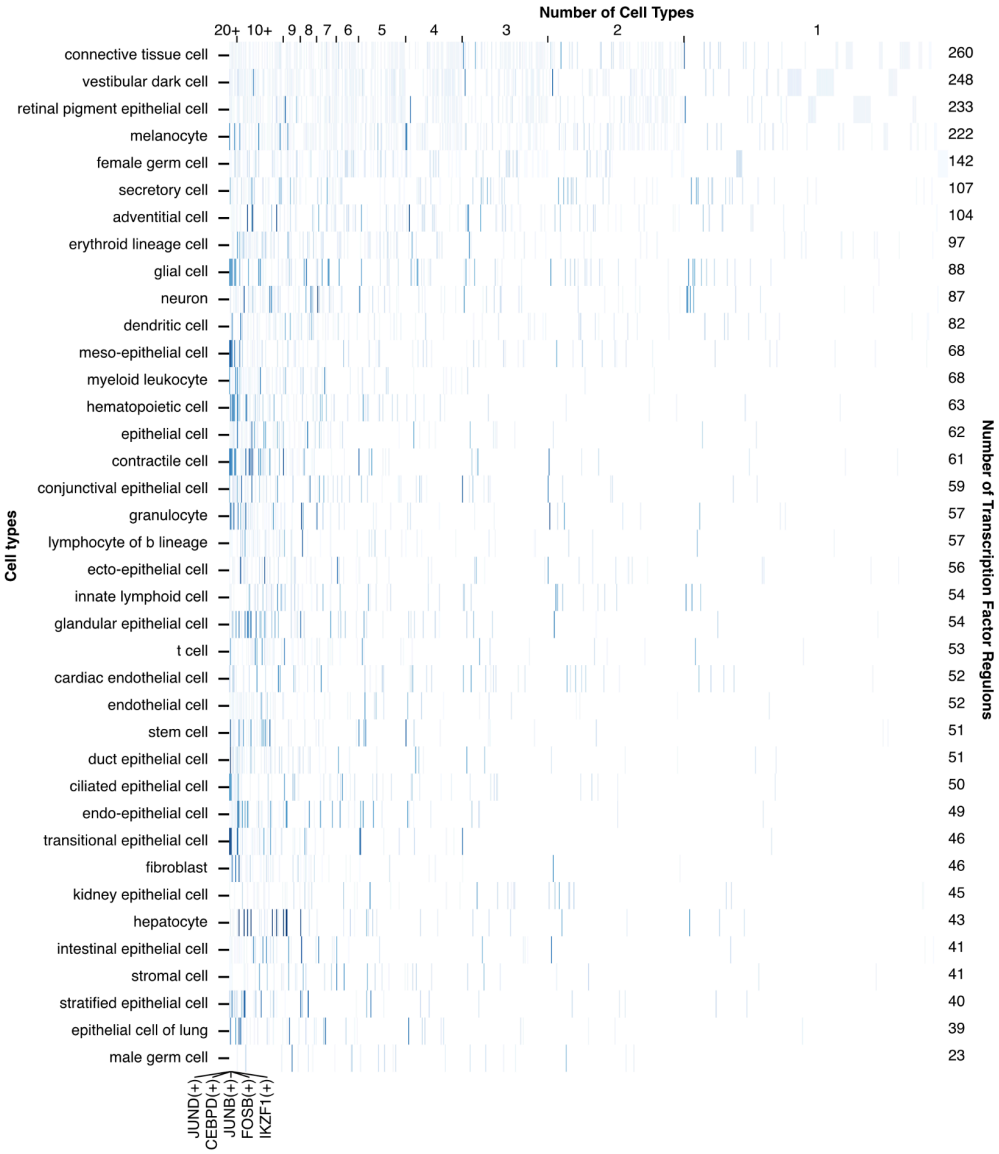


Supplementary Figure 1: **Bubble plot summarizing the distribution of cells across various tissues, donors, ages, and sexes from the Tabula Sapiens 2.0.** Each row corresponds to a different tissue type, with tissues listed on the y-axis. The x-axis represents individual donors (TSP1 to TSP30), with symbols at the bottom indicating donor sex (pink circles for females, blue circles for males) and age group (green diamonds for donors under 40 years old, yellow diamonds for donors aged 40-59, and purple diamonds for donors over 59). The size of the bubbles represents the number of cells sampled from each tissue for each donor, and the color of the bubbles reflects the sex of the donor. On the far right, the total cell count for each tissue is listed, with the sum across all tissues amounting to 1,136,335 cells.

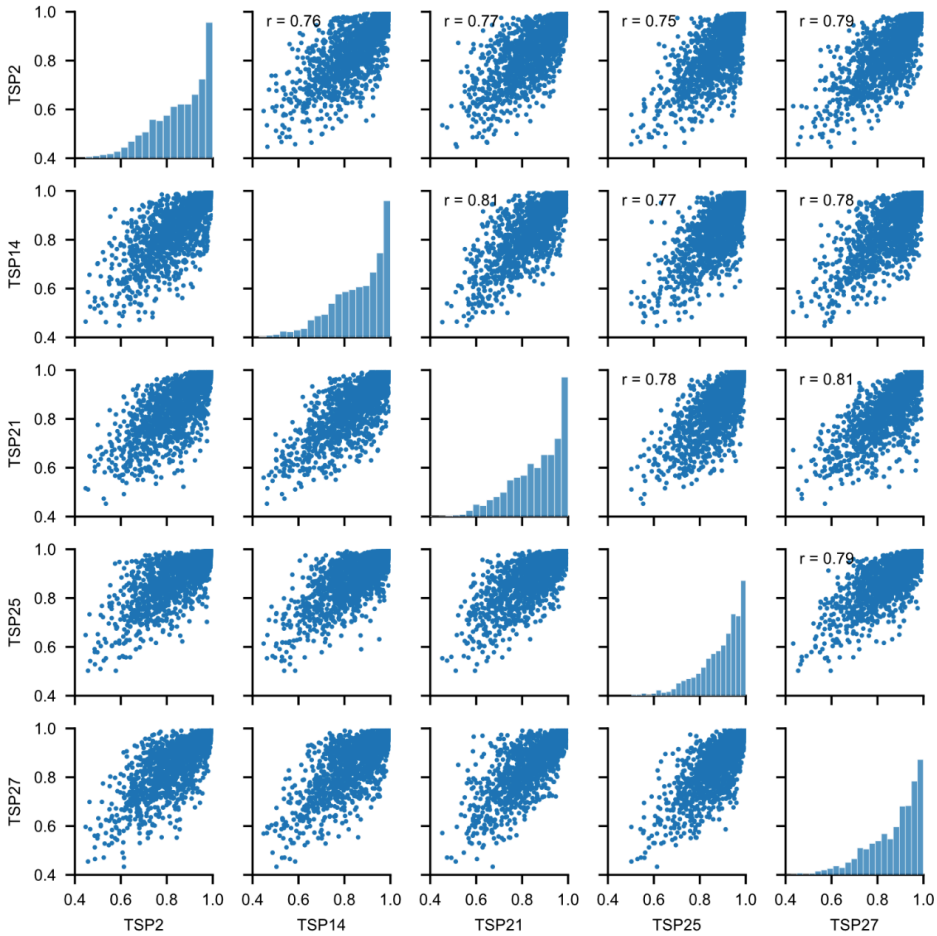


Supplementary Figure 2. UMAP overview of key metadata variables in Tabula Sapiens 2.0. UMAP plots of the expanded Tabula Sapiens 2.0 colored by **A)** Tissue, **B)** Donor, **C)** Age, and **D)** Sex.

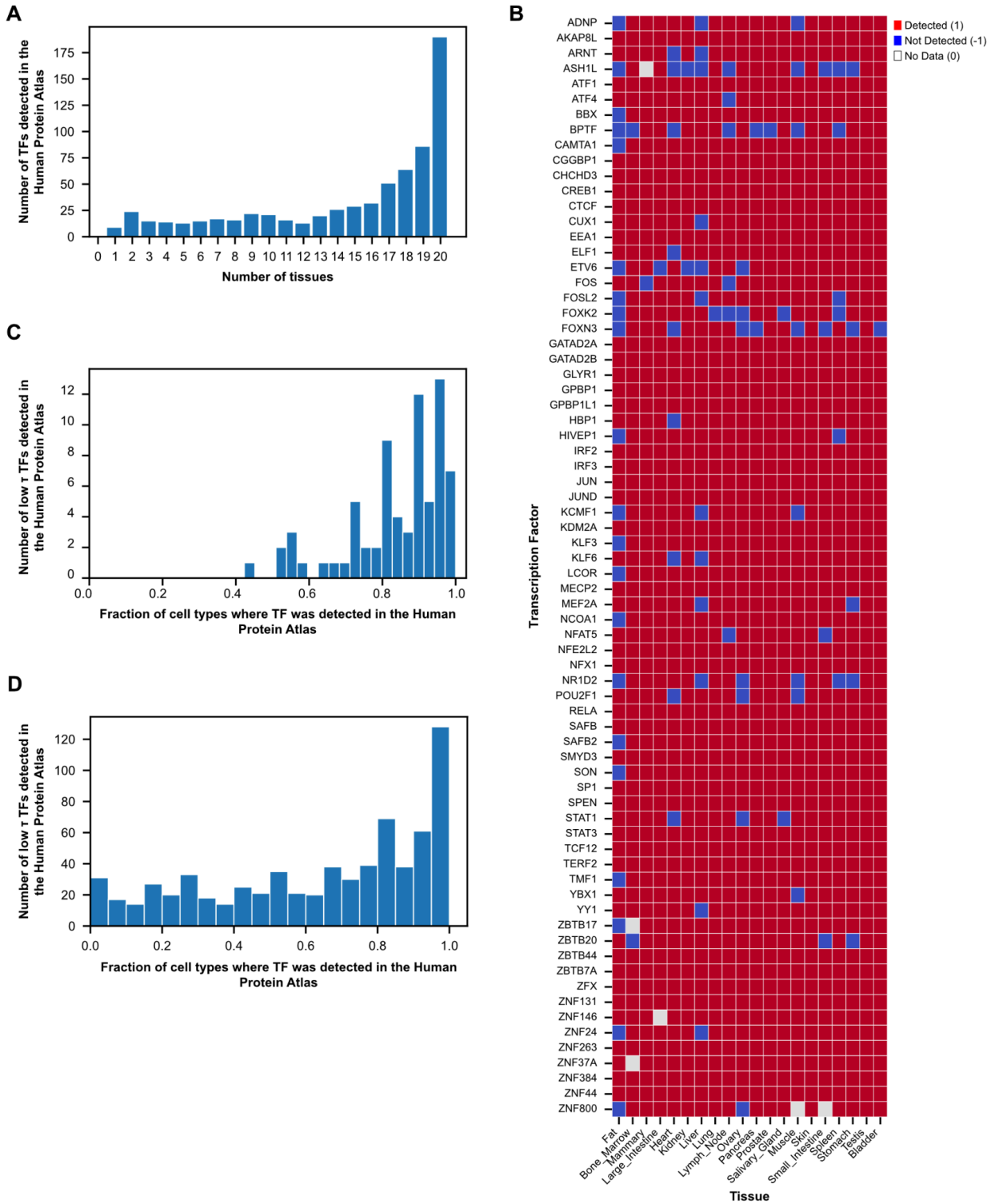
Supplementary Figure 3-6. Data Quality Control.



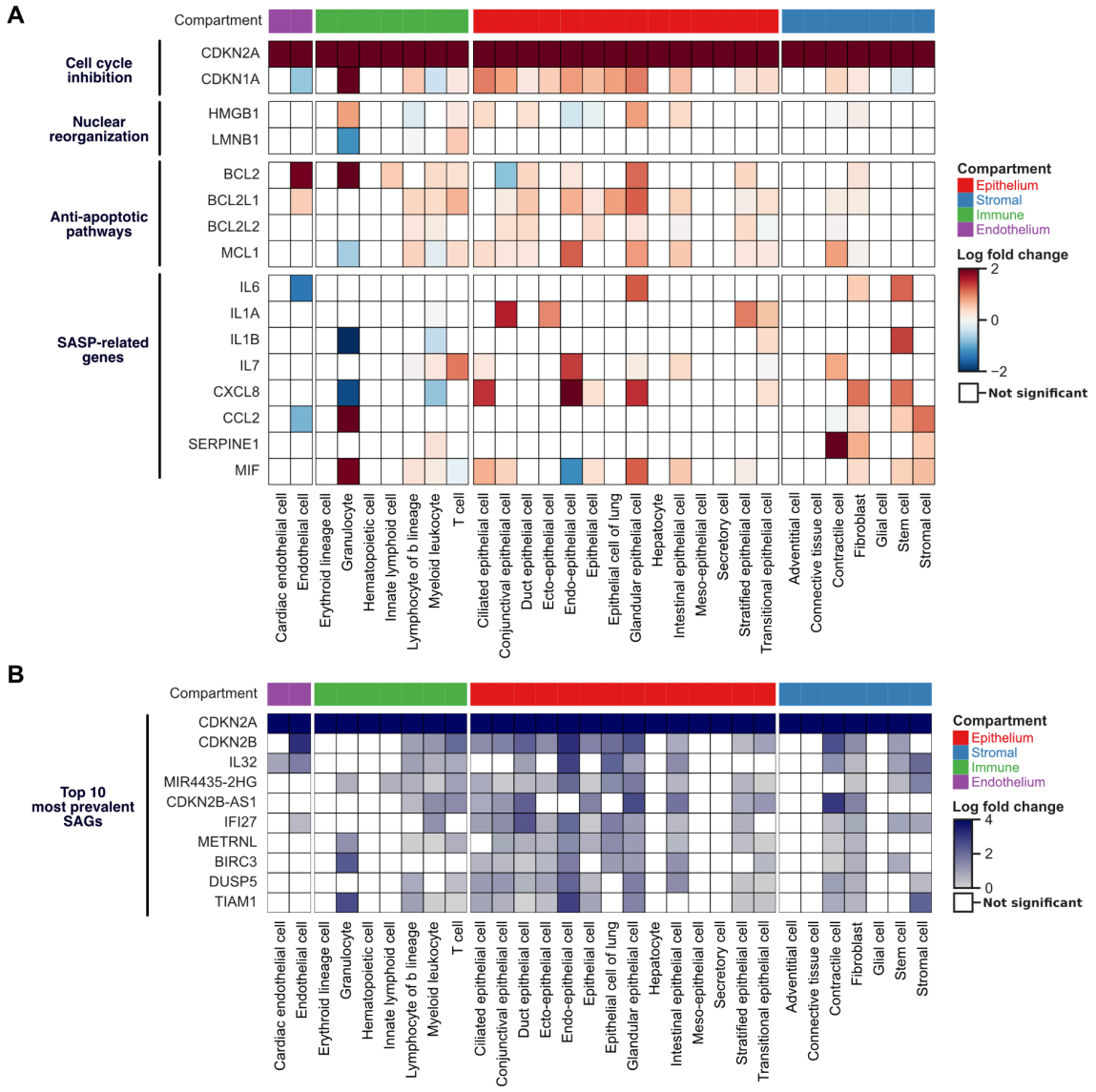
Supplementary Figure 7. Transcription factor activity across cell types. 839 transcription factor regulons identified by SCENIC are shown in descending order by cell type specificity from left to right and cell types are listed in descending order by number of regulons identified for them. Darker shades correspond to higher average activity of a transcription factor within cells of a cell type.



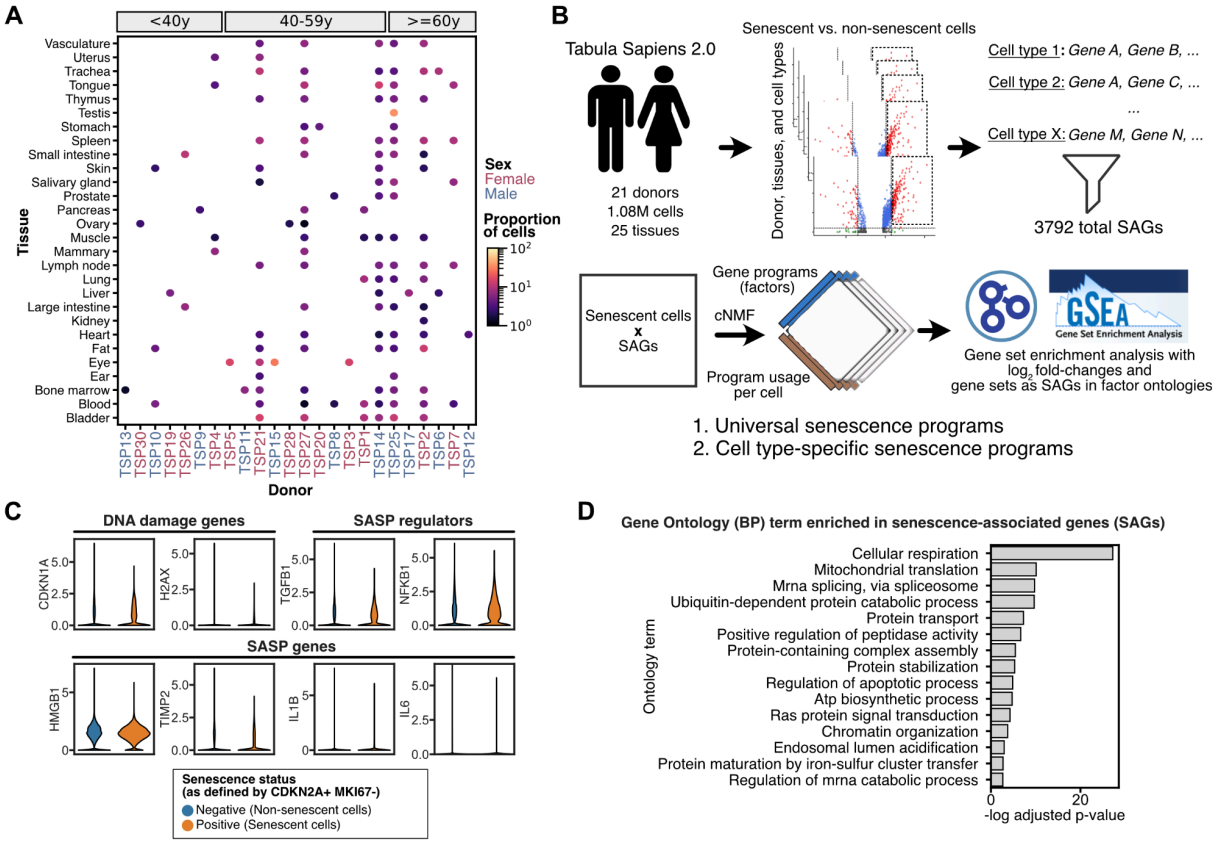
Supplementary Figure 8: Inter-donor pearson correlation of τ specificity. Mean cell type expression was computed on each of donors TSP2, TSP14, TSP21, TSP25 and TSP27, followed by computation of τ cell type specificity for each TF and the pearson correlation between donors.



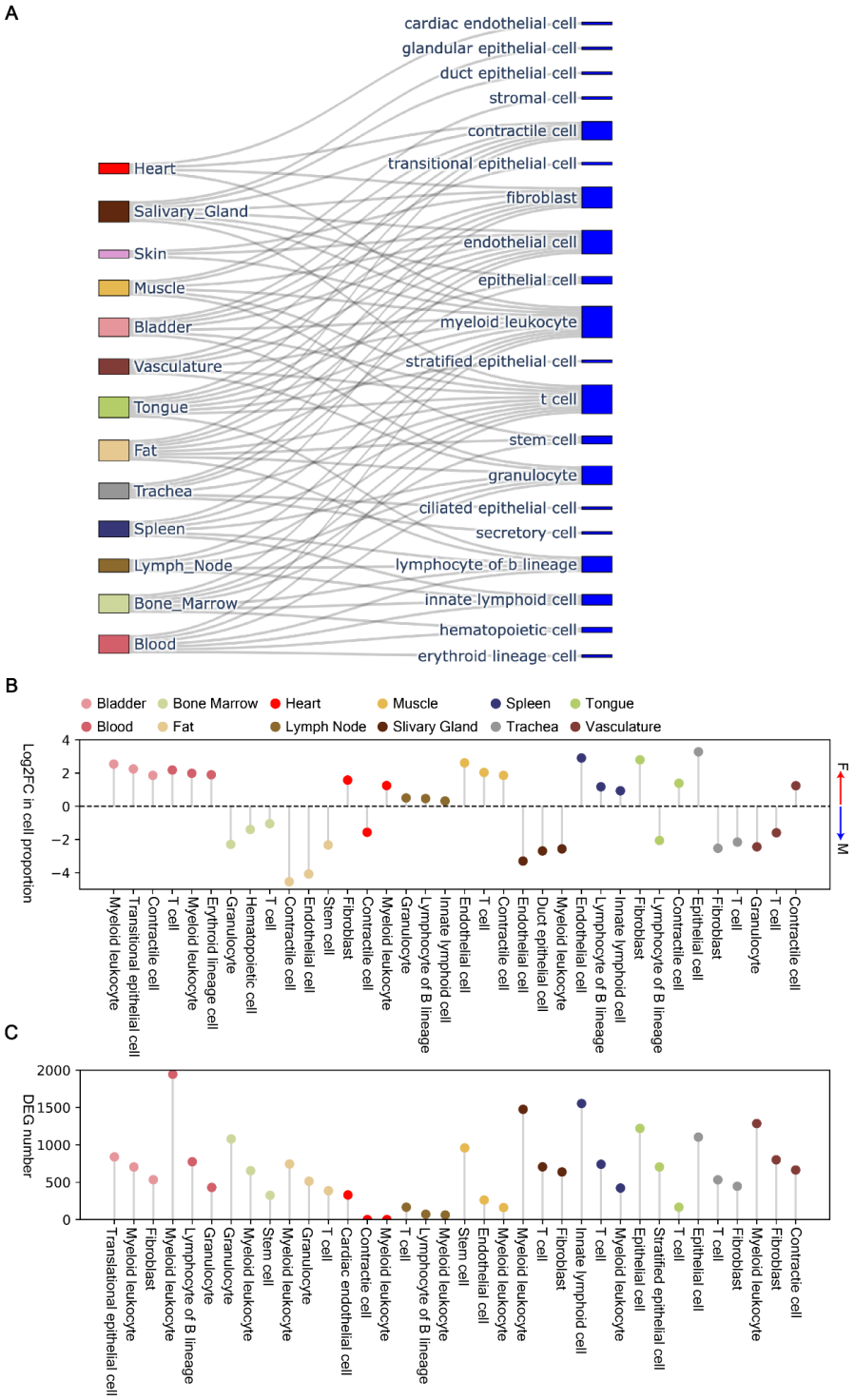
Supplementary Figure 9: Human Protein Atlas immuno detection of transcription factors on a cell-type basis. A) Distribution of the number of tissues in which each of the 699 common TF proteins were detected. **B)** Binary heatmap of detected vs not detected for 72 of the lowest 100 τ TF's in the 20 common tissues. **C)** Distribution of 72 of the 100 lowest τ TF's across the fraction of 32 common cell types in which the TF protein was detected. **D)** The distribution of all 699 TF's tested across the fraction of cell types.



Supplementary Figure 10: **Molecular phenotype of senescent cells across human tissues. A)** Heat map showing the $\log_2$ fold changes for senescence hallmark genes, which were statistically different between senescent and non-senescent cells across broad cell types (adjusted *p*-value < 0.01). The list of hallmark genes was compiled according to SenNet guidelines. Genes are grouped into four distinct hallmarks of senescence, labeled on the left. Broad cell types were grouped into four compartments, as indicated by the color bar at the top. **B)** Heat map showing the $\log_2$ fold changes for the most universal senescence-associated genes (SAGs), which were statistically different between senescent and non-senescent cells across broad cell types (adjusted *p*-value < 0.01). Broad cell types were grouped into four compartments, as indicated by the color bar at the top.

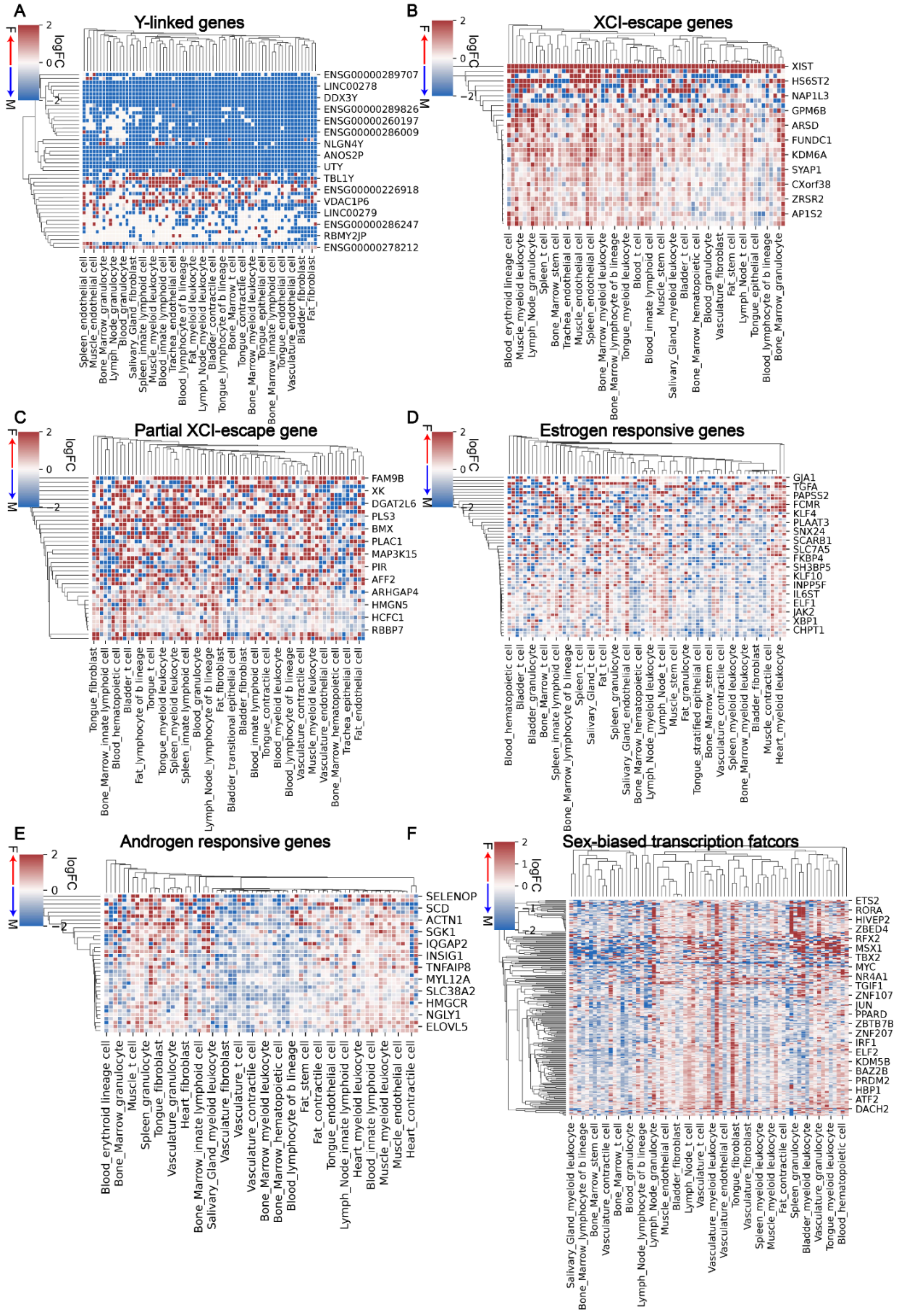


Supplementary Figure 11: **Senescent cells distribution and phenotypes across human tissues. A)** Dot plot showing the proportion of senescent cells, defined as *CDKN2A+ MKI67-* cells, across all tissues and donors. The donor names were colored by sex and were split by age in the annotation bar on top. **B)** Workflow chart showing the methodological steps used to characterize universal and non-universal senescence-associated genes (SAGs). **C)** Violin plots showing the expression of various canonical SAGs in senescent (*CDKN2A+ MKI67-*) and non-senescent cells (*CDKN2A-*) in Tabula Sapiens 2.0 dataset. **D)** Bar plot showing top 15 ontology terms enriched for 3972 senescence-associated genes.

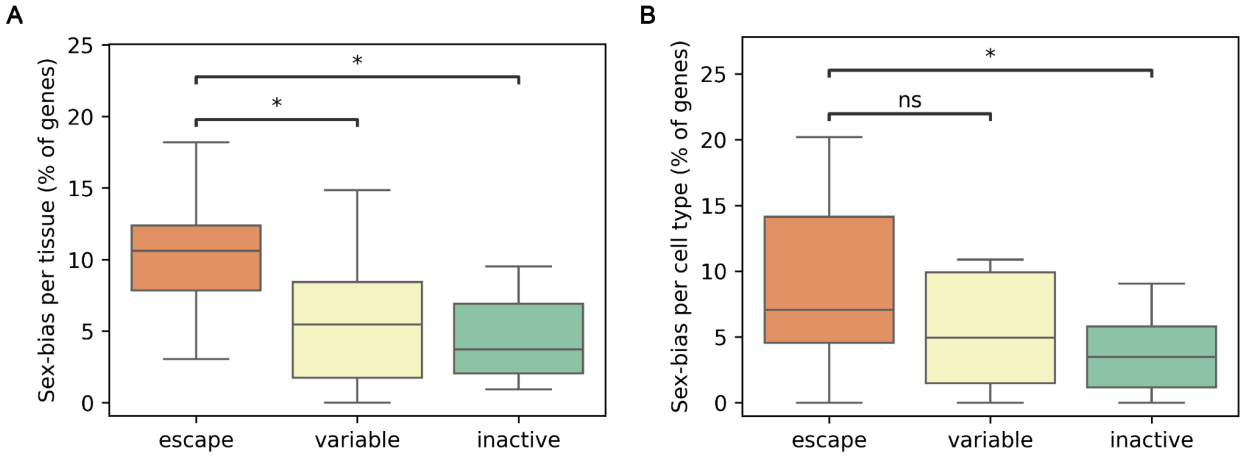


Supplementary Figure 12: **Overview of Tabula Sapiens 2.0 data for sex-biased analysis. A)** Sankey diagram showing the flow of tissue types (left) into specific cell types (right) analyzed for sex-biased expression. Tissues are color-coded and connected to major cell classes, such as cardiac endothelial cells, fibroblasts, contractile cells, epithelial cells, and granulocytes. **B)** Lollipop plot of the log fold change (logFC) for sex-biased gene expression, comparing male and female expression levels across different cell types. Positive logFC values indicate female-biased genes (red), while negative logFC values indicate male-biased genes (blue). **C)** Corresponding plot

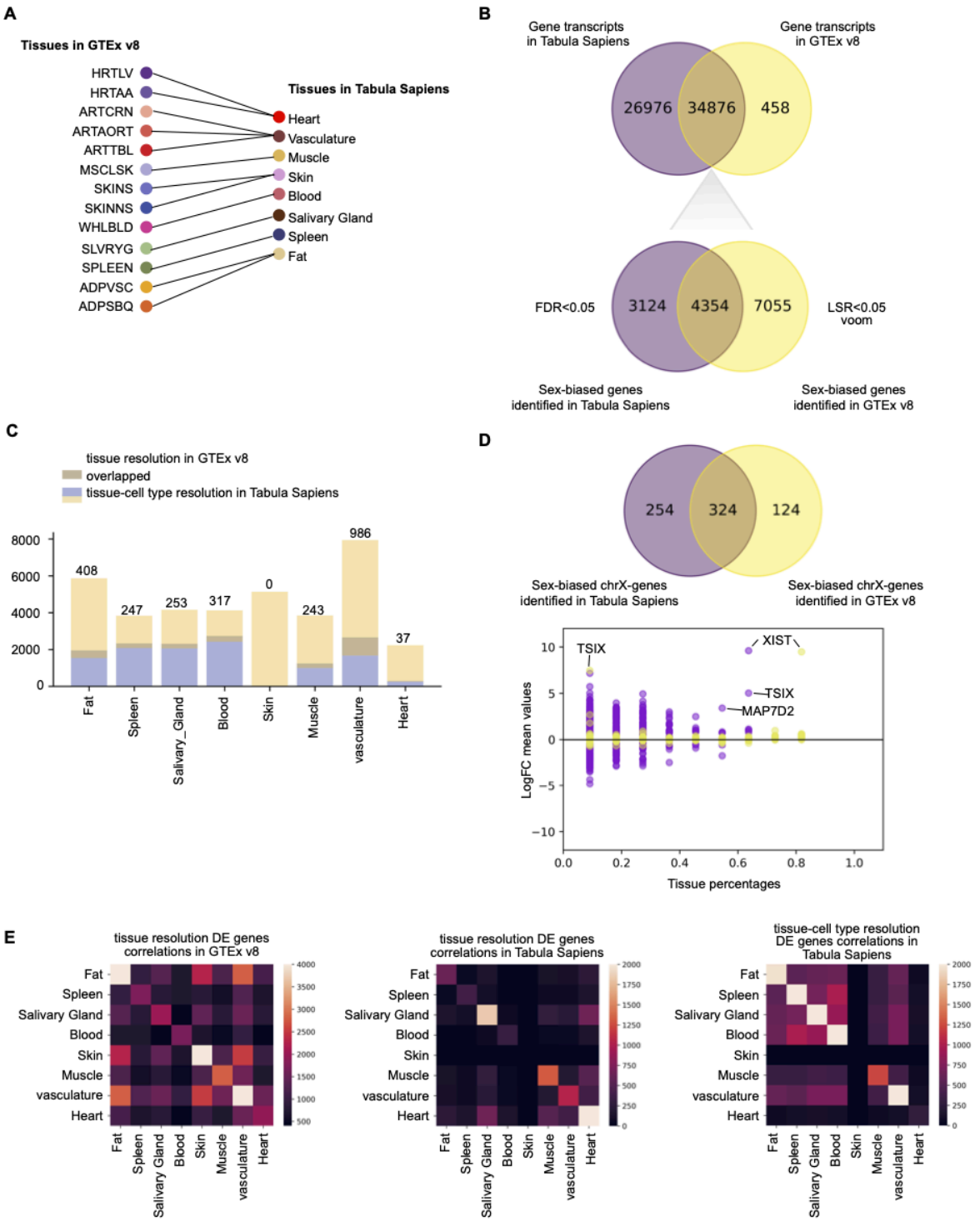
displaying the total number of differentially expressed genes (DEGs) for each cell type. The color of the dots represents the tissue origin, matching the color legend from panel B.


A
Y-linked genes
logFC
F
M
ENSG00000289707
LINC00278
DDX3Y
ENSG00000289826
ENSG00000260197
ENSG00000286009
NLGN4Y
ANOS2P
UTY
TBL1Y
ENSG00000226918
VDAC1P6
LINC00279
ENSG00000286247
RBMY2JP
ENSG00000278212
B
XCI-escape genes
XIST
HS6ST2
NAP1L3
GPM6B
ARSD
FUNDC1
KDM6A
SYAP1
CXorf38
ZRSR2
AP1S2
C
Partial XCI-escape gene
FAM9B
XK
DGAT2L6
PLS3
BMX
PLAC1
MAP3K15
PIR
AFF2
ARHGAP4
HMGN5
HCFC1
RBBP7
D
Estrogen responsive genes
GJA1
TGFA
PAPSS2
FCMR
KLF4
PLAAT3
SNX24
SCARB1
SLC7A5
FKBP4
SH3BP5
KLF10
INPP5F
IL6ST
ELF1
JAK2
XBP1
CHPT1
E
Androgen responsive genes
SELENOP
SCD
ACTN1
SGK1
IQGAP2
INSIG1
TNFAIP8
MYL12A
SLC38A2
HMGCR
NGLY1
ELOVL5
F
Sex-biased transcription fatcors
ETS2
RORA
HIVEP2
ZBED4
RFX2
MSX1
TBX2
MYC
NR4A1
TGIF1
ZNF107
JUN
PPARD
ZBTB7B
ZNF207
IRF1
ELF2
KDM5B
BAZ2B
PRDM2
HBP1
ATF2
DACH2

Supplementary Figure 13. Mean log fold change value of differential gene expression between female to male across. A) Y-linked genes, **B)** chrX inactivation escape genes, **C)** partial escape genes, **D)** estrogen responsive genes and **E)** androgen responsive genes, **F)** sex-biased transcription factors.



Supplementary Figure 14. Barplot of sex-biased per (A) tissue and (B) cell type across XCI escape, variable and inactive status.

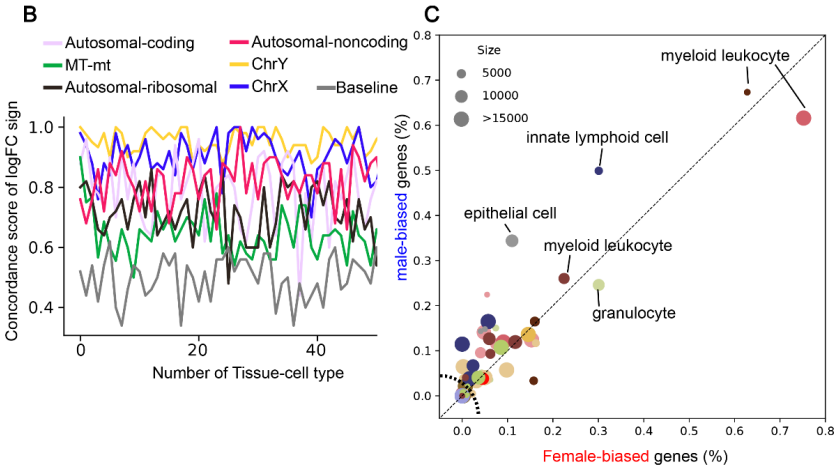


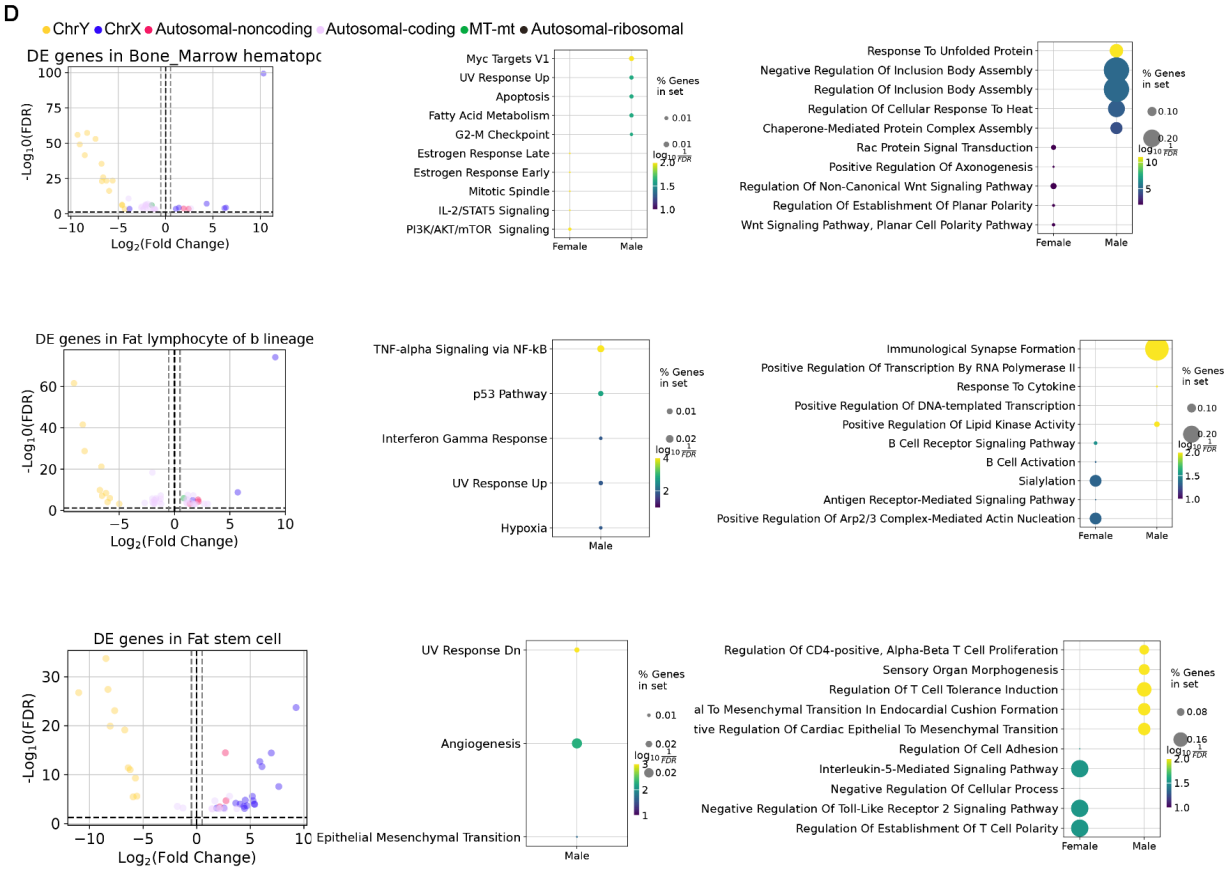
Supplementary Figure 15: **Comparison of differential gene expression across the Tabula Sapiens 2.0 and GTEx Projects. A)** Mapping of overlapping tissues between GTEx v8 and Tabula Sapiens 2.0. Each tissue present in GTEx (left) is connected to its corresponding tissue in Tabula Sapiens (right), including tissues like heart, vasculature, muscle, skin, blood, and fat. **B)** Venn diagrams comparing gene transcripts between Tabula Sapiens 2.0 and GTEx v8. Top: The number of gene transcripts shared between both datasets. Bottom: Comparison of sex-biased genes identified at an FDR < 0.05 in Tabula Sapiens and GTEx using voom analysis. **C)** Bar plot showing the

number of sex-biased genes identified in GTEx (tissue resolution) versus Tabula Sapiens (tissue resolution and tissue-cell type resolution). The orange bars represent genes identified in Tabula Sapiens, yellow bars show genes from GTEx, and the gray bars represent the overlap between the two datasets. **D)** Top: Venn diagram comparing sex-biased X chromosome genes identified in Tabula Sapiens and GTEx, including notable genes such as TSIX and XIST. Bottom: Log fold change (logFC) values for X-linked genes across tissues, highlighting TSIX and MAPT02. **E)** Heatmaps comparing the correlations of differentially expressed genes at the tissue resolution between GTEx v8 (left) and Tabula Sapiens v2 (center), and at the tissue-cell type resolution in Tabula Sapiens 2.0 (right).

A

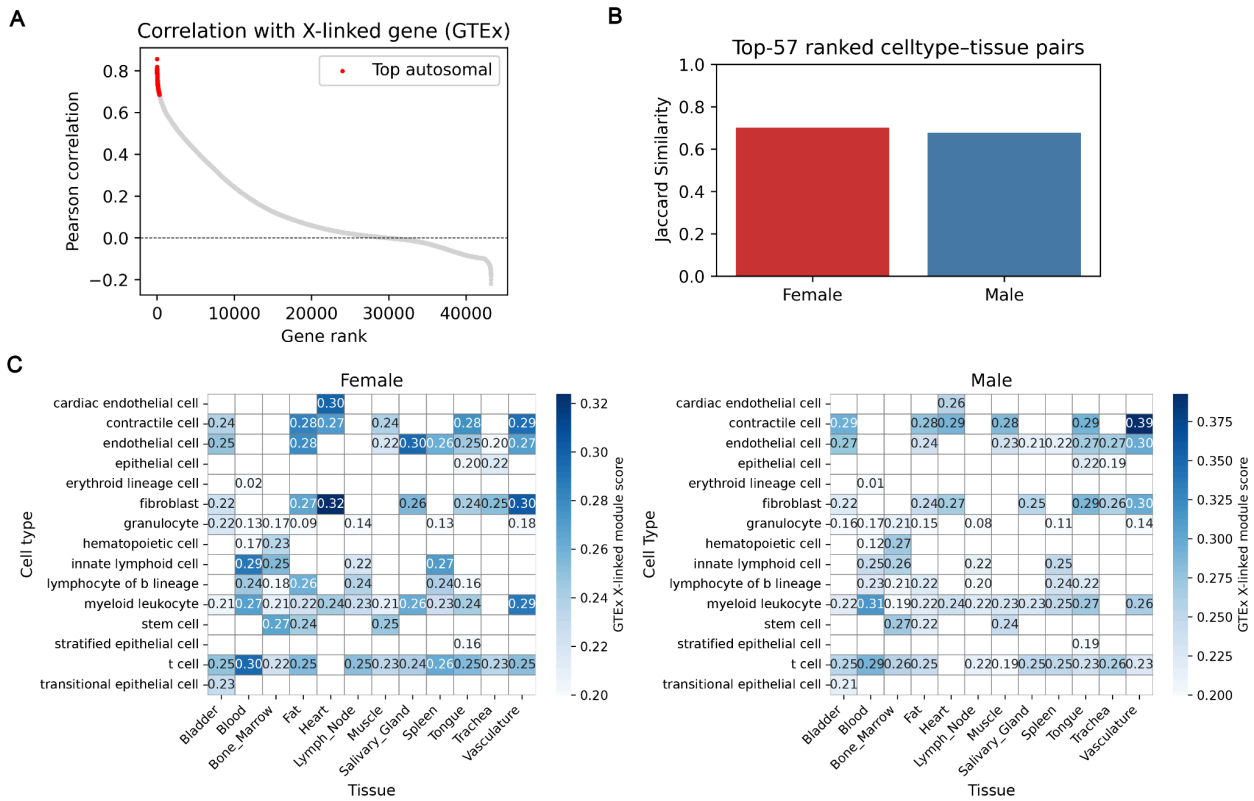
Genes	Tissues	Cell types
HSPA1A	Vasculature Trachea Fat Bone Marrow Bladder	epithelial cell myeloid leukocyte hematopoietic cell t cell fibroblast contractile cell
HSPA1B	Vasculature Tongue Bone Marrow Bladder	epithelial cell stem cell myeloid leukocyte t cell fibroblast
NABP1	Vasculature Trachea Tongue Bone Marrow Bladder	granulocyte stem cell epithelial cell t cell myeloid leukocyte





Supplementary Figure 16. Analysis of sex-biased gene expression across tissue-cell types. A) Table showing representative sex-biased genes identified across various tissues and cell types in the Tabula Sapiens dataset. Tissues include blood, heart, lymph node, salivary gland, and spleen; cell types include B lymphocytes, fibroblasts, contractile cells, and myeloid leukocytes. **B)** Line plot showing concordance scores of differentially expressed genes (DEGs) across tissue-cell types. Concordance is measured for various gene categories, including autosomal-coding, autosomal-noncoding, mitochondrial, chromosome Y (chrY), and chromosome X (chrX). The baseline score is

included for comparison. **C)** Scatter plot displaying the percentage of male-biased versus female-biased genes across cell types. The size of the dots corresponds to the number of sex-biased genes identified in each cell type. Major cell types include B lymphocytes, fibroblasts, contractile cells, and T cells. The plot highlights the distribution of male-biased genes (in blue) and female-biased genes (in red). **D)** Left: Volcano plots illustrating sex-biased genes (logFC) and statistical significance (-log10(FDR)) for select cell types, including T cells, lymphocytes of B lineage, and heart contractile cells. Right: Bubble plots displaying enriched biological pathways for male- and female-enriched genes, with pathway terms related to oxidative phosphorylation, mitotic spindle, UV response, and cellular respiration. Bubble size corresponds to the number of genes involved, and the color represents the logFC direction (female or male).



Supplementary Figure 17: **Autosomal genes co-expressed with X-linked genes across tissues and cell types identified in GTEx are correlated with sex-biased tissue-cell type pairs in Tabula Sapiens 2.0. A)** Top-ranking autosomal genes among the top 300 genes in GTEx coexpressed with X-linked genes. **B)** Jaccard similarity between the top 57 sex-biaed tissue-cell type pairs in GTEx and Tabula Sapiens. **C)** Heatmap of gene expression in Tabula Sapiens 2.0 for autosomal genes identified in GTEx co-expressed with X-linked genes across tissues and cell types.

Full author list

Overall project direction and coordination:
Robert C Jones[1]
Mark Krasnow[2,3]
Angela Oliveira Pisco[4]
Stephen R. Quake[1,5,6]
Julia Salzman[2,7]
Nir Yosef[4,8,9]

Manuscript writing
George Crowley[1]
Siyu He[1,7]
Robert C Jones[1]
Madhav Mantri[1]
Angela Olivera Pisco[4]
Stephen R. Quake[1,5,6]
Anton Thieme[1]

Donor recruitment:
Jessie Aguirre[10]
Ron Garner[10]
Sal Guerrero[10]
William Harper[10]
Resham Irfan[10]
Sophia Mahfouz[10]
Ravi Ponnusamy[10]
Bhavani A. Sanagavarapu[10]
Ahmad Salehi[10]
Ivan Sampson[10]
Chloe Tang[10]

Surgeons:
Alan G. Cheng[11]
James M. Gardner[12,13]
Burnett Kelly[10,14]
Thurman Slone[10]
Zifa Wang[10]

Logistical coordination:
Anika Choudhury[1,4]
Sheela Crasta[1]
George Crowley[1]
Chen Dong[1,4]

Marcus L. Forst[1]
Douglas E. Henze[1]
Jaeyoon Lee[1]
Maurizio Morri[4]
Angela Oliveira Pisco[4]
Serena Y. Tan[15]
Sevahn K. Vorperian[16,17]
Lynn Yang[1,4]

Organ processing:

Marcela Alcántara-Hernádez[18]
Julian Berg[19]
Dhruv Bhatt[20]
Sara Billings[11]
Andrès Gottfried-Blackmore[18,21]
Jamie Bozeman[20]
Simon Bucher[22]
Elisa Caffrey[23]
Amber Casillas[24]
Rebecca Chen[23]
Anika Choudhury[1,4]
Matthew Choi[22]
Sheela Crasta[1]
Rebecca N. Culver
Ivana Cvijovic[1,5]
Ke Ding[26]
Hala Shakib Dhowre[27]
Hua Dong[26]
Kenneth Donaville[22]
Lauren Duan[19]
Xiaochen Fan[20]
Mariko H. Foecke[24]
Francisco X. Galdos[19]
Eliza A. Gaylord[24]
Karen Gonzales[20]
Wiliam R. Goodyer [28]
Michelle Griffin[29]
Yuchao Gu[2,30,31]
Shuo Han[26]
Jun Yan He[23]
Paul Heinrich[19]
Rebeca Arroyo Hornero[18]
Keliana Hui[23]
Juan C. Irwin[24]

SoRi Jang[2]
Annie Jensen[19,32]
Saswati Karmakar[15,25]
Jengmin Kang[33]
Hailey Kang[22]
Soochi Kim[33]
Stewart J. Kim[2,30,31]
William Kong[26]
Mallory A. Laboulaye[26]
Daniel Lee[19]
Gyehyun Lee[34]
Elise Lelou[22]
Anping Li[26]
Baoxiang Li[27]
Wan-Jin Lu[26]
Hayley Raquer-McKay[18]
Elvira Mennillo[34]
Lindsay Moore[35]
Elena Montauti[23]
Karim Mrouj[26]
Shravani Mukherjee[27]
Patrick Neuhöfer [2,30,31]
Saphia Nguyen[22]
Honor Paine[22]
Jennifer B. Parker[26,29]
Julia Pham[23]
Kiet T. Phong[36]
Pratima Prabala[20]
Zhen Qi[26]
Joshua Quintanilla[19,32]
Iulia Rusu[34]
Ali Reza Rais Sadati[19]
Bronwyn Scott[27]
David Seong[18]
Hosu Sin[37]
Hanbing Song[38]
Bikem Soyur[24]
Sean Spencer[18,21]
Varun R. Subramaniam[27]
Michael Swift[1]
Aditi Swarup[27]
Greg Szot[12,13]
Aris Taychameekiatchai[22]
Emily Trimm[20]

Stefan Veizades[19,32]
Sivakamasundari Vijayakumar[26]
Kim Chi Vo[24]
Tian Wang[35]
Timothy Wu[2]
Yinghua Xie[19,32]
William Yue[22]
Zue Zhang[2]

Sequencing:
Angela Detweiler[4]
Honey Mekonen[4]
Norma F. Neff[4]
Sheryl Paul[4]
Amanda Seng[4]
Jia Yan[4]

Biobanking and Histology:
Deana Rae Crystal Colburg[39]
Chen Dong[1,4]
Balint Laszlo Forgo[15]
Luca Ghita[18]
Frank McCarthy[40]
Serena Y. Tan[15]
Lynn Yang[1,4]

cDNA and library prep:
Aditi Agrawal[4]
Anika Choudhury[1,4]
Sheela Crasta[1]
Chen Dong[1,4]
Alina Isakova[1]
Jaeyoon Lee[1]
Maurizio Morri[4]
Kavita Murthy[1]
Alexandra Psaltis[1]
Wenfei Sun[1]
Jia Yan[4]

Data Analysis:
Kyle Awayan[4]
Pierre Boyeau[41]
Robrecht Cannoodt[42–44]
George Crowley[1]

Leah Dorman[4]
Samuel D'Souza[4]
Can Ergen[8,41]
Siyu He[7]
Justin Hong[45]
Harper Hua[7]
Robert C Jones[1]
Jaeyoon Lee[1]
Madhav Mantri[1]
Erin McGeever[4]
Antoine de Morree[33,46,47]
Angela Oliveira Pisco[4]
Luise A. Seeker[1]
Alexander J. Tarashansky[4]
Anton Thieme[1]

Expert cell type annotation:
Marcela Alcántara-Hernádez[18]
Simon Bucher[22]
Ivana Cvijovic[1,5]
Hala Shakib Dhowre[27]
Ke Ding[26]
Hua Dong[26]
Xiaochen Fan[20]
Mariko H. Foecke[24]
Eliza A. Gaylord[24]
Astrid Gillich[2]
Wiliam R. Goodyer [28]
Andrès Gottfried-Blackmore[18,21]
Rebeca Arroyo Hornero[18]
Taha A. Jan[48]
SoRi Jang[2]
Soochi Kim[33]
Saswati Karmakar[15,25]
Stewart J. Kim[2,30,31]
William Kong[26]
Mallory A. Laboulaye[26]
Angela Ling[48]
Wan-Jin Lu[26]
Antoine de Morree[46]
Abhishek Murti[22]
Patrick Neuhöfer[2,30,31]
Kiet T. Phong[36]
Hayley Raquer-McKay[18]

Nikita Sajai[22]
Ryan M. Samuel[49]
David Seong[18]
Hosu Sin[37]
Hanbing Song[38]
Michael Swift[1]
Aris Taychameekiatchai[22]
Stefan Veizades[19,32]
Sivakamasundari Vijayakumar[26]
Juliane Winkler[50,51]
Timothy Wu[2]
Zue Zhang[2]

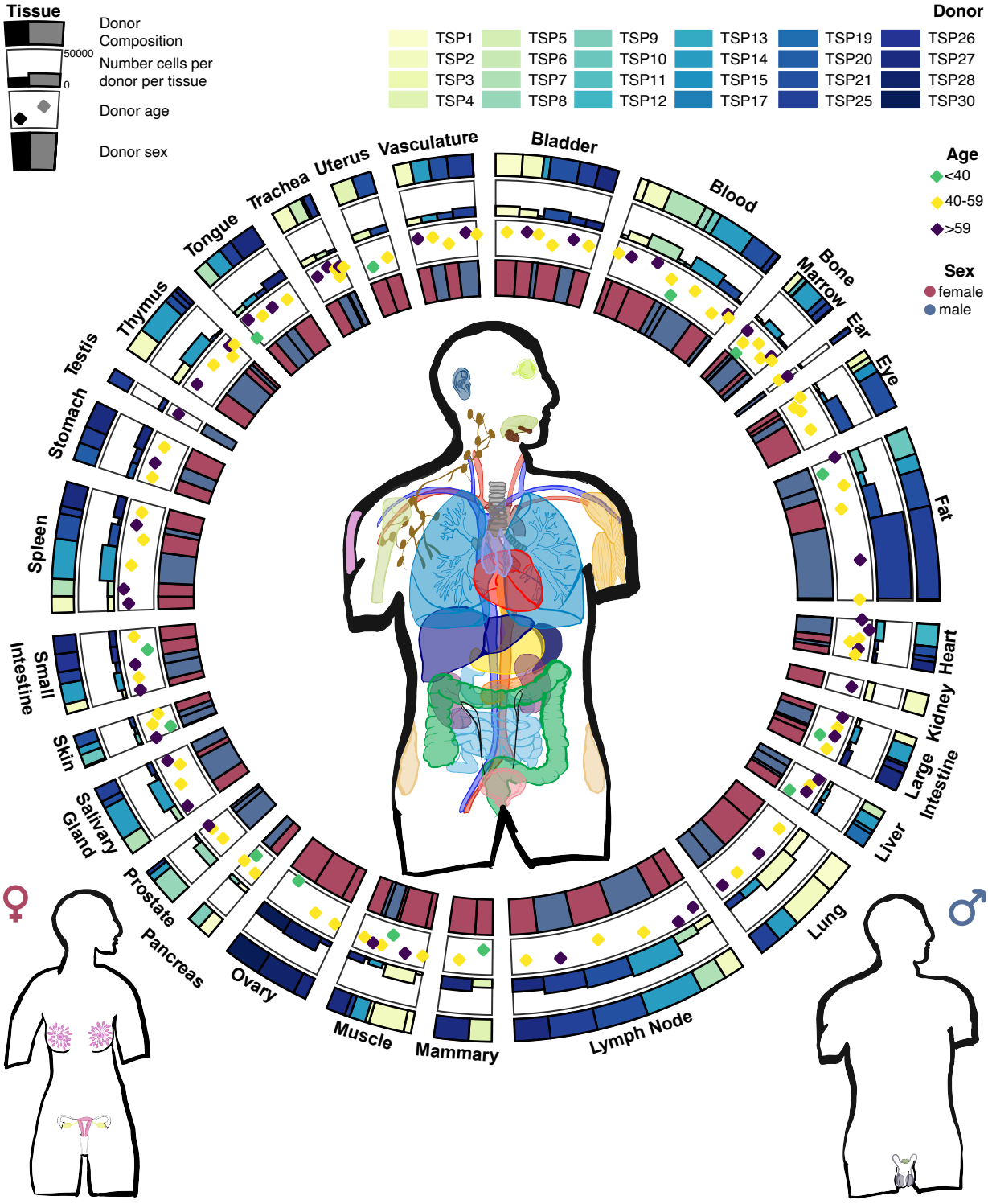
Tissue expert principal investigators:
Steven E. Artandi[2,30,31]
Philip A. Beachy[26,32,52]
Alan G. Cheng[11]
Mike F. Clarke[26]
Zev Gartner[4,53]
Linda C. Giudice[54]
Franklin W. Huang[38,55]
Juliana Idoyaga[18,56]
Michael G. Kattah[34]
Mark Krasnow[2,3]
Christin S. Kuo[57]
Diana J. Laird[24]
Michael T. Longaker[26,58]
Patricia Nguyen[19,32,59]
David Y. Oh[23]
Stephen R. Quake[1,5,6]
Thomas A. Rando[33]
Kristy Red-Horse[20]
Bruce Wang[22]
Albert Y. Wu[27]
Sean M. Wu[19,32]
Bo Yu[37]
James Zou[7,60]

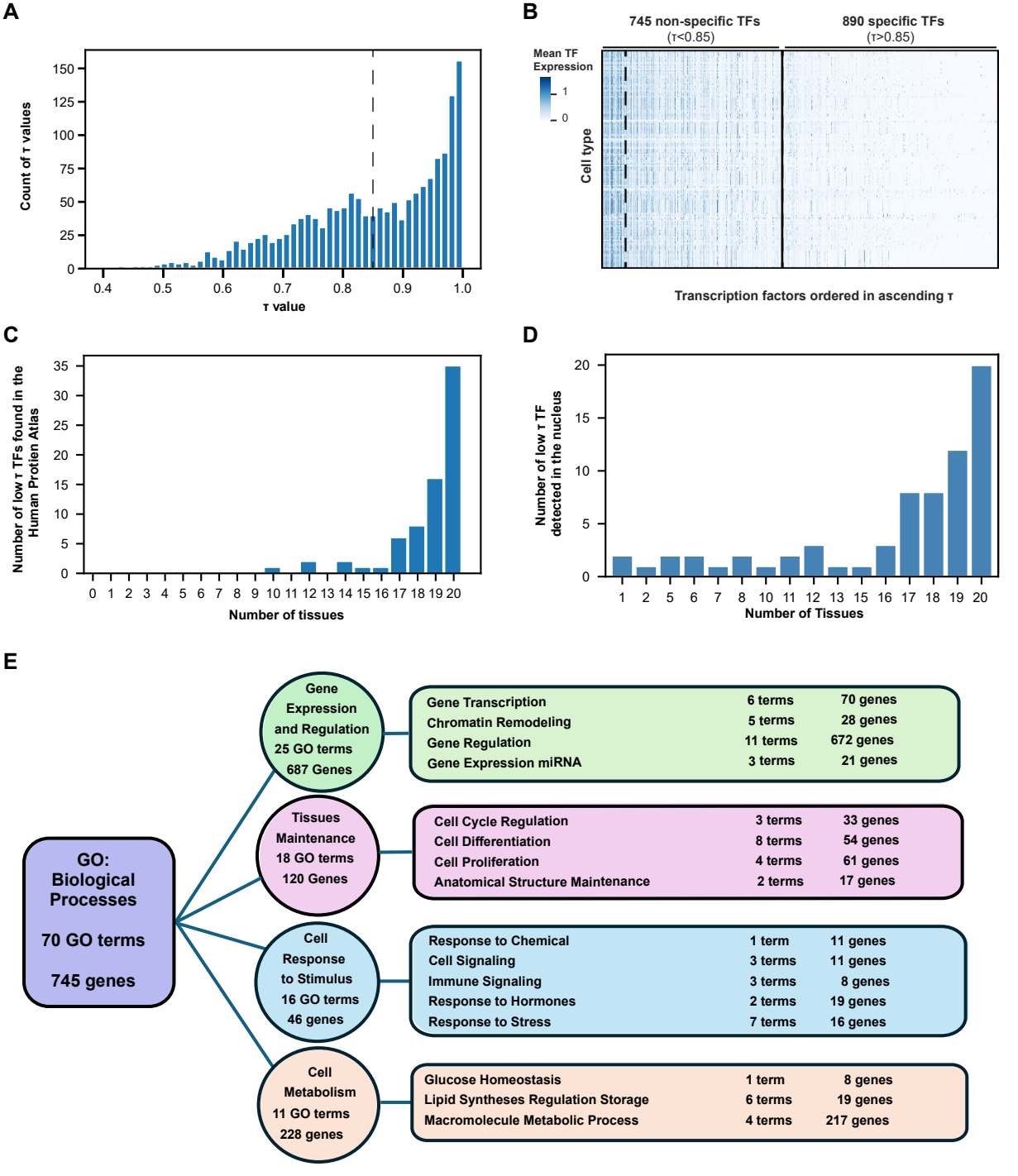
Affiliations

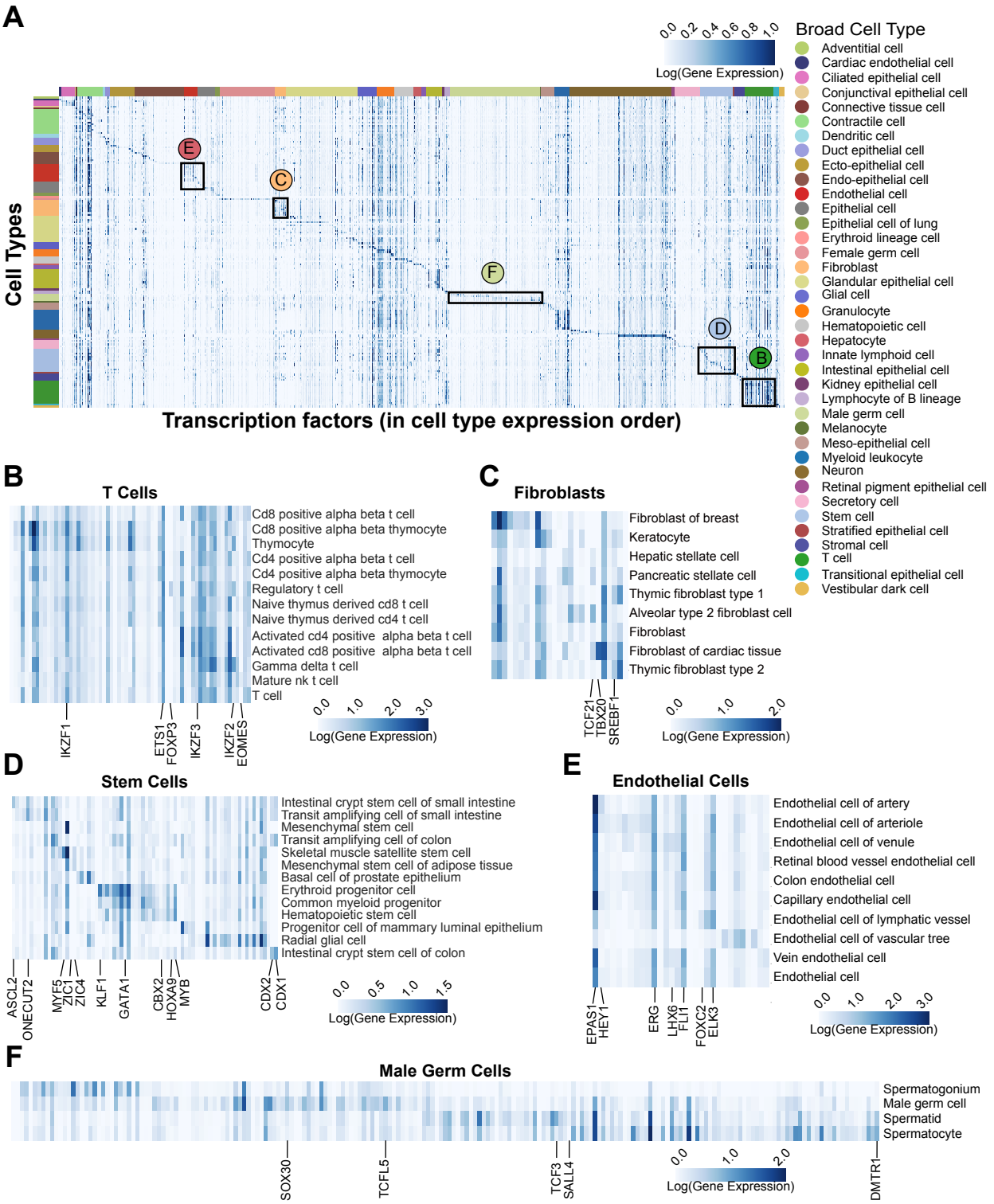
1. Department of Bioengineering, Stanford University; Stanford, CA, USA.
2. Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA.
3. Howard Hughes Medical Institute, USA.
4. Chan Zuckerberg Biohub, San Francisco, CA, USA.

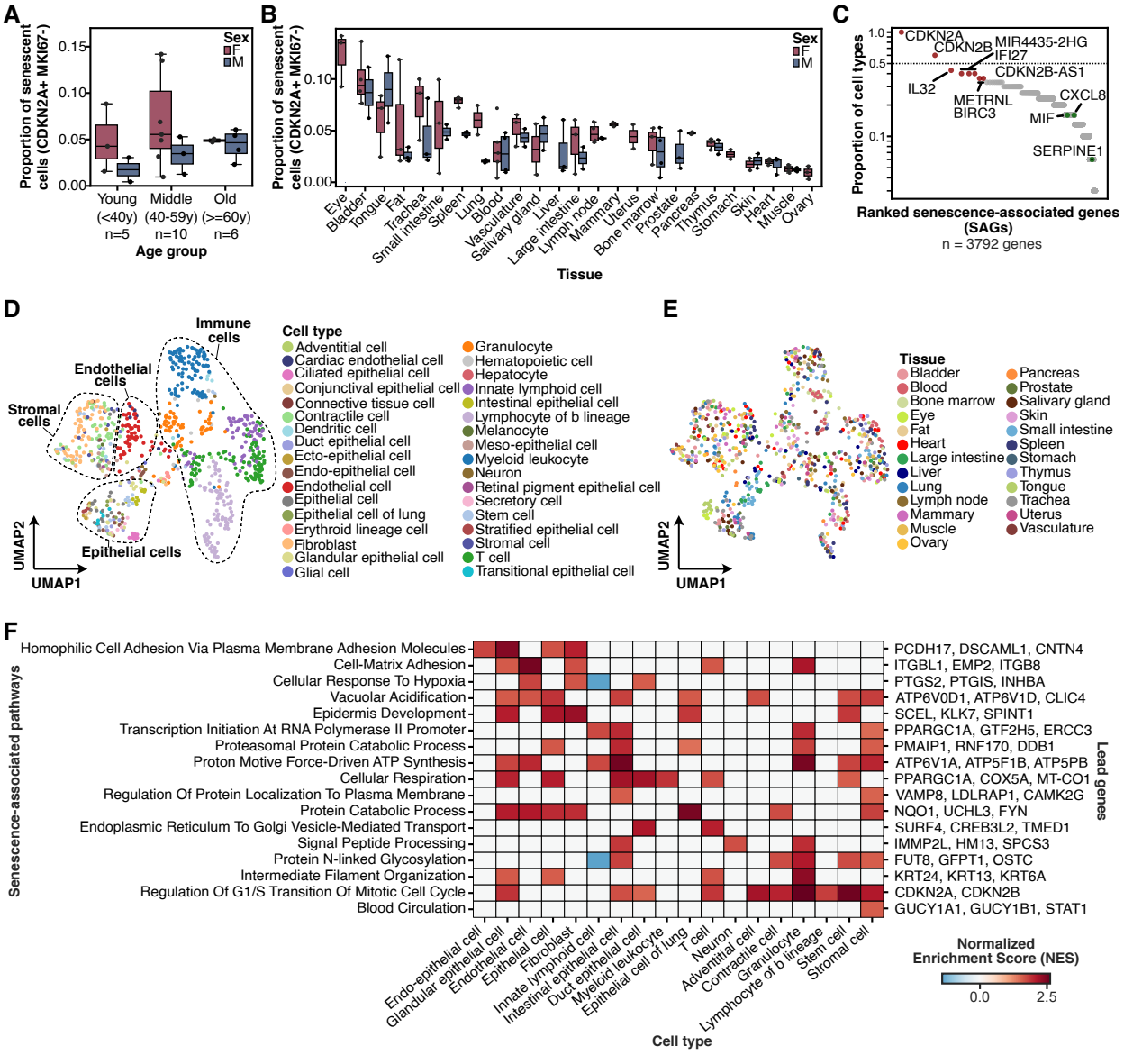
5. Department of Applied Physics, Stanford University, Stanford, CA, USA.
6. The Chan Zuckerberg Initiative, Redwood City, CA, USA.
7. Department of Biomedical Data Science, Stanford University, Stanford, CA, USA.
8. Center for Computational Biology, University of California Berkeley, Berkeley, CA, USA.
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 Chemical Engineering, Stanford University, Stanford, CA, USA.
17. Sarafan ChEM-H, Stanford University, Stanford, CA, USA.
18. Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA.
19. Stanford Cardiovascular Institute, Stanford CA, USA.
20. Department of Biology, Stanford University, Stanford, CA, USA.
21. Division of Gastroenterology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA.
22. Department of Medicine and Liver Center, University of California San Francisco, San Francisco, CA, USA.
23. Division of Hematology/Oncology, Department of Medicine, University of California San Francisco, San Francisco, CA, USA.
24. Department of Ob/Gyn and Reproductive Sciences, Eli and Edythe Broad Center for Regeneration Medicine and Stem Cell Research, University of California, San Francisco.
25. Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA.
26. Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA.
27. Department of Ophthalmology, Stanford University School of Medicine, Stanford, CA, USA.
28. Department of Pediatrics, Division of Cardiology, Stanford University School of Medicine, Stanford, CA, USA.
29. Department of Surgery, Division of Plastic and Reconstructive Surgery, Stanford University School of Medicine, Stanford, CA, USA.
30. Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA, USA.
31. Department of Medicine, Division of Hematology, Stanford University School of Medicine, Stanford, CA, USA.
32. Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA, USA.
33. Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA.
34. Division of Gastroenterology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA.
35. Division of Pediatric Otolaryngology Stanford University School of Medicine, Stanford, CA, USA.

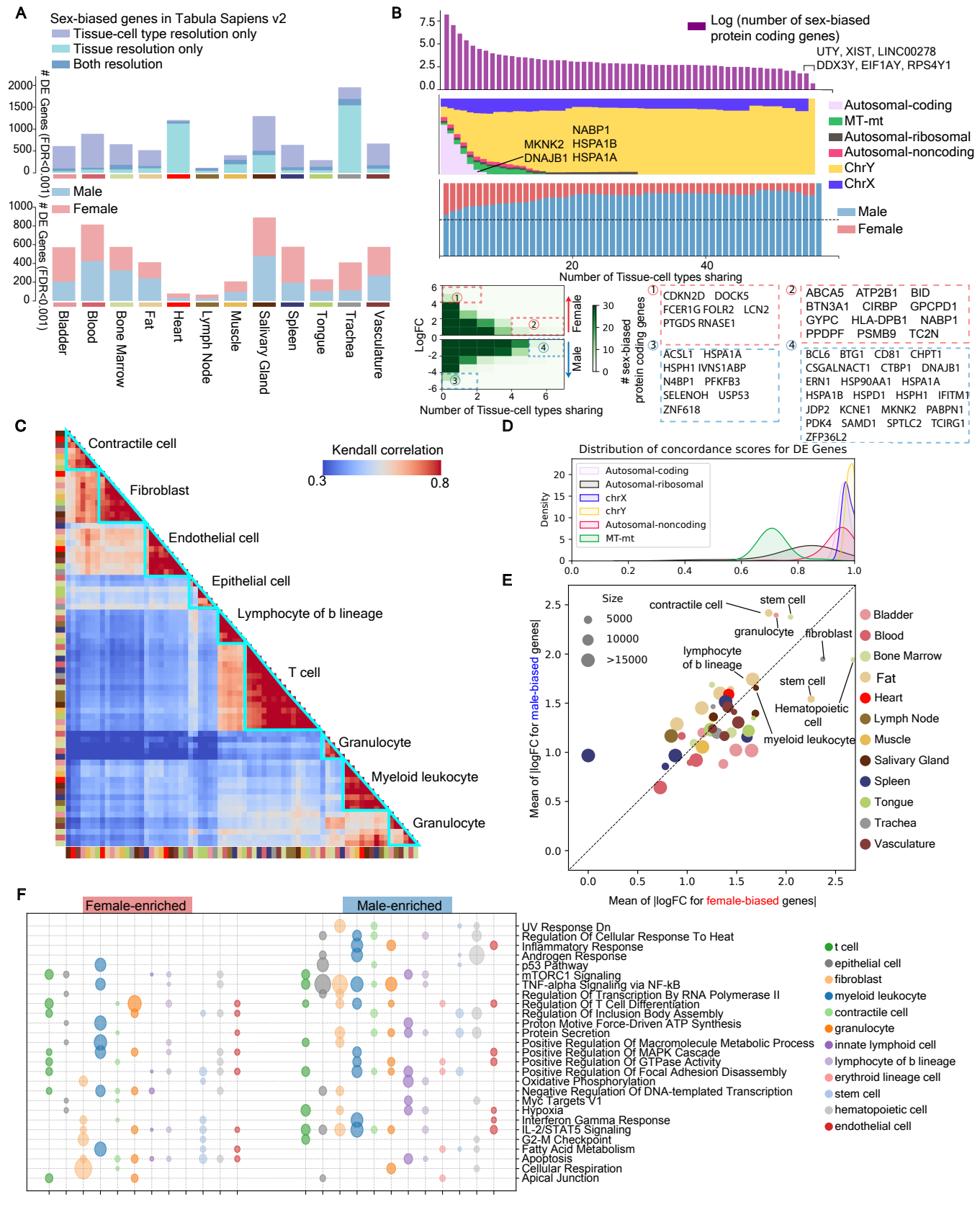
36. Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA.
37. Department of OB/GYN Stanford University, Palo Alto, CA, USA.
38. Division of Hematology and Oncology, Department of Medicine, Bakar Computational Health Sciences Institute, Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA.
39. Stanford Health Care, Stanford CA, USA.
40. Mass Spectrometry Platform, Chan Zuckerberg Biohub, Stanford, CA, USA.
41. Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA, USA.
42. Data Intuitive, Flanders, Belgium.
43. Data Mining and Modelling for Biomedicine group, VIB Center for Inflammation Research, Ghent, Belgium.
44. Department of Applied Mathematics, Computer Science, and Statistics, Ghent University, Ghent, Belgium.
45. Department of Computer Science, Columbia University; New York, NY, USA.
46. Department of Biomedicine, Aarhus University, Aarhus, Denmark.
47. Paul F. Glenn Center for the Biology of Aging, Stanford University School of Medicine, Stanford, CA, USA.
48. Department of Otolaryngology, Vanderbilt University Medical Center, Nashville, TN, USA.
49. Department of Cellular Molecular Pharmacology, University of California, San Francisco, San Francisco, CA, USA.
50. Department of Cell & Tissue Biology, University of California San Francisco, San Francisco, CA, USA.
51. Center for Cancer Research Medical University of Vienna Borschkegasse 8a 1090 Vienna, Austria.
52. Department of Urology, Stanford University School of Medicine, Stanford, CA, USA.
53. Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA.
54. Center for Reproductive Sciences, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco, San Francisco, CA, USA.
55. Division of Hematology/Oncology, Department of Medicine, San Francisco Veterans Affairs Health Care System, San Francisco, CA, USA.
56. Pharmacology and Molecular Biology Departments, Schools of Medicine and Biological Sciences, University of California, San Diego, CA, USA.
57. Department of Pediatrics, Division of Pulmonary Medicine, Stanford University, Stanford, CA, USA.
58. Department of Surgery, Stanford University School of Medicine, Stanford, CA, USA.
59. Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, USA.
60. Department of Computer Science, Stanford University, Stanford, CA.


Tissue
Donor Composition
50000
Number cells per donor per tissue
0
Donor age
Donor sex
Donor
TSP1
TSP2
TSP3
TSP4
TSP5
TSP6
TSP7
TSP8
TSP9
TSP10
TSP11
TSP12
TSP13
TSP14
TSP15
TSP17
TSP19
TSP20
TSP21
TSP25
TSP26
TSP27
TSP28
TSP30
Age
<40
40-59
>59
Sex
female
male
Vasculature
Bladder
Blood
Bone Marrow
Ear
Eye
Fat
Heart
Kidney
Large Intestine
Liver
Lung
Lymph Node
Mammary
Muscle
Ovary
Pancreas
Prostate
Salivary Gland
Skin
Small Intestine
Spleen
Stomach
Testis
Thymus
Tongue
Trachea
Uterus

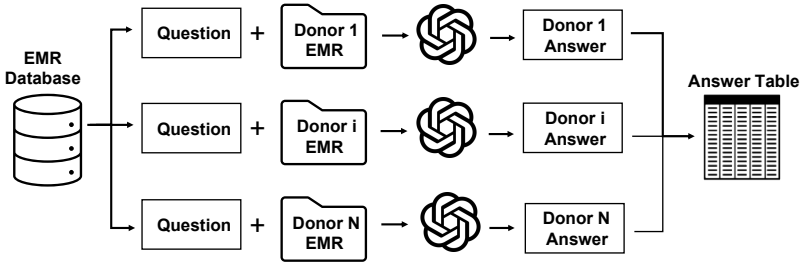

A
Count of τ values
τ value
B
745 non-specific TFs
(τ<0.85)
890 specific TFs
(τ>0.85)
Mean TF Expression
Cell type
Transcription factors ordered in ascending τ
C
Number of low τ TFs found in the Human Protien Atlas
Number of tissues
D
Number of low τ TF detected in the nucleus
Number of Tissues
E
GO: Biological Processes
70 GO terms
745 genes
Gene Expression and Regulation
25 GO terms
687 Genes
Gene Transcription 6 terms 70 genes
Chromatin Remodeling 5 terms 28 genes
Gene Regulation 11 terms 672 genes
Gene Expression miRNA 3 terms 21 genes
Tissues Maintenance
18 GO terms
120 Genes
Cell Cycle Regulation 3 terms 33 genes
Cell Differentiation 8 terms 54 genes
Cell Proliferation 4 terms 61 genes
Anatomical Structure Maintenance 2 terms 17 genes
Cell Response to Stimulus
16 GO terms
46 genes
Response to Chemical 1 term 11 genes
Cell Signaling 3 terms 11 genes
Immune Signaling 3 terms 8 genes
Response to Hormones 2 terms 19 genes
Response to Stress 7 terms 16 genes
Cell Metabolism
11 GO terms
228 genes
Glucose Homeostasis 1 term 8 genes
Lipid Syntheses Regulation Storage 6 terms 19 genes
Macromolecule Metabolic Process 4 terms 217 genes


A
Log(Gene Expression)
0.0 0.2 0.4 0.6 0.8 1.0
Broad Cell Type
Adventitial cell
Cardiac endothelial cell
Ciliated epithelial cell
Conjunctival epithelial cell
Connective tissue cell
Contractile cell
Dendritic cell
Duct epithelial cell
Ecto-epithelial cell
Endo-epithelial cell
Endothelial cell
Epithelial cell
Epithelial cell of lung
Erythroid lineage cell
Female germ cell
Fibroblast
Glandular epithelial cell
Glial cell
Granulocyte
Hematopoietic cell
Hepatocyte
Innate lymphoid cell
Intestinal epithelial cell
Kidney epithelial cell
Lymphocyte of B lineage
Male germ cell
Melanocyte
Meso-epithelial cell
Myeloid leukocyte
Neuron
Retinal pigment epithelial cell
Secretory cell
Stem cell
Stratified epithelial cell
Stromal cell
T cell
Transitional epithelial cell
Vestibular dark cell
Cell Types
Transcription factors (in cell type expression order)
E
C
F
D
B
B
T Cells
Cd8 positive alpha beta t cell
Cd8 positive alpha beta thymocyte
Thymocyte
Cd4 positive alpha beta t cell
Cd4 positive alpha beta thymocyte
Regulatory t cell
Naive thymus derived cd8 t cell
Naive thymus derived cd4 t cell
Activated cd4 positive alpha beta t cell
Activated cd8 positive alpha beta t cell
Gamma delta t cell
Mature nk t cell
T cell
IKZF1
ETS1
FOXP3
IKZF3
IKZF2
EOMES
0.0 1.0 2.0 3.0
Log(Gene Expression)
C
Fibroblasts
Fibroblast of breast
Keratocyte
Hepatic stellate cell
Pancreatic stellate cell
Thymic fibroblast type 1
Alveolar type 2 fibroblast cell
Fibroblast
Fibroblast of cardiac tissue
Thymic fibroblast type 2
TCF21
TBX20
SREBF1
0.0 1.0 2.0
Log(Gene Expression)
D
Stem Cells
Intestinal crypt stem cell of small intestine
Transit amplifying cell of small intestine
Mesenchymal stem cell
Transit amplifying cell of colon
Skeletal muscle satellite stem cell
Mesenchymal stem cell of adipose tissue
Basal cell of prostate epithelium
Erythroid progenitor cell
Common myeloid progenitor
Hematopoietic stem cell
Progenitor cell of mammary luminal epithelium
Radial glial cell
Intestinal crypt stem cell of colon
ASCL2
ONECUT2
MYF5
ZIC1
ZIC4
KLF1
GATA1
CBX2
HOXA9
MYB
CDX2
CDX1
0.0 0.5 1.0 1.5
Log(Gene Expression)
E
Endothelial Cells
Endothelial cell of artery
Endothelial cell of arteriole
Endothelial cell of venule
Retinal blood vessel endothelial cell
Colon endothelial cell
Capillary endothelial cell
Endothelial cell of lymphatic vessel
Endothelial cell of vascular tree
Vein endothelial cell
Endothelial cell
EPAS1
HEY1
ERG
LHX6
FLI1
FOXC2
ELK3
0.0 1.0 2.0 3.0
Log(Gene Expression)
F
Male Germ Cells
Spermatogonium
Male germ cell
Spermatid
Spermatocyte
SOX30
TCFL5
TCF3
SALL4
DMTR1
0.0 1.0 2.0
Log(Gene Expression)


A
Proportion of senescent cells (CDKN2A+ MKI67-)
Sex
F
M
Young (<40y) n=5
Middle (40-59y) n=10
Old (>=60y) n=6
Age group
B
Tissue
Eye
Bladder
Tongue
Fat
Trachea
Small intestine
Spleen
Lung
Blood
Vasculature
Salivary gland
Liver
Large intestine
Lymph node
Mammary
Uterus
Bone marrow
Prostate
Pancreas
Thymus
Stomach
Skin
Heart
Muscle
Ovary
C
Proportion of cell types
CDKN2A
CDKN2B
MIR4435-2HG
IFI27
CDKN2B-AS1
IL32
METRNL
BIRC3
CXCL8
MIF
SERPINE1
Ranked senescence-associated genes (SAGs)
n = 3792 genes
D
Immune cells
Endothelial cells
Stromal cells
Epithelial cells
UMAP2
UMAP1
Cell type
Adventitial cell
Cardiac endothelial cell
Ciliated epithelial cell
Conjunctival epithelial cell
Connective tissue cell
Contractile cell
Dendritic cell
Duct epithelial cell
Ecto-epithelial cell
Endo-epithelial cell
Endothelial cell
Epithelial cell
Epithelial cell of lung
Erythroid lineage cell
Fibroblast
Glandular epithelial cell
Glial cell
Granulocyte
Hematopoietic cell
Hepatocyte
Innate lymphoid cell
Intestinal epithelial cell
Lymphocyte of b lineage
Melanocyte
Meso-epithelial cell
Myeloid leukocyte
Neuron
Retinal pigment epithelial cell
Secretory cell
Stem cell
Stratified epithelial cell
Stromal cell
T cell
Transitional epithelial cell
E
UMAP2
UMAP1
Tissue
Bladder
Blood
Bone marrow
Eye
Fat
Heart
Large intestine
Liver
Lung
Lymph node
Mammary
Muscle
Ovary
Pancreas
Prostate
Salivary gland
Skin
Small intestine
Spleen
Stomach
Thymus
Tongue
Trachea
Uterus
Vasculature
F
Senescence-associated pathways
Homophilic Cell Adhesion Via Plasma Membrane Adhesion Molecules
Cell-Matrix Adhesion
Cellular Response To Hypoxia
Vacuolar Acidification
Epidermis Development
Transcription Initiation At RNA Polymerase II Promoter
Proteasomal Protein Catabolic Process
Proton Motive Force-Driven ATP Synthesis
Cellular Respiration
Regulation Of Protein Localization To Plasma Membrane
Protein Catabolic Process
Endoplasmic Reticulum To Golgi Vesicle-Mediated Transport
Signal Peptide Processing
Protein N-linked Glycosylation
Intermediate Filament Organization
Regulation Of G1/S Transition Of Mitotic Cell Cycle
Blood Circulation
Lead genes
PCDH17, DSCAML1, CNTN4
ITGBL1, EMP2, ITGB8
PTGS2, PTGIS, INHBA
ATP6V0D1, ATP6V1D, CLIC4
SCEL, KLK7, SPINT1
PPARGC1A, GTF2H5, ERCC3
PMAIP1, RNF170, DDB1
ATP6V1A, ATP5F1B, ATP5PB
PPARGC1A, COX5A, MT-CO1
VAMP8, LDLRAP1, CAMK2G
NQO1, UCHL3, FYN
SURF4, CREB3L2, TMED1
IMMP2L, HM13, SPCS3
FUT8, GFPT1, OSTC
KRT24, KRT13, KRT6A
CDKN2A, CDKN2B
GUCY1A1, GUCY1B1, STAT1
Endo-epithelial cell
Glandular epithelial cell
Endothelial cell
Epithelial cell
Fibroblast
Innate lymphoid cell
Intestinal epithelial cell
Duct epithelial cell
Myeloid leukocyte
Epithelial cell of lung
T cell
Neuron
Adventitial cell
Contractile cell
Granulocyte
Lymphocyte of b lineage
Stem cell
Stromal cell
Cell type
Normalized Enrichment Score (NES)
0.0
2.5


A
Sex-biased genes in Tabula Sapiens v2
Tissue-cell type resolution only
Tissue resolution only
Both resolution
DE Genes (FDR<0.001)
Male
Female
Bladder
Blood
Bone Marrow
Fat
Heart
Lymph Node
Muscle
Salivary Gland
Spleen
Tongue
Trachea
Vasculature
B
Log (number of sex-biased protein coding genes)
UTY, XIST, LINC00278
DDX3Y, EIF1AY, RPS4Y1
Autosomal-coding
MT-mt
Autosomal-ribosomal
Autosomal-noncoding
ChrY
ChrX
NABP1
MKNK2 HSPA1B
DNAJB1 HSPA1A
Male
Female
Number of Tissue-cell types sharing
LogFC
sex-biased protein coding genes
CDKN2D DOCK5
FCER1G FOLR2 LCN2
PTGDS RNASE1
ACSL1 HSPA1A
HSPH1 IVNS1ABP
N4BP1 PFKFB3
SELENOH USP53
ZNF618
ABCA5 ATP2B1 BID
BTN3A1 CIRBP GPCPD1
GYPC HLA-DPB1 NABP1
PPDPF PSMB9 TC2N
BCL6 BTG1 CD81 CHPT1
CSGALNACT1 CTBP1 DNAJB1
ERN1 HSP90AA1 HSPA1A
HSPA1B HSPD1 HSPH1 IFITM1
JDP2 KCNE1 MKNK2 PABPN1
PDK4 SAMD1 SPTLC2 TCIRG1
ZFP36L2
C
Kendall correlation
0.3
0.8
Contractile cell
Fibroblast
Endothelial cell
Epithelial cell
Lymphocyte of b lineage
T cell
Granulocyte
Myeloid leukocyte
Granulocyte
D
Distribution of concordance scores for DE Genes
Density
Autosomal-coding
Autosomal-ribosomal
chrX
chrY
Autosomal-noncoding
MT-mt
E
Size
5000
10000
>15000
contractile cell
stem cell
granulocyte
fibroblast
lymphocyte of b lineage
stem cell
Hematopoietic cell
myeloid leukocyte
Mean of |logFC for male-biased genes|
Mean of |logFC for female-biased genes|
Bladder
Blood
Bone Marrow
Fat
Heart
Lymph Node
Muscle
Salivary Gland
Spleen
Tongue
Trachea
Vasculature
F
Female-enriched
Male-enriched
UV Response Dn
Regulation Of Cellular Response To Heat
Inflammatory Response
Androgen Response
p53 Pathway
mTORC1 Signaling
TNF-alpha Signaling via NF-kB
Regulation Of Transcription By RNA Polymerase II
Regulation Of T Cell Differentiation
Regulation Of Inclusion Body Assembly
Proton Motive Force-Driven ATP Synthesis
Protein Secretion
Positive Regulation Of Macromolecule Metabolic Process
Positive Regulation Of MAPK Cascade
Positive Regulation Of GTPase Activity
Positive Regulation Of Focal Adhesion Disassembly
Oxidative Phosphorylation
Negative Regulation Of DNA-templated Transcription
Myc Targets V1
Hypoxia
Interferon Gamma Response
IL-2/STAT5 Signaling
G2-M Checkpoint
Fatty Acid Metabolism
Apoptosis
Cellular Respiration
Apical Junction
t cell
epithelial cell
fibroblast
myeloid leukocyte
contractile cell
granulocyte
innate lymphoid cell
lymphocyte of b lineage
erythroid lineage cell
stem cell
hematopoietic cell
endothelial cell

A



B

	Donor ID	Heart Disease History
1	TSP1	No
2	TSP2	Yes. The donor has a cardiovascular mechanism of injury and has been treated for NSTEMI (Non-ST-Elevation Myocardial Infarction), as indicated by the administration of ASA (Aspirin) and heparin. Additionally, the ECG showed ST elevation in AVR with ST depressions in II, III, AVF, V5, V6, which are signs associated with heart problems, directly pointing to a history of coronary artery disease (CAD).
3	TSP3	Yes. Supporting Information: The donor has a history of mitral valve regurgitation, described in the admission course comments, indicating a form of heart disease. Additionally, echo/cardiac cath showed severe mitral valve regurgitation and hemodynamics consistent with sever cardiogenic shock and biventricular pressure overload.