Whole body cell map tracks tissue-specific immune cell accumulation and plasticity loss through ageing

Mangiola S[1,2], Milton M[1], Ranathunga N[1,2,*], Li-Wai-Suen CSN[1,2,*], Yang E[1,*], Hutchison W[1,2], Garnham A[1,2], Iskander J[1], Pal B[3], Yadav V[4], Carey VJ[5], Morgan M[8,#], Bedoui S[6,#], Kallies A[6,#], Papenfuss AT[1,2,8,#,¶]

1. The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia
2. Department of Medical Biology, University of Melbourne, Melbourne, Victoria, Australia.
3. Olivia Newton-John Cancer Research Institute, Heidelberg, Victoria, Australia.
4. Systems Biology of Aging Laboratory, Columbia University; New York, USA.
5. Channing Division of Network Medicine, Mass General Brigham, Harvard Medical School, Harvard University, Boston, USA
6. The Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia.
7. Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, NY, USA
8. Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia.

* These authors contributed equally

\# These authors contributed equally

¶ To whom correspondence should be addressed

Abstract

The presence of immune cells in non-lymphoid tissues throughout the body is vital in the fight against infections and cancer. Yet, a detailed map of immune cell distribution and interactions throughout the body is lacking. To address this gap, we harmonised and annotated 29 million cells across 12,981 single-cell RNA sequencing samples covering 45 anatomical sites to create a comprehensive compositional and communication map of the healthy immune system. This resource represents a 30-fold increase in cell count and a 10-fold increase in tissue coverage compared to the existing state-of-the-art. We used this resource to model compositional changes in the immune system with age and diversity across demographic groups and to investigate the impact of different technologies using a novel multilevel Bayesian model. Our analysis revealed patterns of progressive tissue-specific inflammation, loss of plasticity with age, and significant differences between ethnicities. This study represents a comprehensive healthy reference for precision medicine, immunology, infectious disease and cancer.

Keywords: Human cell atlas, Compositional Analysis, Immune Body Map, Immune Communication, CELLxGENE, single-cell RNA sequencing, Bayes, ageing, ethnicity, sex

Main

The optimal function of the immune system hinges on its decentralised anatomy. While lymphoid organs, such as the spleen or lymph nodes, represent key hubs in which immune cells are activated and interact, many immune cells reside in non-lymphoid organs[1]. The presence of immune cells in non-lymphoid organs is considered vital in the fight against infections and cancer and forms the basis of a fast-growing number of therapeutic approaches[2] and diagnostic tools[3]. Previous multi-tissue characterisations of single cells have been limited to <200,000 cells[4], included only a few tissues[5], and have been mainly focused on cell type classification and early development rather than ageing or diversity. A comprehensive analysis of immune cell distribution at the single-cell level across tissues, age[6], sex[7], and ethnicity[8,9] in health and disease will be invaluable for improving immunotherapy and precision medicine[10], but is currently missing.

Recent large-scale single-cell sequencing initiatives, such as the Human Cell Atlas[7], create an unprecedented opportunity to develop a comprehensive map of immune cells in health and disease across the body, including their distribution, state and communication. Moreover, the number of studies using single-cell sequencing to understand immune cell functions is growing rapidly, providing a continuous stream of additional data for such a map. However, the disparate nature of the individual data sources, the use of different sequencing approaches and variations in experimental designs of the respective studies provide significant barriers to data analyses. Overcoming these significant hurdles requires novel methodologies for data harmonisation and scalable analysis.

Here, we describe a map of immune cell composition in the human body generated by harmonising transcriptomic data from 29 million cells from 12,981 samples into a single comprehensive resource, the CuratedAtlasQuery. A user-friendly interface facilitates the exploration of this resource. Our cross-study immune cell profiling was made possible by reannotating 11 million immune cells using a novel multi-method consensus approach that allowed us to classify cell-type annotations confidence. We used these data to quantify

immune composition and cell communication changes associated with ageing, including loss of immune plasticity and disruption of central immune processes across 31 anatomical locations and a population of 8,400 individuals. Furthermore, we uncovered significant diversity in tissue immune composition across ethnic groups despite substantial sample bias toward individuals with European ancestry in the data collection. This whole-body reference of immune cell composition and communication provides the first resource of its kind, which will facilitate research into precision medicine, immunology, infectious disease and cancer and serve as a comprehensive healthy reference.

Results

Data harmonisation

To comprehensively characterise the immune system across tissues, age, sex and ethnicity, we first harmonised single-cell transcriptomic data from the CELLxGENE database. We collected 28,975,366 cells, spanning 156 tissues (excluding cell cultures) and 12,981 samples and 324 studies (Figure 1A). We standardised sample identifiers, tissue labels (grouping them to 45 based on anatomy; Figure 1A), age, and other metadata variables, and transformed gene-transcript abundance to a common scale, resulting in a harmonised dataset including all human developmental stages from embryonic to geriatric. These data include a wide range of non-lymphoid tissues and anatomical compartments associated with the immune system, such as blood, thymus, bone marrow, spleen and lymph nodes.

The hierarchical dependence structure of these data provides a major hurdle to systematic analysis. To overcome this, we developed and applied a multi-level statistical model[11] to the data (Figure 1B). Data from each tissue is sourced from multiple studies, each containing multiple samples (Figure 1B), with 26.2% of studies including multiple tissues (up to 49). Samples are characterised by differences in age, sex, ethnicity, disease status, and single-cell technologies (e.g. 10X and Smart-Seq2[12]). While our harmonised database includes individuals with underlying conditions or diseases, all following analyses were limited to healthy individuals and primary data (Figure 1C). The median number of cells assayed in each sample differs between technologies, ranging from <10 cells by Patch-Seq to 19,411 by Slide-Seq (Figure 1D, top). The atlas incorporates 19 distinct protocols (Figure 1D,

bottom). Nine used 10X microfluidics, representing 70.2% of all cells. Other highly represented technologies include sci-RNA-seq[13] and Microwell-seq. A total of 814 tissue specimens were assayed with multiple technologies (Figure 1F), making this atlas a powerful resource for technology benchmarking.

The resulting database constitutes a comprehensive resource facilitating large-scale studies of human biology at the single-cell level across human developmental stages and demographic groups.

Demographic trait and tissue biases

To evaluate the strengths and weaknesses of the database, we analysed the abundance of data across demographic traits and tissues. Tissues are not equally represented in the atlas, with two orders of magnitude separating the most and least represented (number of samples; Figure 1E). Blood is the most represented, with 3,582 samples (28% of the atlas), followed by brain and heart with how many samples (together making up 30%). The appendix vermiformis, gall bladder and rectum are among the least represented (together 0.02%). Tissues have differential age distribution (Figure 1H, right), with blood, lung, and kidney covering broad ranges, in contrast to adipose and lymph nodes with a more limited age range.

We also identified biases in ethnicity. Excluding samples of unknown origin, Europeans represent 82.3% of samples, followed by those of Asian descent (9%; Figure 1G). Other ethnic groups, such as African, represent 3%, Hispanic and Greater Middle Eastern individuals represent less than 1% of all samples, while Aboriginal Australians are absent. These biases are amplified across diseases, where non-Europeans lack representation among 74% of diseases in the database. Ethnic biases are also reflected in human development coverage (Figure 1H, left). Europeans are represented across a wide age range (7 months to 96 years old), while most other ethnic groups only consist of samples from adults. These observations help to guide the interpretation of results and indicate the need for future data collection that is representative of human ethnic diversity.

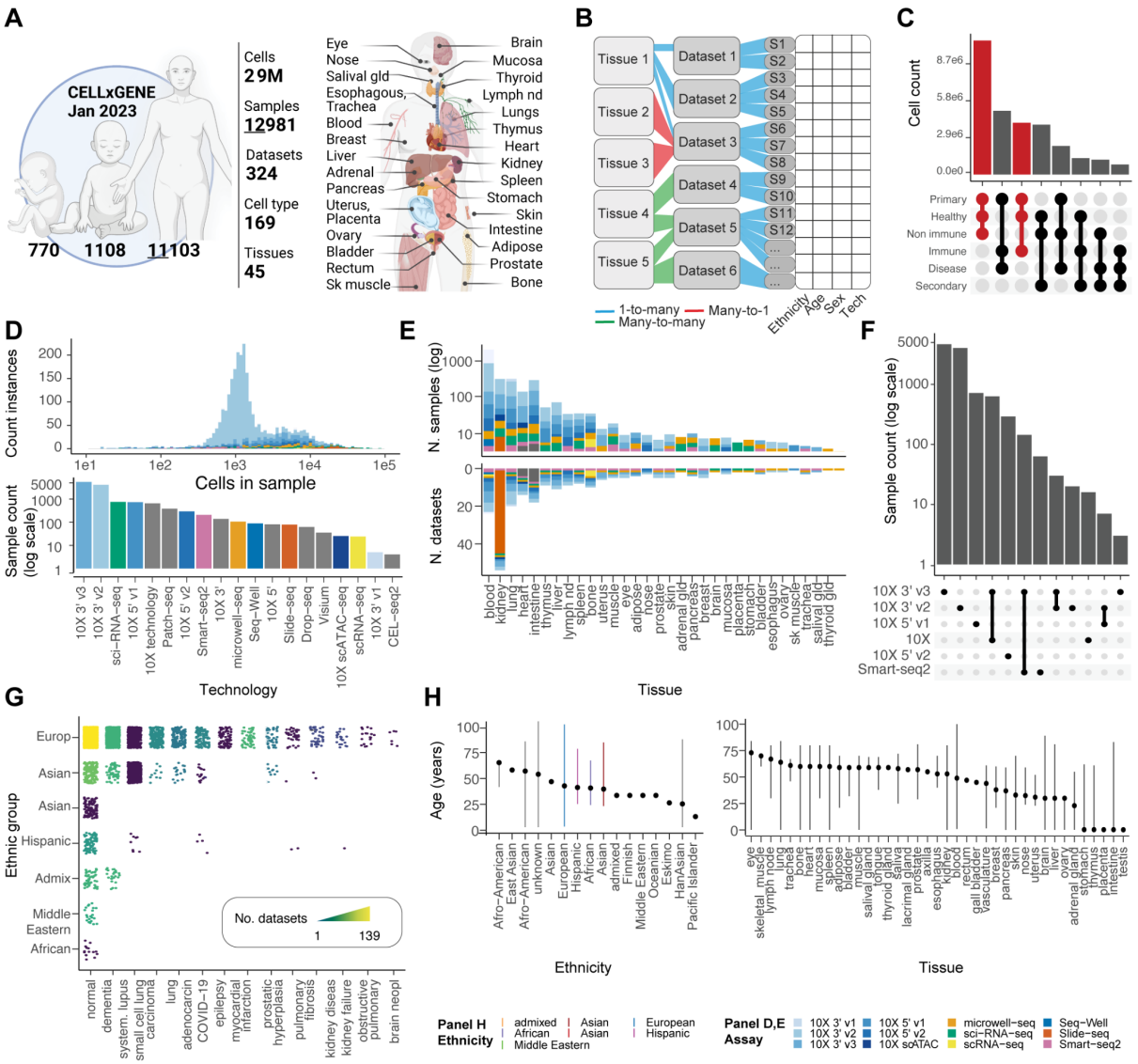

Figure 1. Overview of the CELLxGENE human cell atlas data. **A:** Contents of the harmonised database across ages and tissues. **B:** Schematic showing the multilevel structure of the data. Relationships between data strata are colour coded. Tissues and datasets are modelled here as a hierarchical grouping. **C:** Data selected (red) for statistical analyses. The label "Primary" represents data not reanalysed or uploaded as a separate collection. **D:** Distribution of the number of cells in each sample, coloured by technology (top); and the number of samples by technology (bottom). **E:** Summary statistics of the data content by tissue (using here the harmonised tissue label). The bottom panel y-axis is log-scaled. The bars are coloured by sample processing technology. **F:** Counts of samples across multiple technologies. **G:** Representation of ethnicity across diseases. Dots represent samples. Quadrants are coloured according to the number of datasets. **H:** Age distribution across ethnicities (left) and tissues (right, using here the harmonised tissue label).

Cell-type label confidence varies with technology, demography and tissue type

To model heterogeneity and dynamic transcriptional changes of the immune system across studies, we adopted a consistent immune cell-type ontology (i.e. cell type classes) appropriate for lymphoid and non-lymphoid tissues. We applied a consensus cell labelling strategy to minimise biases in immune cell classification from study-specific standards. The evidence included original and novel annotations, including Azimuth PBMC[14], Blueprint[15] and Monaco[16] references. Our analysis revealed that 87% of cell identities had complete consensus among the original and new annotations (Figure 2B). The lack of consensus for some cells prompted us to score cell identity in confidence tiers (Figure 2B, Methods). High confidence (tier 1) represents a complete consensus of all annotations. In contrast, confidence tiers 2, 3, and 4 indicate a lack of consensus for one, two, or all annotations, respectively. The confidence tier distribution varies across technologies that probe single-cell gene transcription (Figure 2C). While 10X-based datasets, Smart-seq2, Seq-well and micro-well reached a complete consensus for most cells (75%), 10X ATAC-seq, Sci-seq, Slide-seq, and 10X Visium did not reach 25%. Confidence tier distribution is also diverse across ethnicities, tissues, and cell types (Figure 2C, Supplementary Figure S1). Low confidence in cell labelling is not significantly associated with cell library size (Figure S2).

To determine whether the lack of consensus was associated with specific cell types, we ranked the cell type combinations using confidence tiers (Figure 2G). Most cell label disagreements represented activation states and subsets of common cell lineages (e.g., CD4 memory and naive T cells). However, disagreement was also prominent between more distantly related but transcriptionally convergent cell types such as NK and CD8 memory T cells.

Our re-annotation and consensus analyses highlight the utility of consistent ontology and a multi-reference consensus approach. The high-confidence annotation set supports more reliable analyses of the immune system and the creation of a unified body map.

Exploration of the curated and reannotated atlas enabled by CuratedAtlasQuery

To support data access and programmatic exploration of the harmonised atlas, we developed CuratedAtlasQuery (Figure 2J). This interface allows the programmatic exploration of our harmonised database. Cells of interest can be selected based on their ontology, tissue of

origin, demographics and disease. For example, the user can select CD4 T cells across lymphoid tissues in health and diseases, including COVID-19. The data for the selected cells can be downloaded locally into popular single-cell data containers[17]. Pseudobulk counts are also available to facilitate large-scale, summary analyses of transcriptional profiles.

This platform offers a standardised workflow for accessing atlas-level datasets programmatically and reproducibly.

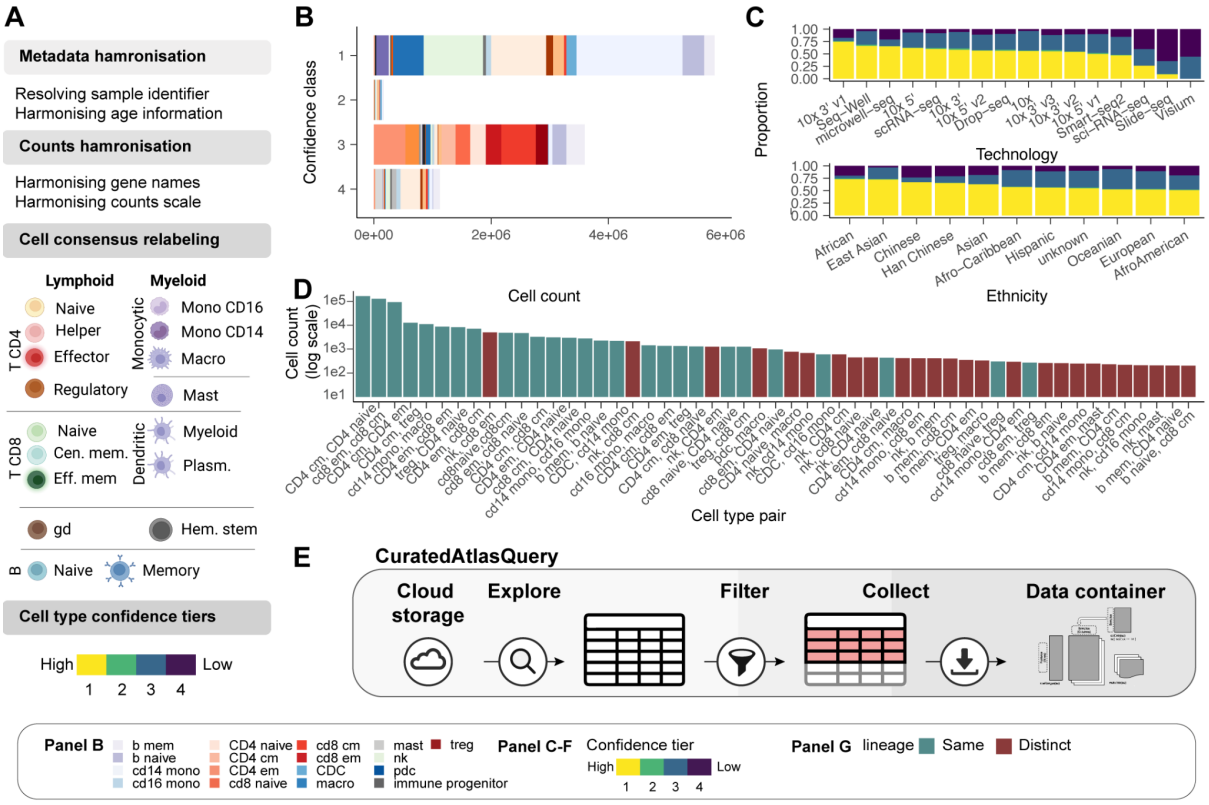

Figure 2. Data curation and query interface. **A:** Schematic of the curation pipeline. **B:** Number of immune cells in each confidence tier. **C:** Proportion of cells in each confidence tier across technologies and ethnicity. **D:** Co-occurrence of cell type annotations (between methods) for cells in confidence tiers 2-4. For example, Azimuth may annotate a cell as CD4 T, while SingleR with Monaco reference may annotate it as CD8 T. **E:** Schematic of the CuraterAtlasQuery interface.

Biases in tissue composition associated with technology

Given the diversity of technologies used to probe cell RNA, we sought to identify biases in cell type representation across tissues independently from other technical and biological factors, including sex, ethnicity, age and dataset (Figure 3, grey box, Supplementary File 1).

The principal component analysis of the cellular composition, adjusted for unwanted variation, shows that the 10x technologies were consistent and did not show significant compositional biases. On the contrary, Smart-seq2, microwell-seq[18] and sci-seq showed significant differences along the second principal components (Figure 3A). These differences highlight the importance of modelling diverse technologies in interpreting compositional analyses. We then investigated which cell types mainly contributed to the technology bias.

Compared to the most common technology, 10x 3' version-2, we detected several significant differences with other assays (Figure 3B). The number of significant differences per technology reflected the distances observed in the principal component plot. Sci-seq and microwell-seq showed the most significant differences with 9 and 7, respectively. The set of cell types having significant differences between these two methods did not show considerable overlap. Despite Smart-seq2 being a fundamentally different technology to 10x, it did not show significant differences at the cell-type level.

Consistency and biases in probing cell types of interest should be considered when choosing the technological platform. Therefore, we tested the differential variability of cellularity composition across tissues (Figure 3C, Supplementary Figure S3). 10x technologies were the most homogeneous (low variability), independent of other biological batch factors (Figure 3, grey box). Similarly, Sci-seq[19] showed low variability, although the uncertainty around the variability estimate is high due to the more limited data availability. Smart-seq2 and microwell-seq show the biggest compositional variability (Figure 3).

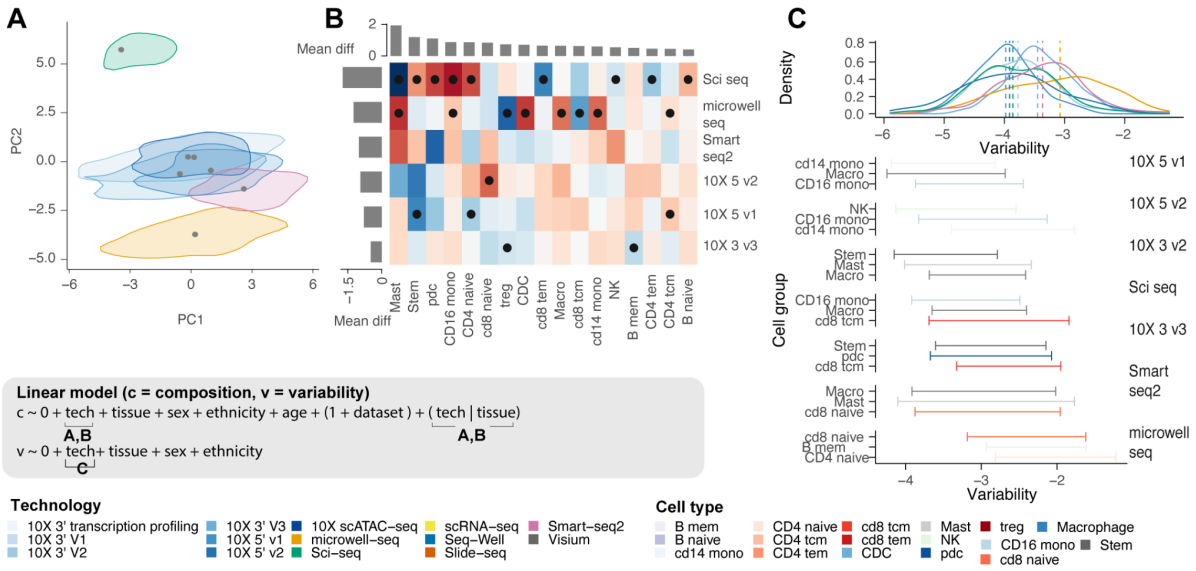

Figure 3. Unwanted variation in composition due to technology. **A:** Principal component analysis of the immune system composition for the seven most represented technologies across tissues, independent of datasets, sex, age, and ethnicity (linear model described in the legend). Dots represent the expected dimensional value. The density areas represent the uncertainty around the expected value (density 0.02). **B:** Biases in cell type relative abundance across technologies compared to the most used 10X 3' v2 technology. The bar plots represent the mean significant absolute change across technologies (rows) and cell types (columns). The dots represent the significant differences (false-discovery rate < 0.05). **C:** Consistency (i.e. differential variability) in cell type relative abundance across technologies independently from datasets, sex, age, ethnicity and tissue. The top panel represents the mean variability across cell types. The dashed lines represent the expected value, while the density lines represent the uncertainty. The bottom panel shows the variability 95% credible interval for the three least consistent (most variable) cell types per technology.

The landscape of immune cell composition across tissues in the healthy immune system

To provide a comprehensive reference of the distribution of immune cells throughout the body, we quantified the immune composition of 31 tissues, excluding major technical (e.g. technology and dataset) and biological (e.g. age, sex, ethnicity) sources of variation (Figure 4A; Supplementary Figure S4 and SE, Supplementary File 1, Data availability section). The principal component analysis (PCA) of immune composition showed the immune cellularity as the main source of variation associated with the first principal component (Figure S5). As expected, blood was located far from all other tissues because of the absence of non-immune cells. Primary and secondary lymphoid tissues, including bone marrow, thymus, spleen, and

lymph nodes, show compositional similarity (Figure 4B). Other similarities include small and large intestines and skeletal and heart muscles.

We then estimated the immune compositional landscape across tissues (Figure 4), with the data modelled to represent a 42-year-old individual of European ethnicity (as it is the most data-rich ethnicity), averaging sex effects. This landscape shows that, on average, monocytic cells (22.5%) are the most abundant across all tissues, followed by CD4 T cells (19.6%), CD8 T cells (18.0%), B cells (9.8%), and NK/ILC (4.9%.). Clustering tissues by composition, we identified 7 major clusters. Cluster 1 is characterised by a relative CD8 T cell enrichment and includes tissues like the ovary, small intestine, and prostate. Cluster 2 is characterised by relative enrichment of Tregs and CD8 T cells, including tissue like the thymus, lymph node, skin, and tongue. Cluster 3 is characterised by classical monocyte and macrophage enrichment but a relatively low overall immune infiltration. Cluster 4 is B-cell enriched and includes tissues of the digestive system, such as the stomach, pancreas, salival gland, oesophagus and appendix. The large intestine, although in a different cluster, is also highly enriched for B cells. Cluster 5 is enriched in CD8 T cells and NK/ILC, including lymphoid tissues such as the spleen and bone, but also the liver, adrenal gland and kidney. Cluster 6 is enriched in CD8 T cells, CD4 T helper 1/17 and T gamma delta, including tissues such as the large intestine, breast, adipose tissue and uterus. Cluster 7 is CD8 T and CD14 monocyte enriched, including blood, heart, and lung.

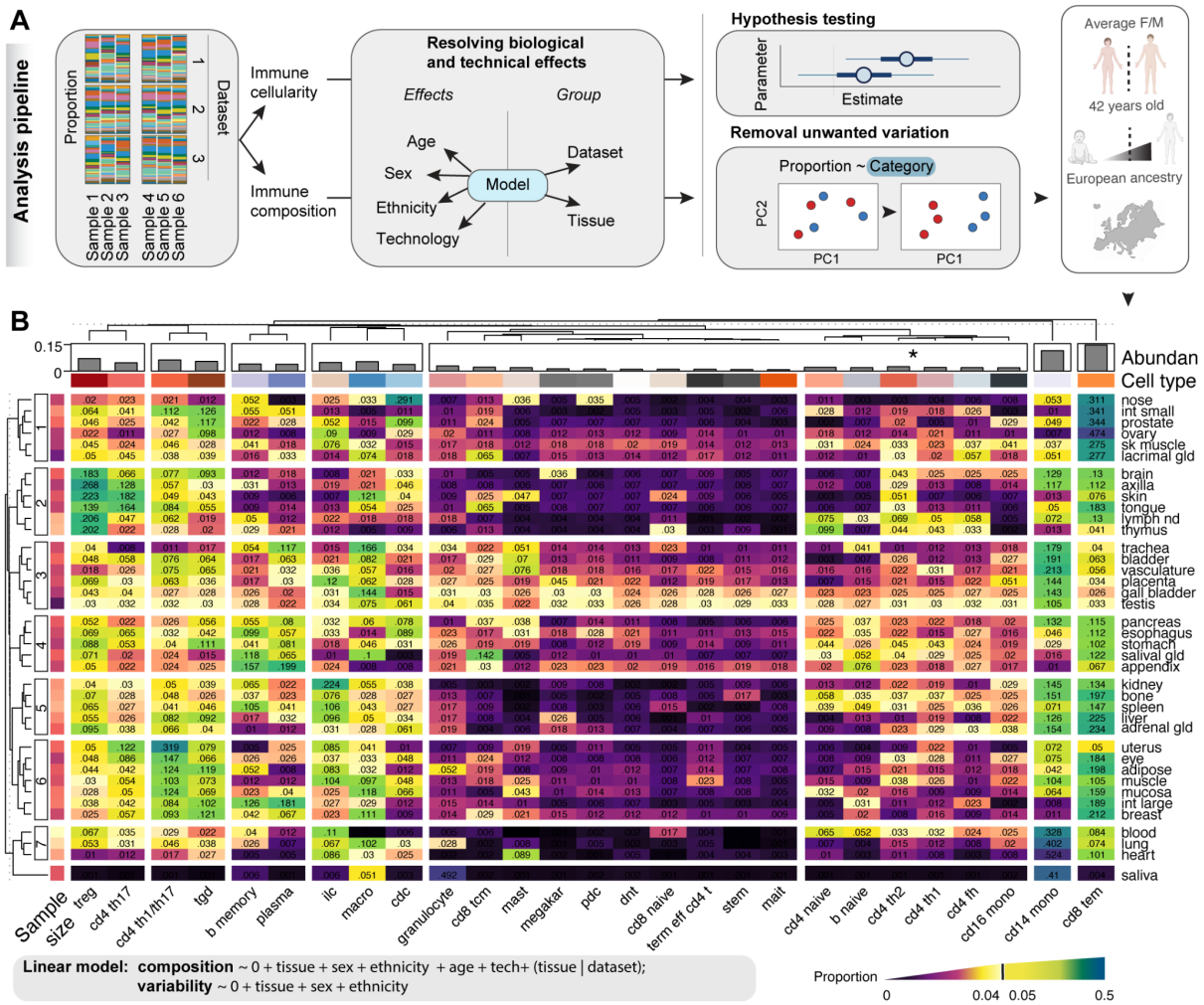

Figure 4. The immune landscape of tissues. **A:** Pipeline for modelling immune tissue cellularity and composition. From left to right, we show data, statistical modelling of biological and technical effects, testing of effect significance, removing unwanted variation, and estimating demographically defined immune profiles. **B:** Estimated proportions of immune cell types (columns) across tissues (rows), factoring out group (random)-level effect of datasets, and sex, age, and ethnicity. Rows are ordered by their hierarchical clustering. Columns are ordered by the mean abundance of a cell type. The model's baseline (intercept terms; white rectangle on the bottom) represents an average between male and female profiles of 42 years-old and European descendants (data-rich), assayed with the 10X droplet technology. The star indicates clusters defined manually from the dendrogram for clarity.

Ageing is associated with global immune cell accumulation

To track the evolution of the immune system through age at the body and tissue levels, excluding other biological or technical sources of variation, we developed a multi-level

compositional analysis model tailored to single-cell data (Figure 5B, Supplementary Methods, Supplementary File 1). We included individuals from 1 to 84 years of age while excluding embryonic stages. Globally, we observe a significant immune accumulation in non-lymphoid tissues, with immune cellularity increasing by 1.2±1.3% per ten years (95% CI; FDR 0.04 for a greater-than-zero effect, and FDR 0.06 for a >1% change from 1 to 84 years of age; Figure 5B).

This trend is reflected locally, with 60% of tissues showing increases in immune cells with age (Figure 5C). The prostate with an 8.9±3.2 fold increase (95% CI; FDR <1x10^-5) between 30 and 83 years of age, the brain with an 8.1±8.4 fold increase between 30 and 75 years of age (95% CI; FDR 0.047), and the kidney with a 3.1±2.5 fold increase over the life-span (95% CI; FDR <1x10^-5.024) are the most immune-infiltrate tissues. In contrast, other tissues show significantly decreased immune cellularity with age. For example, the skin shows a 1.6±0.9 fold decrease (95% CI; FDR 2×10^{-4}) from 22 to 70 years of age, the heart has a modest decrease of 0.6±0.23 fold (95% CI; FDR 2×10^{4}) over the life-span, and the intestine has a 1.9±0.8 fold decrease (95% CI; FDR 1×10^{3}) from 25 to 61 years of age.

These findings recapitulate and quantify key aspects of inflammaging and suggest a major role in immune cell accumulation.

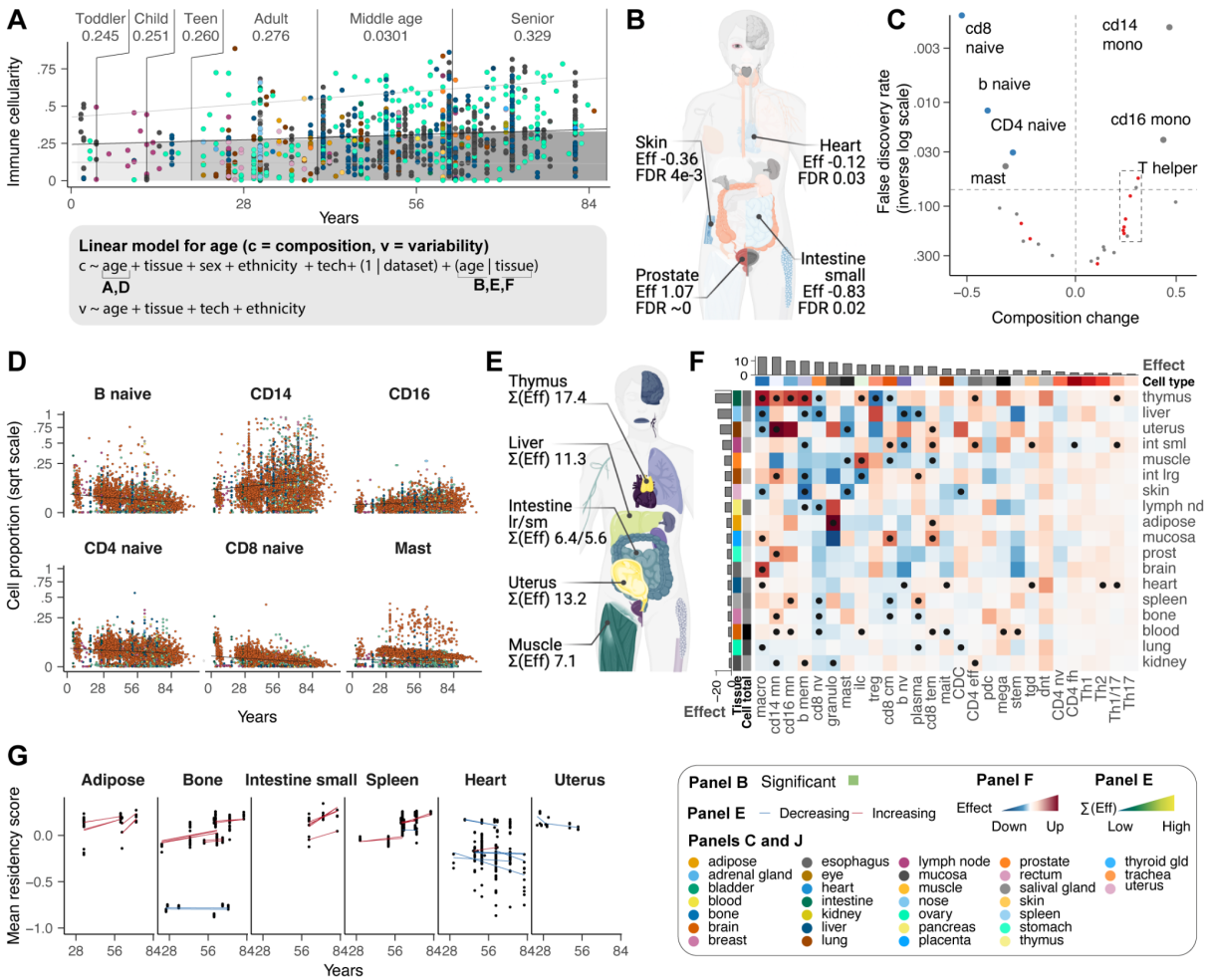

Figure 5. The landscape of immunity changes with age. **A:** Positive association between immune cellularity and age across tissues, independently from probing technology, sex, ethnicity and batch datasets. The statistical model is described within the grey box. The immune cellularity is represented as a proportion adjusted for the unwanted biological and technical variation. The mean estimated immune cellularity is provided for each stage of life. Colours represent tissues/tissue. **B:** Association between immune cellularity and age for each tissue, independent of other biological and technical factors. Colours represent the effect from negative (blue) to positive (red) effects. Significant associations are labelled as text. **C:** Enriched and depleted cell types through age. **D:** Association between immune composition and age across tissues. Only cell types with significant associations (false-discovery rate < 0.05) are displayed. **E:** Average (across cell types) immune compositional changes associated with age for each tissue. Yellow represents large compositional changes. Significant associations are labelled as text. **F:** Detailed view of age-related immune compositional changes for each cell type and tissue. The red colours of the heatmap represent a positive association. Cell types and tissues/tissue are colour labelled as previously described. Histograms represent the magnitude of composition changes (sum of absolute significant effects) across tissue for rows and cell types for columns. **G:** Association between residency

score of T cells and age. Red and blue represent positive and negative associations, respectively. Trend lines are shown across datasets.

Age-related global depletion of naive lymphocytes and mast cells is matched by an increase of antigen-experienced lymphocytes and monocytes

To resolve further the overall immune accumulation in ageing, we estimated age-related changes in immune composition at the body and tissue levels. At the body level, we were able to recapitulate and quantify the well-documented[20–22] differentiation of the lymphoid population away from the naive state (Figure 5, panels D and E). We quantified the reduction of naive B cells of 1.6±1.4 fold (95% CI; FDR 5×10^{-3}), CD4 T cells of 2.2±1.6 fold (95% CI; FDR 2×10^{-3}) and CD8 T cells of 1.3±0.7 fold (95% CI; FDR <1x10^-5). Similarly, mast cells decreased with age, observable after childhood[23] (Figure 5E), with an overall fold-change of 1.9 (95% CI; FDR 2×10^{-2}). The age-related decrease of mast cells is most evident in the uterus, skin and muscle (Figure 5G). In contrast, at the body level, monocytes are the cells that most significantly increase in frequency through ageing, with a fold change of 0.9±0.6 (95% CI; FDR 3×10^{-4}) and 0.7±0.7 (95% CI; FDR 8×10^{-3}) for CD14 and CD16 monocytes, respectively, which is consistent with their role in tissue inflammation[22]. Similarly, CD4 T cells are consistently enriched through ageing (although non-significantly when helper subtypes are tested separately, Figure 5D, dashed rectangle).

This finding recapitulates and quantifies, for the first time at the body scale, some well-known age-related phenotype shifts of immune populations[22], identifies monocytes as the primary contributor to immune accumulation in non-lymphoid tissues, and highlight the significant depletion of mast cells.

Ageing is characterised by changes in immune composition in the thymus and other tissues

To test whether global changes in composition in ageing are consistent across tissues, we build a multi-tissue map of composition and residency. Our analysis at the tissue level revealed marked heterogeneity across anatomical sites and immune cell types. The tissue most affected by age is the thymus (Figure 5, panels F and G). This observation aligns with the well-known atrophy of the thymus[24,25]. The ageing thymus was enriched in monocytes

(CD14, CD16, and macrophages) but not dendritic cells. Memory B cells and plasma cells were also significantly enriched. On the contrary, regulatory T cells (Tregs) were the most depleted cell type through the ageing of the thymus, with a proportionality approaching 0 at old age (FDR <1x10^-5). Although the decrease in immune diversity, the loss of naive T cells, and the qualitative changes of Tregs through ageing[26] are well documented[21,25], the observation that the Treg compartment is characterised by the largest changes is novel. Other significant changes occur in the liver, uterus, muscle and intestine. The most significant age-related change in blood is the enrichment of monocytes and innate lymphocytes (including NK cells).

To investigate molecular contributors to age-related immune enrichment across tissues, we scored the CD4 and CD8 effector T-cell repertoire for residency[27] (Figure 5G). Our results show that the increase in effector T-cell infiltration in the intestine is linked to the upregulation of residency markers. On the contrary, this effect is not observed in adipose tissue, bone marrow, and spleen. These observations are consistent with what is often described as inflammaging[28].

These results provide a tissue-resolved compositional reference for the compositional contribution to inflammaging, which ranks the thymus as the most relevant among the lymphoid tissues.

Cell communication tracks with age-related systemic inflammation and loss of immune plasticity

To characterise and quantify age-related effects at the molecular level that can complement changes in cellular composition, we profiled immune cell communication across tissues. We focussed on tissues represented by a sufficiently large number of samples and with many cells per cell type, including blood, heart, lung, lymph node, kidney, and thymus. For those tissues, we quantified the average strength[29] with which one cell type communicated with all others relying on well-known communication axes (i.e. ligand-receptor pairs). For significant axes, we resolved the communication between specific cell types.

The loss of compositional homeostasis of Tregs in the thymus (Figure 5G) co-occurs with the alteration of essential communication patterns across tissues (Figure 6A). Blood and lymph nodes were the most affected tissues. Beyond communication axes that were

dysregulated globally, some communication axes, including SELPLG and ICAM, were selectively downregulated between Treg and antigen-primed T cells (e.g. CD8 memory and CD4 helper), while they were upregulated toward naive T cells (Figure 6B, Supplementary Figure S6).

Classical dendritic cells and macrophages show a consistent and strong decrease in communication overall with all other cell types (Figure 6C). While we observed an overall decrease in communication for several cell types, we did not observe any consistent increase in communication for any cell type. These changes point toward an overall loss of immune cell communication with age. Blood, heart and lung are the tissues for which we found evidence for the disruption of overall communication for monocytic cells. The downregulated communication axes are mostly linked to dendritic cell migration and monocytic antigen clearance (e.g. ITGB2, ICAM, PECAM1, SELPLG and MIF), antigen presentation[30] (e.g. RETN), cytokine production in dendritic cells[31] (e.g. CD48). The downregulated monocyte communication (Figure 6E, Supplementary Figure S6) primarily targets T cells (e.g. for RETN, CD48) and other monocytes (e.g., for CD22, PECAM1, SELPLG, and MIF).

Our results highlight the age-related alteration of five main immune features, contributing to inflammation and loss of immune specificity and plasticity. These features include Treg thymic abundance and modulation of lymphocytes, monocytic increased tissue abundance, decreased migration, antigen clearance, and antigen presentation (Figure 6E).

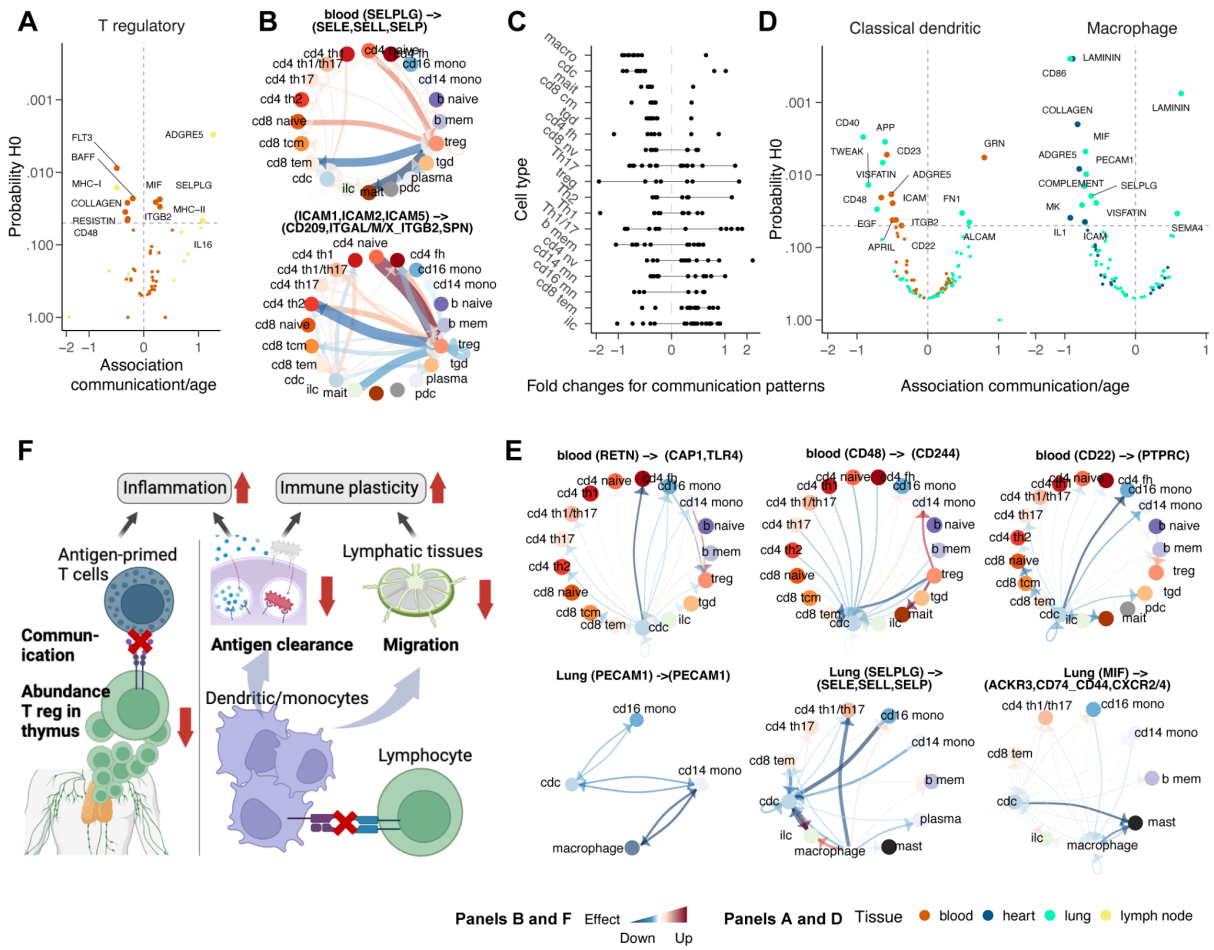

Figure 6. Changes in cell-cell communication with age. **A:** Association between communication axes (which underlie several ligand and receptor genes) transcription abundance and ageing for T regulatory cells. **B:** Estimated changes in communication strength across cell types through ageing, mainly in T regs. Red and blue represent increase and decrease, respectively. **C:** Overall communication changes across cell types. Dots represent communication axes. Lines represent the 95% credible interval. **D:** Association between communication axes transcription abundance and ageing for classical (myeloid) dendritic cells and macrophages. **E:** Estimated changes in communication strength across cell types through ageing, mainly in professional antigen-presenting cells. Red and blue represent increase and decrease, respectively. **F:** Cartoon of the major mechanisms of inflammaging and losing immune plasticity through ageing.

Immune composition profiles are associated with ethnicity more than sex

To test systematically whether sex and ethnic background are drivers of diverse immune profiles, we estimated the association of immune cellularity and composition with these demographic traits. We found no evidence of overall immune cellularity differences between

the sexes (Figure 7A, Supplementary File 1). Similarly, no significant differences in immune cellularity were estimated at the tissue level. In contrast, we identified differences in cell type composition (Figure 7B). Males showed an overall enrichment of antigen-experienced T CD4 cells, including Th2, Th1/17 and Th17, CD4 terminal effector and gamma delta. Females are enriched in immune progenitor cells and megakaryocytes.

Ethnicities show immunological heterogeneity both in cellularity and composition. Across tissues, we observe a small but significant (FDR = 0.028, for effect > 0.1) increase in immune cellularity for European ethnicity compared with the average of the other ethnicities (Figure 7C) and a decrease for African ethnicity (FDR = 0.048). Between ethnicities, the differential variability analysis did not show major differences; however, samples with African ethnicity show a lower variability compared to the average of the other ethnicities. At the tissue level, we estimated ethnic groups to be highly heterogeneous in immune cellularity for the breast, kidney and heart (Figure E). European ancestry shows an increased infiltration compared with the average of the other ethnicities. African ethnicity shows lower immune cellularity in the kidney, while the heart shows diverse immune cellularity for Hispanic (increased) and African (decreased) ethnicity. At the immune-composition level, Asians and Europeans show the most differences compared to other ethnicities (Figure 7F). We observed a consistent enrichment for antigen-experienced T cells, including CD8 and CD4 T cells, and a relative depletion of monocytic populations, including dendritic cells and classical monocytes in the Asian group. In contrast, European ethnicity was relatively depleted in CD4 T helper cells compared to the mean of the other ethnicities.

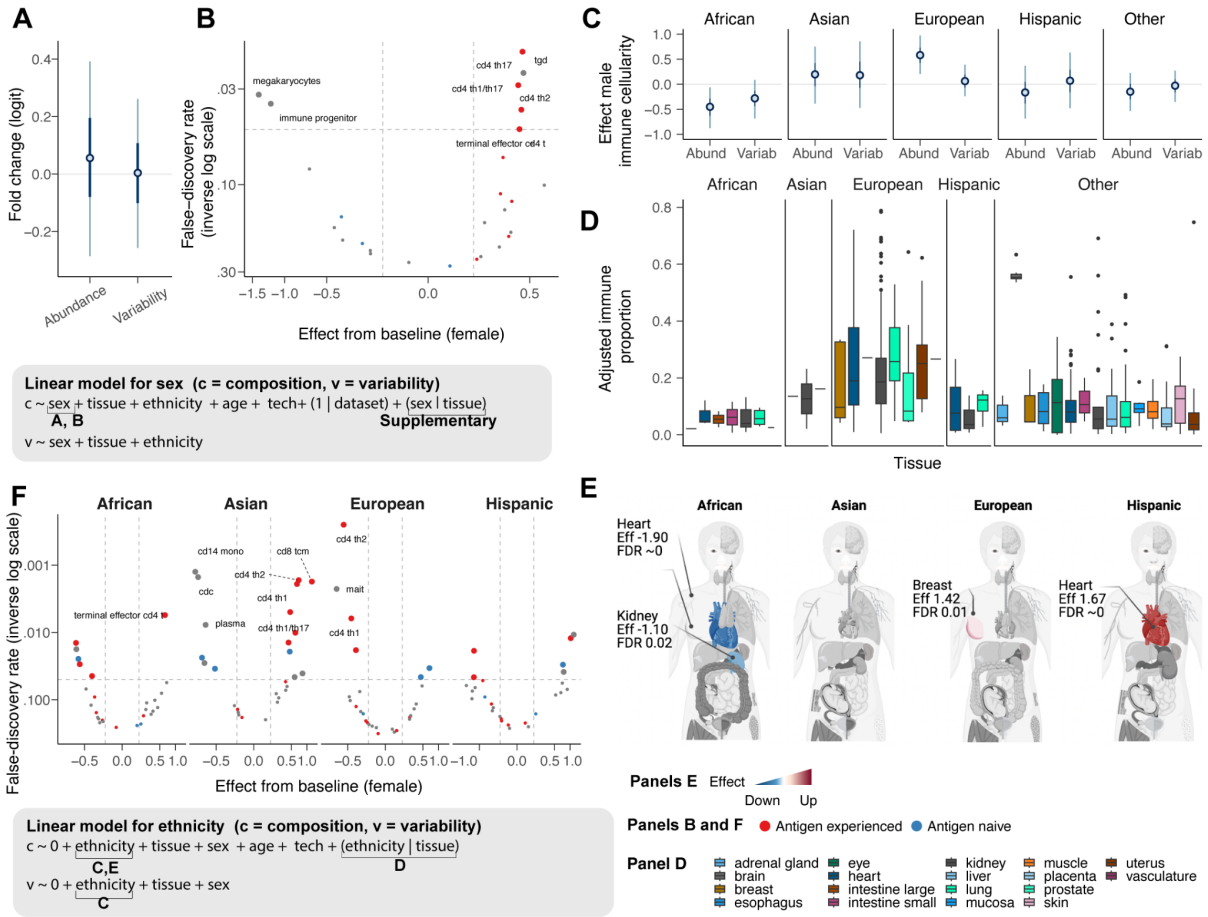

Figure 7. Changes in immune cell composition with sex and ethnicity. **A:** Overall effect of sex on immune cellularity and its variability with females as the baseline. The mean, 50% and 95% credible intervals are visualised. **B:** Tissue-specific differences of immune cellularity between sexes. Left: cartoon of the human body coloured by significant effect. Right: Immune cellularity of samples, with non-sex effects removed. **C:** Cell types differences across tissues and sexes (male compared to a female baseline). Cell-type antigen experience is colour-coded. Size represents significance, and a dashed line marks the threshold of 0.05. **D:** Overall effect on immune cellularity and its variability of ethnicity, compared to the mean of other ethnicities. **E:** Tissue-specific differences of immune cellularity among ethnicities (each compared with the mean of the others). Top: cartoon of the human body coloured by significant effect. Bottom: Immune cellularity of samples, with non-ethnicity effects removed. **F:** Cell types difference, overall across tissues, across ethnicities (each compared with the mean of the others). Size represents significance, and a dashed line marks the threshold of 0.05.

Discussion

This study comprehensively maps the human immune system compiling single-cell data, resolving the effect of ageing, sex and demography. Compared with previous multi-organ studies[4,5], we increased tissue coverage, age range, and ethnicity representation. We modelled a healthy reference body map of immune composition and uncovered immune changes in ageing and across demographics. We share this data with the community through a programmatic query interface in R and Python, CuratedAtlasQuery.

Our immune-cell curation and annotation highlight that while most cells are consistently classified across references, 46% lack complete consensus. This lack is prominent for CD4 and CD8 T cell activation phenotypes and monocyte polarisation (CD14 or CD16). We also observed a frequent lack of consensus between transcriptionally overlapping cytotoxic lymphocytes, including CD8 T cells and natural killer cells. The lack of consensus is not randomly distributed across cell types, tissues, ethnicities or technologies, and this bias could be partly due to a lack of domain-specific reference data. These observations will help guide future studies on cell annotation for challenging cell types or data sources.

We have introduced a method for compositional analyses across a heterogeneous atlas. This method extends sccomp[32], a robust and flexible Bayesian model able to estimate differences in relative abundance and variability in an outlier-insensitive manner. Our method, applied to this large-scale dataset, allowed extracting age, sex and ethnicity-related effects from other sources of biological or technical variability. In our healthy immune body composition model, age emerged as the strongest predictor of composition. Across tissues, we estimated a global accumulation of immune cells in the body. The loss of naive lymphocytes in favour of antigen-experienced phenotypes has been previously described[22] and represents one of the most significant global effects (across tissues). Also, the increase of T cells (CD8 memory) we identified in adipose tissue recapitulates previous findings[33].

Among all tissues, the thymus is the most altered through age in terms of immune composition, recapitulating the well-known progressive atrophy of the thymus[24,25] and providing evidence for the robustness of our study. As the top compositional change, we observed the depletion of regulatory T cells. While the enrichment of regulatory T cells has

been observed in lymph nodes (also seen in this study) and in spleen[34–37], the predominant loss of Tregs from the thymus has not been reported before. This observation is important as the generation of regulatory T cells in the thymus is crucial for immune homeostasis and self-tolerance[38]. We also identified an age-related alteration of cell communication patterns in blood and lymph nodes (Figure 6A); for example, the significant decrease in communication patterns between regulatory T cells and antigen-experienced and the significant increase with naive T cells (e.g. Figure 6B). Cell composition or communication alterations may decrease Treg suppressive function, migration, and localisation within the immune system[39].

A substantial depletion of cell communication emerged for professional phagocytes, including macrophages and dendritic cells in the lung and blood (Figure 6F). The disrupted communication axes included antigen presentation and tissue trafficking. This observation supports that the decreased antigen clearance by macrophages and dendritic cells is a critical mechanism in aging[40–42]. Concurrent with the downregulation of pathways linked to the ability of myeloid cells to clear and present antigens to adaptive immune cells, we observed a significant and global increase in monocyte infiltration across tissues (Figure 5 panels E and F). We provide clarity over the controversy of whether aged dendritic cell subtypes demonstrate specific alterations in antigen presentation capacity[22]. Age-related inflammation (inflamaging[28]) is a well-studied phenomenon[22]; however, its investigation is spread across clinical areas without a systematic and quantitative investigation at the population and tissue levels. Our age-resolved tissue map of the immune system pushes the idea that inflammaging and immune plasticity loss are intertwined and synergic. As non-linear age-related effects have been identified in mice[33] future work will involve modelling such trends in large-scale, hierarchical human cell atlases.

Additionally, this study highlights the existence of biases in data sources and coverage. We show a strong bias across ethnic groups skewed toward European ethnicity. At the same time, we showed that the immune microenvironment is highly diverse across ethnicities[8,9]. The lower variability we estimated for the African cohort can indicate a sample selection that needs to prioritise diversity, which is especially important for underrepresented ethnicities. Independently of these differences being linked to genetics or geographical origin, this finding represents a call for a more representative data collection strategy which would

be required for precision medicine in the coming years[43]. Our analysis also highlights the gaps in the age representation across tissues and can be used to place priorities to improve the current map. For example, blood, lung, brain, kidney, and spleen are well represented from birth to old age. In contrast, among disease-prone tissue locations, the intestine, ovary, thyroid gland and adipose tissue lack age spread.

Despite the robustness of our model to outliers and the exclusion of samples with immune fraction enrichment of depletion by experimental design, some tissue results should be interpreted with caution. For example, estimating immune cellularity for non-lymphoid tissues will be impacted by the depletion of stromal and tissue cells during tissue digestion and library preparation. Similarly, tissue blood contamination may represent a source of bias, such as for the blood, lung, and heart cluster in the compositional landscape (Figure 3). Also, some immune cell types are difficult to sample with single-cell technologies (e.g. neutrophils). Although these challenges impact the estimation of an immune baseline, they will not impact the differential analyses, as tissue or cell-specific tissue extraction efficiencies are taken into account by our models.

Our map of immune cell composition and communication captures the intricate immunological landscape of diverse human populations. Using a Bayesian model to account for the complex dependence structure of the data, we resolved changes in the immune system associated with age, gender, ethnicity and biases due to technology. A key contribution of this work is to make single-cell data in the Human Cell Atlas more accessible to users via our CuratedAtlasQuery interface. This resource has the potential to provide new insights into the healthy immune system. By sharing our harmonised atlas with the scientific community, we hope to catalyse further breakthroughs in cancer, infectious disease, immunology and precision medicine.

Methods

Harmonisation and curation of the CELLxGENE atlas metadata

We downloaded the metadata and gene-transcript abundance from the CELLxGENE database (cellxgene.cziscience.com) through the R package cellxgenedp (10.18129/B9.bioc.cellxgenedp). We then harmonised and merged the metadata from 324

datasets, obtaining a dataset with consistent data types and ontologies. The abundance matrices were harmonised to have the same gene set, following the HUGO Gene Nomenclature[44]. We standardised the gene-transcript abundance distribution, defined in the real positive scale (e.g. not logged or square rooted). We standardised sample identifiers (named `sample_` in our database) using a combination of study-specific identifiers and sample metadata. For example, cells grouped under the same identifiers, but sourced from multiple tissues or individuals of distinct ages, were split accordingly, as belonging to several samples by definition. We consistently represented age as days (named `age_days` in our database). Where a categorical value was provided, we converted it into equivalent days using publicly available reference[45]. For example "adolescent stage" was converted to 15 years old (= 5475 days). To increase the interpretability of results, we grouped 156 tissue labels (excluding nine cell cultures) in 45 tissue classes (named `tissue_harmonised` in our database).

Cell type annotation

Immune cells were labelled using Seurat Azimuth mapping to the PBMC reference[46] and SingleR[16] with the Blueprint[47] and Monaco[48] references[16]. To identify a consensus, we compared and contrasted the high-resolution labels `predicted.celltype.l2` for Seurat Azimuth and `label.fine` Blueprint and Monaco references. When possible, the reference-specific cell-type labels were standardised under a common ontology (Table SX). Where the resolution of transcriptionally similar cell types was uncertain with the given tools, cell types were labelled with a coarser resolution. For example, innate lymphoid and natural killer cells were grouped under "innate lymphoid". The cell type curation was performed to obtain a high confidence, meaningful representation of the immune system's heterogeneity, allowing data-rich cell types whose tissue composition can be modelled probabilistically; rather than aiming for the finest resolution possible.

The original annotation provided by the studies was integrated with the three new annotation sources to identify a total or partial consensus. We created confidence tiers based on how many annotation sources agreed. We gave a cell a confidence tier 1 if all four annotations agreed and a confidence tier of 2, 3, or 4 if 1, 2 or 3 annotations disagreed. Where not even a partial agreement across annotations was found, cells were labelled "other

immune". These cells were included in the compositional modelling but not tested for changes.

Harmonised atlas query interface

To support data access and programmatic exploration of the harmonised atlas, we developed R and Python interfaces (CuratedAtlasQuery, Figure 2J). The query workflow is divided into exploration, filtering and data collection. The metadata is explorable in R using dbplyr (CRAN.R-project.org/package=dbplyr), which follows tidy principles[49]. The cell metadata is displayed in the form of a tidy table[49]. Cells represented in the table can be filtered for any condition based on the information relative to cells (e.g. cell type), samples (e.g. ethnicity), and datasets (e.g. tissue). The filtered table with cell identifiers can be used to download transcript abundance for the cells of interest. The transcript abundance is packaged as SingleCellExperiment (R), Seurat (R) or AnnData (Python) containers. When the container allows it, transcript abundance is saved into H5-based matrices, which can be analysed without overloading the machine's memory[17]. Both the original transcript abundance and counts-per-million can be queried. Pseudobulk counts are also available to facilitate large-scale, summary analyses of transcriptional profiles. The metadata and transcript abundance matrices exist in the cloud; and are hosted in the Object Store in the ARDC Nectar Research Cloud (for CuratedAtlasQuery version 0.99.1). The metadata is stored in the parquet format (parquet.apache.org).

Data preparation and filtering for compositional analyses

The harmonised CELLxGENE atlas metadata was used for cellularity (proportion of immune cells among all cells) compositional analyses of immune cells. We selected primary (no re-analysed data or collections) physiological samples (i.e. no disease). We also excluded samples with less than 30 cells. We excluded erythrocytes and platelets from the analyses as they are not of interest. To increase the sample size per demographic group, we merged Asian descendants (labelled in CELLxGENE as Asian and Chinese). We excluded samples whose age was unknown (n=699) and scaled the age for the remaining samples (n=12282).

For immune cellularity analyses, we only considered non-blood and non-lymphoid tissues. We observed a bimodal distribution for the observed immune cellularity

(arithmetically calculated). The modes of this distribution approach 0 (absence of immune cells) and 1 (absence of non-immune cells), where the valley of the distribution was at 0.75. Although our method is insensitive to outliers, we excluded samples that were experimentally enriched or depleted for immune cells (Supplementary File 2) with observed cellularity bigger than 0.75 to avoid large-scale biases. Similarly, to avoid samples that might have been experimentally depleted of immune cells, we selected those with one or more immune cells. For immune composition analyses, we filtered out non-immune cells. To ensure high-confidence analyses, we selected immune cells with confidence tier 1, 2 or 3.

For modelling immune cellularity, datasets with less than three samples were excluded. To model immune composition, we used more stringent filtering (compared with immune cellularity) for datasets with more than 15 samples, as the data availability (cell counts) for several cell types is lower than for the two immune and non-immune cell classes.

Overview of the analysis design

In total, we produced eight models. Across the four factors: age, sex and ethnicity and technology, we modelled immune cellularity (proportion of immune cells among all cells) and composition. For each of these models, we test two main hypotheses, including the existence of a local (a.k.a. group or random) effect for each tissue and a global (a.k.a. population or fixed) effect at the body level. For the hypothesis testing of changes at the body level, just the population-level effect was considered. In contrast, the testing at the tissue level was done considering the sum of the population-level and group-level effects.

The reason for testing our hypotheses on distinct models (rather than testing all hypotheses on one only) is that distinct hypotheses might require distinct data subsets and intercepts (baseline) terms. For example, the age effect (as detailed below) is tested only in post-embryonic samples (> 1 year after conception); similarly, the sex effect is tested only for samples for which sex is known. Also, in visualising the data with the unwanted variation removed, it is helpful to choose an appropriate intercept (or no intercept) depending on the factors of interest. Additionally, reducing the model complexity for factors of no interest (which vary across hypotheses) is beneficial. This approach improves execution times and model convergence. For example, in testing for age, we do not need to resolve all cell-probing technologies; therefore, we can group the 10X chemistry versions, as the goal is

to capture the variability at the technology level and not to estimate the effect for every chemistry. Following this rationale, the ethnicity label was simplified. The unknown or admixed ethnicities and those with less than 100 samples each were grouped within the label "other", including Eskimo, Finnish, Pacific Islander, Oceanian, and Greater Middle Eastern. Similarly, unless the technology was a factor of interest, the 10X single-cell RNA sequencing chemistries and versions were grouped under the same class.

Unless stated otherwise, datasets (made unique if spanning more than one tissue) were modelled as group-level (a.k.a. random) intercept, and tissue was used as a local (a.k.a. group-level, random) effect for ageing. The global (a.k.a. population-level, fixed) effects were age, tissue, sex, ethnicity, and technology. The statistical inference was performed with a multi-level extension of sccomp (Supplementary Methods[32]) based on the Stan language[50].

Removal of unwanted variation

The removal of unwanted variation is based on the calculation of residuals[51]. The estimates are first produced for the model of choice. A target model is specified for the effects of interest. Then, the residuals are calculated between the observed data and the data generated from the target model. For example, if the fitted model is *~ `factor of interest` + `unwanted variation` + (1 + `factor of interest` | grouping)*, and the effects to preserve are defined as *~ `factor of interest`*, data will be generated from the latter model excluding hierarchical and non-hierarchical unwanted variability. If the effects of interest include random effects, the desired formula is *~ `factor of interest` + (1 + `factor of interest` | grouping)*.

Modelling age effects

To model age association with immune cellularity (proportion of immune cells among all cells) and composition, samples less than one year after conception were excluded as the analysis focused on old-age ageing rather than early development. The variability was modelled to be associated with age and tissue. To visualise the trend of immune cellularity in ageing, we removed all variations in the estimated abundance and variability, except for age and the population effect.

Modelling sex effects

To model sex association with immune cellularity (proportion of immune cells among all cells) and composition, samples with unknown sex were excluded. Tissues with less than one represented sex were excluded. To increase the signal-to-noise ratio, as sex differences are most observed from adolescence, samples less than 15 years old were excluded. The model was set with females as intercept (baseline). The variability was modelled to be associated with sex and tissue. To visualise the trend of immune cellularity between sexes, we removed all variation in the estimated abundance and variability, except for the sex and the population effect.

Modelling ethnicity effects

To model ethnicity association with immune cellularity (proportion of immune cells among all cells) and composition, samples with unknown ethnicity were excluded. The model was set to intercept free. The variability was modelled to be associated with ethnicity and tissue. To visualise the trend of immune cellularity between sexes, we removed all variation in the estimated abundance and variability, except for the sex and the population effect.

Modelling the tissue immune landscape

To create a reference map of immune cellularity and composition across tissues, we estimated those quantities after accounting for unwanted biological (i.e. age, sex and ethnicity) and technical (i.e. technology and dataset) variability. The model was set to intercept free. Datasets were used for estimating group-level tissue effects. The estimated cellularity and immune composition per tissue were calculated to the exclusion of technology and group-level dataset effects. The estimated cellularity was converted to proportion with an inverse softmax transformation. The dimensionality reduction was performed on the original data, and the data with unwanted variation was removed, as described in the Methods subsection (Removal of unwanted variation). Dimensionality reduction was performed with tidybulk[52].

Cell communication analyses

We tested changes in communication patterns in ageing across tissues. To ensure high-confidence analyses, we selected immune cells with confidence tier 1, 2 or 3. After

filtering, for each sample, we excluded cell types with less than ten cells for each sample. Then we excluded samples with less than two remaining immune cell types. The scoring of cell communication was done with CellChat[29] with default parameters. For each sample, we calculated the communication weight across communication axes. We included cell-cell interaction, extracellular-matrix interaction (for solid tissues), and secreted factors. Two weights were calculated as the overall communication strength of a cell type to all other cell types and the communication strength between each cell type pair. To ensure quality results, as the inference of cell communication adds uncertainty to the hypothesis testing, we limited our modelling to the tissues with a large sample size after filtering, for which communication was detected. We included blood, lymph nodes, heart and lung. To account for group-level (a.k.a. random) effects, we used brms for multilevel statistical modelling[11]. We used a beta noise model, given that the weight was bounded between 0 and 1.

Tissue-residency analyses

We scored tissue residency for T cells using Singscore[53] and the residency reference Mckay et al.[27], on the scaled counts (counts per million). For comparative purposes, we estimated the residency score using the signature proposed by XXX[53], which showed a high correlation (Supplementary Figure S7). As residency and exhaustion are related phenomena, we used Singscore to score T cell exhaustion. However, exhaustion was not included in the analyses as no cell showed a higher-than-background score. (Supplementary Figure S8)

Data manipulation and visualisation

The data analysis was performed in R[54]. Data wrangling was done through tidyverse[49]. Single-cell data analysis and manipulation were done through Seurat[55] (version 4.0.1), tidyseurat[56] (version 0.3.0), tidySingleCellExperiment (version 1.9.4), and tidybulk[52] (version 1.6.1). Colour-blind-friendly colouring was aided by dittoseq[57]. Heatmaps were created with ggplot2[58] and tidyHeatmap[59,60]. GGupset was used for upset plots. Parallelisation was achieved through makeflow[61].

Data availability

The count data used for this study can be downloaded using the query interface CuratedAtlasQuery can be found in Bioconductor and at github.com/stemangiola/CuratedAtlasQueryR. The proportion estimates can be found at github.com/stemangiola/immuneHealthyBodyMap

Code availability

The code used for the analyses and the figures can be found at github.com/stemangiola/immuneHealthyBodyMap. Our multilevel model can be found at github.com/stemangiola/sccomp/tree/multilevel.

Acknowledgements

We thank the Human Cell Atlas Consortium for the CELLxGENE data source. We thank the Chan Zuckerberg Initiative for the formatting and organisation of the data used in this study. We thank all individuals who donated tissues. We acknowledge all WEHI's ITS and Research Computing Platform members for their computing support. We thank the Stan community for their constant support. This research was supported by the Nectar Research Cloud, a collaborative Australian research platform supported by the NCRIS-funded Australian Research Data Commons (ARDC). We acknowledge the Chan Zuckerberg Initiative for the data infrastructure we used and the Human Cell Atlas for the data source.

Ethics declarations

Authors have no competing interests to declare.

Contributions

SM conceived, designed and implemented the study and performed the analyses under the equal supervision of ATP, Martin M, SB and AK; Michael M. and YE implemented the CuratedAtlasQuery interface; NR, CL and WS performed data analyses for Figures 1 and 2. VY and BP contributed to the revision of the analyses and the manuscript. WH supported data curation. All authors contributed to the writing of the article.

Funding

S.M. was supported by the Victorian Cancer Agency Early Career Research Fellowship (ECRF21036). A.T.P. was supported by an Australian National Health and Medical Research Council (NHMRC) Senior Research Fellowship (1116955). S.M. and A.T.P. were supported by the Lorenzo and Pamela Galli Next Generation Cancer Discoveries Initiative. The research benefited from the Victorian State Government Operational Infrastructure Support and Australian Government NHMRC Independent Research Institute Infrastructure Support.

References

1. Elmentaite, R., Domínguez Conde, C., Yang, L. & Teichmann, S. A. Single-cell atlases: shared and tissue-specific cell types across human organs. *Nat. Rev. Genet.* **23**, 395–410 (2022).
2. Esfahani, K. *et al.* A review of cancer immunotherapy: from the past, to the present, to the future. *Curr. Oncol.* **27**, S87–S97 (2020).
3. Schofield, J. B. & Wright, N. A. Immunodiagnosis of Cancer (second edition). *Br. J. Cancer* **66**, 1205–1205 (1992).
4. Karlsson, M. *et al.* A single-cell type transcriptomics map of human tissues. *Sci Adv* **7**, (2021).
5. Suo, C. *et al.* Mapping the developing human immune system across organs. *Science* **376**, eabo0510 (2022).
6. Sadighi Akha, A. A. Aging and the immune system: An overview. *J. Immunol. Methods* **463**, 21–26 (2018).
7. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).
8. González, L. A., Toloza, S. M. A., McGwin, G., Jr & Alarcón, G. S. Ethnicity in systemic lupus erythematosus (SLE): its influence on susceptibility and outcomes. *Lupus* **22**, 1214–1224 (2013).
9. Clark, L. T. *et al.* Increasing Diversity in Clinical Trials: Overcoming Critical Barriers. *Curr. Probl. Cardiol.* **44**, 148–172 (2019).
10. Rood, J. E., Maartens, A., Hupalowska, A., Teichmann, S. A. & Regev, A. Impact of the Human Cell Atlas on medicine. *Nat. Med.* **28**, 2486–2496 (2022).
11. Bürkner, P.-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).
12. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

13. Martin, B. K. *et al.* An optimized protocol for single cell transcriptional profiling by combinatorial indexing. *arXiv [q-bio.GN]* (2021).
14. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
15. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
16. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
17. HDF5Array. *Bioconductor* https://bioconductor.org/packages/release/bioc/html/HDF5Array.html.
18. Chen, H. *et al.* High-throughput Microwell-seq 2.0 profiles massively multiplexed chemical perturbation. *Cell Discov* **7**, 107 (2021).
19. Martin, B. K. *et al.* Optimized single-nucleus transcriptional profiling by combinatorial indexing. *Nat. Protoc.* (2022) doi:10.1038/s41596-022-00752-0.
20. Lazuardi, L. *et al.* Age-related loss of naïve T cells and dysregulation of T-cell/B-cell interactions in human lymph nodes. *Immunology* **114**, 37–43 (2005).
21. Goronzy, J. J., Fang, F., Cavanagh, M. M., Qi, Q. & Weyand, C. M. Naive T cell maintenance and function in human aging. *J. Immunol.* **194**, 4073–4080 (2015).
22. Mogilenko, D. A., Shchukina, I. & Artyomov, M. N. Immune ageing at single-cell resolution. *Nat. Rev. Immunol.* **22**, 484–498 (2022).
23. Wilgus, T. A., Ud-Din, S. & Bayat, A. A Review of the Evidence for and against a Role for Mast Cells in Cutaneous Scarring and Fibrosis. *Int. J. Mol. Sci.* **21**, (2020).
24. Pearse, G. Histopathology of the thymus. *Toxicol. Pathol.* **34**, 515–547 (2006).
25. Goronzy, J. J., Lee, W.-W. & Weyand, C. M. Aging and T-cell diversity. *Exp. Gerontol.* **42**, 400–406 (2007).
26. Yang, S., Fujikado, N., Kolodin, D., Benoist, C. & Mathis, D. Immune tolerance. Regulatory T cells generated early in life play a distinct role in maintaining self-tolerance. *Science* **348**, 589–594 (2015).
27. Mackay, L. K. *et al.* Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science* **352**, 459–463 (2016).
28. Pan, W. Chapter 7 - Aging and the immune system. in *Molecular, Cellular, and Metabolic Fundamentals of Human Aging* (eds. Fang, E. F., Bergersen, L. H. & Gilmour, B. C.) 199–224 (Academic Press, 2023).
29. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).

30. Pine, G. M., Batugedara, H. M. & Nair, M. G. Here, there and everywhere: Resistin-like molecules in infection, inflammation, and metabolic disorders. *Cytokine* **110**, 442–451 (2018).
31. Morandi, B. *et al.* Distinctive lack of CD48 expression in subsets of human dendritic cells tunes NK cell activation. *J. Immunol.* **175**, 3690–3697 (2005).
32. Mangiola, S. *et al.* Robust differential composition and variability analysis for multisample cell omics. *bioRxiv* 2022.03.04.482758 (2022) doi:10.1101/2022.03.04.482758.
33. Schaum, N. *et al.* Ageing hallmarks exhibit organ-specific temporal signatures. *Nature* **583**, 596–602 (2020).
34. Han, G.-M., Zhao, B., Jeyaseelan, S. & Feng, J.-M. Age-associated parallel increase of Foxp3(+)CD4(+) regulatory and CD44(+)CD4(+) memory T cells in SJL/J mice. *Cell. Immunol.* **258**, 188–196 (2009).
35. Lages, C. S. *et al.* Functional regulatory T cells accumulate in aged hosts and promote chronic infectious disease reactivation. *J. Immunol.* **181**, 1835–1848 (2008).
36. Nishioka, T., Shimizu, J., Iida, R., Yamazaki, S. & Sakaguchi, S. CD4+CD25+Foxp3+ T cells and CD4+CD25-Foxp3+ T cells in aged mice. *J. Immunol.* **176**, 6586–6593 (2006).
37. Sharma, S., Dominguez, A. L. & Lustgarten, J. High accumulation of T regulatory cells prevents the activation of immune responses in aged animals. *J. Immunol.* **177**, 8348–8355 (2006).
38. Hsieh, C.-S., Lee, H.-M. & Lio, C.-W. J. Selection of regulatory T cells in the thymus. *Nat. Rev. Immunol.* **12**, 157–167 (2012).
39. Raffin, C., Vo, L. T. & Bluestone, J. A. Treg cell-based therapies: challenges and perspectives. *Nat. Rev. Immunol.* **20**, 158–172 (2020).
40. Aprahamian, T., Takemura, Y., Goukassian, D. & Walsh, K. Ageing is associated with diminished apoptotic cell clearance in vivo. *Clin. Exp. Immunol.* **152**, 448–455 (2008).
41. Chougnet, C. A. *et al.* Loss of Phagocytic and Antigen Cross-Presenting Capacity in Aging Dendritic Cells Is Associated with Mitochondrial Dysfunction. *J. Immunol.* **195**, 2624–2632 (2015).
42. Butcher, S. K. *et al.* Senescence in innate immune responses: reduced neutrophil phagocytic capacity and CD16 expression in elderly humans. *J. Leukoc. Biol.* **70**, 881–886 (2001).
43. Cheung, A. T. M. *et al.* Racial and ethnic disparities in a real-world precision oncology data registry. *NPJ Precis Oncol* **7**, 7 (2023).
44. Seal, R. L. *et al.* Genenames.org: the HGNC resources in 2023. *Nucleic Acids Res.* **51**, D1003–D1009 (2023).
45. Balasundaram, P. & Avulakunta, I. D. *Human Growth and Development*. (StatPearls Publishing,

2023).
46. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
47. Fernández, J. M. *et al.* The BLUEPRINT Data Analysis Portal. *Cell Syst* **3**, 491–495.e5 (2016).
48. Xu, W. *et al.* Mapping of γ/δ T cells reveals Vδ2+ T cells resistance to senescence. *EBioMedicine* **39**, 44–58 (2019).
49. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).
50. Carpenter, B. *et al.* Stan: A Probabilistic Programming Language. *Journal of Statistical Software* vol. 76 Preprint at https://doi.org/10.18637/jss.v076.i01 (2017).
51. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* **2**, lqaa078 (2020).
52. Mangiola, S., Molania, R., Dong, R., Doyle, M. A. & Papenfuss, A. T. tidybulk: an R tidy framework for modular transcriptomic data analysis. *Genome Biol.* **22**, 42 (2021).
53. Bhuva, D. D. *et al.* Using singscore to predict mutation status in acute myeloid leukemia from transcriptomic signatures. *F1000Res.* **8**, 776 (2019).
54. Becker, R. A., Chambers, J. M. & Wilks, A. R. The new S language. *Pacific Grove, Ca.: Wadsworth & Brooks, 1988* (1988).
55. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
56. Mangiola, S., Doyle, M. A. & Papenfuss, A. T. Interfacing Seurat with the R tidy universe. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab404.
57. Bunis, D. G., Andrews, J., Fragiadakis, G. K., Burt, T. D. & Sirota, M. DittoSeq: Universal user-friendly single-cell and bulk RNA sequencing visualization toolkit. *Bioinformatics* **36**, 5535–5536 (2020).
58. Villanueva, R. A. M. & Chen, Z. J. ggplot2: Elegant Graphics for Data Analysis (2nd ed.). *Measurement: Interdisciplinary Research and Perspectives* vol. 17 160–167 Preprint at https://doi.org/10.1080/15366367.2019.1565254 (2019).
59. Mangiola, S. & Papenfuss, A. T. tidyHeatmap: an R package for modular heatmap production based on tidy principles. *Journal of Open Source Software* **5**, 2472 (2020).
60. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
61. Albrecht, M., Donnelly, P., Bui, P. & Thain, D. Makeflow: a portable abstraction for data

intensive computing on clusters, clouds, and grids. in *Proceedings of the 1st ACM SIGMOD Workshop on Scalable Workflow Execution Engines and Technologies* 1–13 (Association for Computing Machinery, 2012).