Single-cell atlas of the human immune system reveals sex-specific dynamics of immunosenescence

Maria Sopena-Rios[1,2*], Aida Ripoll-Cladellas[1,2*], Fatemeh Omidi[3], Sara Ballouz[4], Jose Alquicira-Hernandez[5,6,7,8,9], Roy Oelen[10], Lude Franke[10], Monique G.P. van der Wijst[10], Joseph E. Powell[9,11] & Marta Melé[1$]

1. Life Sciences Department, Barcelona Supercomputing Center (BSC), Barcelona, Spain
2. Universitat de Barcelona, Barcelona, Spain
3. Research and Development Department, qGenomics Laboratory, Barcelona, Spain.
4. School of Computer Science and Engineering, University of New South Wales, Sydney, Australia
5. Center for Data Sciences, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.
6. Divisions of Rheumatology and Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA
7. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
8. Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
9. Translational Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia.
10. Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands
11. UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, Australia

*Contributed equally

$Correspondence: marta.mele@bsc.es

Abstract

Immunosenescence, the gradual deterioration of the immune system with age, leads to an increased susceptibility to a range of diseases associated with immune dysfunction. Notably, sex is an important variable underlying how immune aging unfolds, as, for instance, autoimmunity develops with aging differently between males and females. Even though some clinical and molecular differences have been identified between male and female immunosenescence, it is not known to what extent sex affects the dynamic composition of immune cells over time. Here, we analyze a large single-cell RNA-sequencing dataset of peripheral blood mononuclear cells from a sex-balanced cohort of 982 human donors providing novel transcriptional and cellular insights into immune aging at an unprecedented resolution. We uncover that aging induces cell type-dependent and sex-specific transcriptional shifts that translate into a differential abundance of distinct immune cell subpopulations. These shifts predominantly involve translation-related genes, indicating a strong link between transcriptional and translational throughput with cell function and consequent immune cell composition. This sexual dimorphism overlaps known autoimmune disease-related genetic variants and results in the differential enrichment of functionally distinct immune populations. Specifically, we uncover that a cytotoxic CD8+ T effector memory subpopulation with an NK-like phenotype accumulates with age only in females and identify a distinct B cell subpopulation that expands with age exclusively in males. These cell subpopulations represent novel sex-specific hallmarks of immune aging. Our findings underscore the hidden complexity of immune aging and demonstrate the value of high-resolution, single-cell analyses in large population cohorts. This research paves the way for future sex-specific interventions targeting immunosenescence to ultimately promote a personalized approach to foster healthy aging.

Introduction

Aging of the immune system, or immunosenescence, refers to the gradual decline of the immune system, which predisposes to multiple diseases, including infection, cancer, autoimmune, and vascular diseases[1,2,3]. This decline hampers the capacity to mount an effective immune response against

pathogens and is associated with a persistent low-level inflammatory state[4]. Given the pervasive influence of the immune system throughout the body, immunosenescence also contributes to age-related alterations in other tissues[5,6]. As a result, the immune system is emerging as a potential biomarker for aging, serving as a proxy to predict biological aging and longevity[7–9], and it has a great potential to develop healthcare strategies towards healthy aging[10].

The age-related decline of the immune system is driven by a combination of changes in the composition of immune cell types and their functional decay[11–16]. However, most studies to date have focused on only one of these aspects. For instance, aging shifts the balance from naive to more memory T and B cells[16–19], and certain cytotoxic T cell subtypes become more prevalent[11–14]. Moreover, frail elderly individuals, who are at an increased risk of adverse health outcomes, accumulate a specific monocyte subset marked by the high expression of specific long non-coding RNAs[12]. Research on transcriptional changes also revealed that supercentenarians have delayed age-related expression changes[20], while frail elderly individuals often show signs of T cell exhaustion[12]. Finally, a recent study demonstrated gene expression and abundance changes are mostly in T cells[16]. However, these studies are limited by small sample sizes, and thus, lack the statistical power to detect changes in minor immune cell types[11,12,16,20]. This narrow focus overlooks the diversity of immune subpopulations, making the high-resolution potential of single-cell transcriptomics essential to capture key processes of immunosenescence.

Historically, immunosenescence research has focused more on male subjects[21]. Consequently, and despite growing recognition of sex differences in immunosenescence[22,23], many studies still disregard these disparities[14,16,24]. Importantly, males and females have distinct immune responses. Females typically have stronger immune responses, which confers them greater resistance to infections and improved response to vaccinations[25,26]. However, it also results in 80% higher incidence of autoimmune diseases[22,27]. Additionally, aging also introduces sex-specific susceptibilities to other diseases, such as cardiovascular disorders[22]. Recent work showed that males exhibit an age-dependent increase in proinflammatory responses and NK-mediated immunity[11], accompanied by a more pronounced decline in B-cell specific genes and higher chromatin remodeling[23]. Conversely, females have a more pronounced aging T cell transcriptional signature of increased cytotoxicity and loss of stemness[28]. However, whether immune subpopulations experience sex-specific changes with age is currently unknown. To address this gap, larger, more inclusive studies are needed to explore the transcriptional and cellular landscape of immune aging, with a focus on specialized subpopulations and sex-specific differences.

In our study, we perform a large-scale characterization of the impact of aging on the circulating immune system in over one million peripheral blood mononuclear cells (PBMCs) from 982 individuals across sexes (416 males and 565 females)[29]. We characterize the main transcriptional and abundance changes with age across 24 immune cell subpopulations. We identify a divergent aging transcriptional signature across cell types that is strongly associated with changes in cell composition. Moreover, we show that these cell type composition changes with age are in some instances sex-specific and may contribute to disease susceptibility differences between sexes. Our findings offer promising avenues for developing strategies to enhance healthy immune aging and mitigate immune age-related diseases.

Results

Most immune cell types show expression shifts towards either up- or down-regulation with age

To explore the impact of aging on the circulating immune system at single-cell resolution, we analyzed single-cell RNA sequencing (scRNA-seq) data of over one million PBMCs from 982 donors from the OneK1K cohort[29] **(Fig. S1).** To find global patterns of immune aging that are common between sexes and maximize statistical power, we initially analyzed both males and females together. To identify age-related gene expression changes, we performed differential gene expression analysis with age, controlling for biological (sex) and technical (sequencing pool) covariates, within each of the 24 immune

cell subpopulations identified by Azimuth[30,31] (see **Methods** and **Fig. S1A-C, Supplementary Table 1, Supplementary Table 2)**. We identified age-differentially expressed genes (age-DEGs) in almost all tested cell types (18 out of 24 cell types, 75%) (False Discovery Rate (FDR) < 0.05) **(Fig. 1A, Supplementary Table 2).** Naive CD8+ T have the largest number of age-DEGs (2,805 DEGs, FDR < 0.05), followed by naive CD4+ T cells (1,923 DEGs, FDR < 0.05), which is the most abundant cell type. These results are consistent when downsampling to the same number of donors per cell type (see **Methods** and **Fig. S2A, Supplementary Table 2**). We observe that age-DEGs are significantly biased towards either up- or down-regulation in most cell types (15 out of 18 cell types, 83%) (Binomial test, FDR < 0.05) (**Fig. 1A)**. The most significant biases are in CD8+ T naive, B memory, and mucosal-associated invariant T (MAIT) cells, which all have two orders of magnitude more down-regulated age-DEGs than up-regulated (Binomial test; CD8+T naive-FDR = $2.22 \cdot 10^{-308}$, B memory-FDR = $7.75 \cdot 10^{-234}$, and MAIT-FDR = $1.06 \cdot 10^{-131}$). Conversely, Natural Killers (NK), CD8+ T effector memory (CD8+ TEM) cells and CD16+ Monocytes have ten times more up-regulated age-DEGs than down-regulated (Binomial test; NK-FDR = $3.72 \cdot 10^{-109}$, CD8+ TEM-FDR = $8.56 \cdot 10^{-120}$, CD16+ Monocyte = $2.99 \cdot 10^{-163}$).

We next explored whether these age-DEGs were shared across cell types. We found that 53% of the age-DEGs (2,407 out of 4,555 unique age-DEGs, 53%) were differentially expressed in at least two cell types. However, in many cases, shared age-DEG did not change in the same direction. For example, 58% of genes shared in more than six cell types have expression changes in opposite directions across cell types (hereafter referred to as 'shared non-concordant') **(Fig. 1B,C, Supplementary Table 2).** Clustering cell types by the proportion of shared age-DEGs that were concordant identified two distinct clusters (see **Methods** and **Fig. 1D**). One cluster contained cell types that had a bias towards up-regulation with age (NK, CD8+ TEM, CD4+ T central memory (CD4+ TCM), CD4+ T cytotoxic, Monocytes, and Treg cells) and another cluster with cell types down-regulated (naive CD8+ and CD4+ T, CD8+ TCM, CD4+ TEM, MAIT, and B cells) **(Fig. 1A).** Notably, these two clusters of cell types are independent of cell lineage and type of immune response (i.e., innate or adaptive immunity).

These results suggest that aging induces divergent gene expression shifts that may impact the expression of the same genes in different directions depending on the cell type.

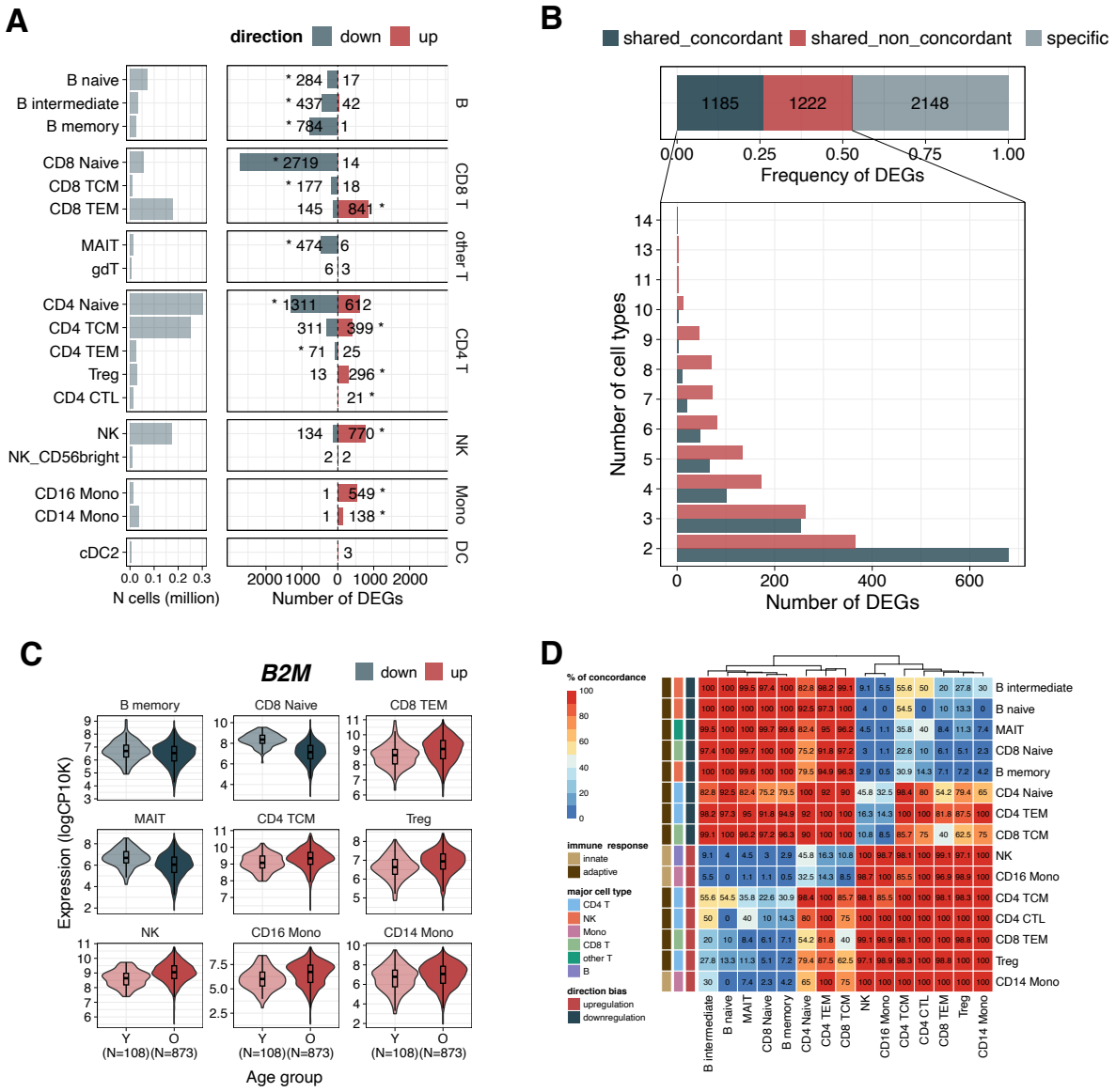

Fig. 1. Aging changes expression of the same genes across cell types in opposite directions. **A.** Left. Total number of cells analyzed per cell type. Right. Number of up- and down-regulated age-DEGs (FDR < 0.05) per cell type. Asterisk indicates a significant bias towards up- or down-regulation (Binomial test; FDR < 0.05). **B.** Top. Proportion of cell type-specific age-DEGs or shared concordant (same direction) and non-concordant (different direction) age-DEGs among cell types. Bottom. Bar plot of the number of age-DEGs shared (concordant and non-concordant) among cell types. **C.** Gene expression of *B2M* gene, a highly cell type-shared age-DEGs (age-DEGs in nine cell types). For visualization purposes, the donors have been classified into young (Y) (≤ 40 years old) and old (O) (> 40 years old) categories. **D.** Heatmap of the percentage of cell type-shared age-DEGs that are concordant between cell type pairs. Cell types (rows and columns) are sorted according to hierarchical clustering analysis. The row annotation bars show specific cell type-features, such as the major cell type the cell type belong to, the type of immune response, and whether the cell type showed an up- or down-regulation bias with age in A.

Age-expression shifts are associated with changes in gene expression throughput and immune functional decay

The observation that aging impacts the same genes in opposite directions depending on the cell type was a bit unexpected and thus, we next wanted to further examine the function of those genes. To do this, we selected age-DEGs shared in a non-concordant manner across at least six cell types and performed hierarchical clustering on the magnitude of age-expression changes (i.e., age-log fold change (FC)) across cell types (see **Methods**). We identified two main gene clusters **(Fig. 2A)**. One gene cluster was enriched for translation-related processes, while the other was associated with immune activation (Gene Ontology (GO) Enrichment; FDR < 0.05) **(Fig. 2B, Supplementary Table 2)**. Notably, translation-related genes in the first cluster were consistently up-regulated in the cell types with the age-up-regulation shift and down-regulated in those with a shift towards down-regulation **(Fig 1A)**. These results are robust to selecting age-DEGs shared in different numbers of cell types (see **Methods** and **Fig. S2B-E)**. We replicated these findings in an external sex-balanced scRNAseq PBMCs dataset from 72 healthy donors[32] with a similar age distribution (mean±SD = 56.83 ± 12.73) to our discovery cohort [32], where cell types also cluster into two similar groups (Cophenetic, R = 0.74) (see **Methods** and **Fig. S2F)** and the shared age-DEGs are also enriched in ribosomal pathways and immune activation genes **(Fig. S2G, Supplementary Table 2)**. The observation that changes in age-expression bias always involves a respective up- or down-regulation of the translation machinery could indicate changes in the transcriptional and translational throughput of the cells with age.

Then, we aimed to explore the functions of the cell type-specific age-DEGs (i.e., age-DEGs only in one cell type) as immunosenescence is associated with functional decay of specific immune cell types. The majority of cell types (12 out of 18 cell types with at least one age-DEG, 67%) had cell type-specific age-DEGs **(Fig 2C).** We then performed functional enrichment analyses and found significant enrichments for up-regulated age-DEGs in CD14+ Monocytes and for down-regulated age-DEGs in five immune cell subpopulations **(Fig 2D, Supplementary Table 2)**. Up-regulated CD14+ Monocytes-specific age-DEGs are enriched in terms related to pattern recognition receptor activity, consistent with previous work showing alterations in pattern recognition receptors responses in Monocyte subsets from old individuals[33]. Down-regulated NK-specific age-DEGs are enriched in leukocyte migration, a critical function for NK defense mechanisms[34] which is known to decay with age in mice[35,36]. Lastly, B intermediate cells are depleted in B cell activation pathways, consistent with an age-associated decline in B cell functionality[37].

Overall, we observe shared age-related changes in transcriptional and translational throughput across immune cell types, along with gene expression alterations that likely contribute to cell type-specific functional decay with age.

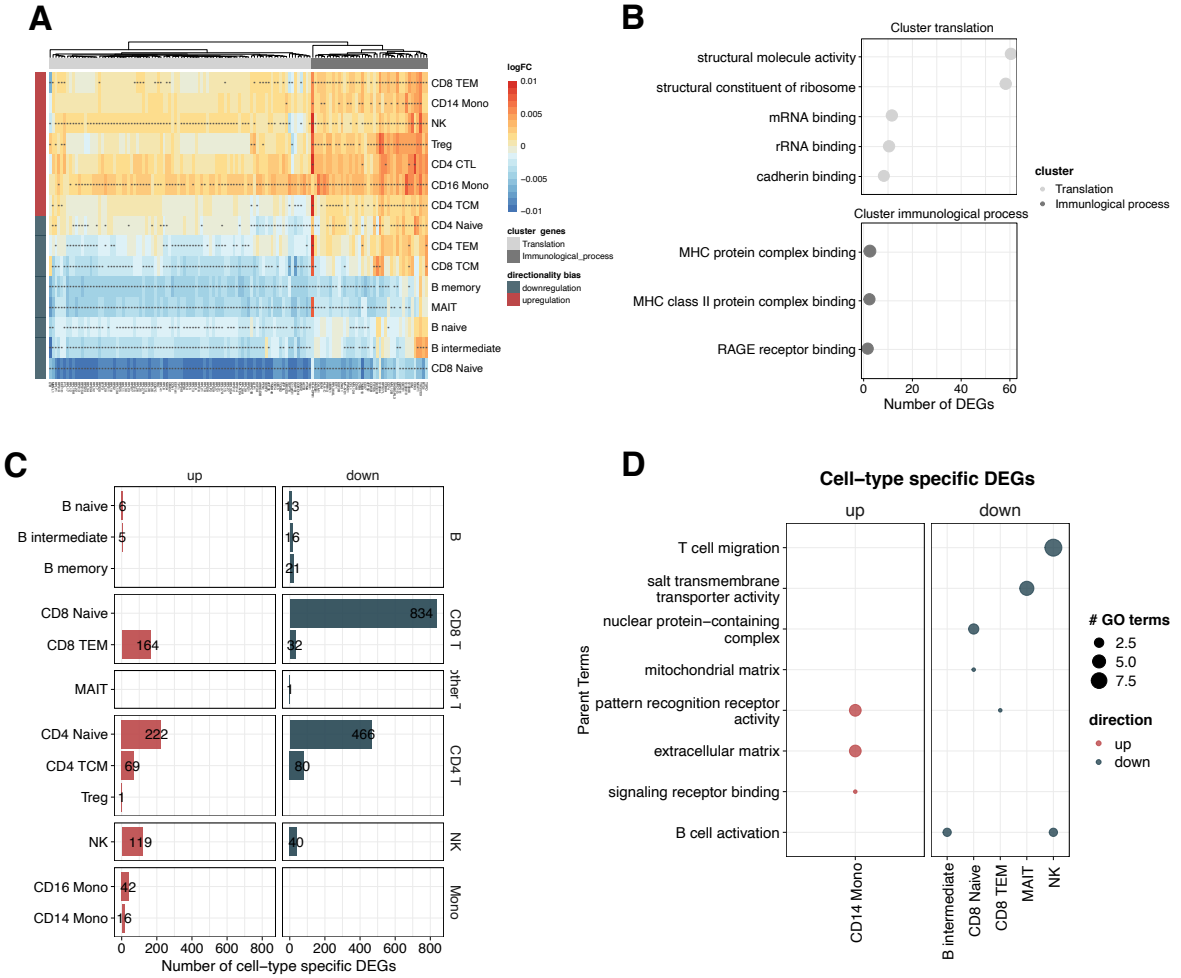

Fig 2. Aging has a widespread impact on translation and immune-related functions with cell-type specific effects A. Heatmap of age-logFC of the high cell type-shared age-DEGs (non-concordant age-DEGs in > 6 cell types) per cell type. Asterisk indicates the gene is differentially expressed with age (FDR < 0.05) in that cell type. Cell types (rows) and genes (columns) are sorted according to hierarchical clustering analysis. The row annotation bar show whether the cell type showed an up- or down-regulation bias with age in Fig. 1A. The column annotation bar show in which of the two main clusters the gene belongs to, which are functionally annotated in B. **B.** Enriched GO terms (FDR < 0.02) of the two main clusters of highly cell type-shared age-DEGs (age-DEGs in > 6 cell types) obtained in A. **C.** Number of up- and down-regulated cell type-specific age-DEGs (age-DEGs only in one cell type). **D.** Non-redundant GO term set from enriched GO terms (FDR < 0.05) of cell type-specific age-DEGs.

Strong association between age-expression shifts and changes in cell type abundance

Aging-related immune decline is driven by both alterations in cellular function and changes in cell type composition[38]. Thus, we wanted to explore whether the impact of aging on gene expression is somehow associated with changes in cellular abundance. First, to identify the age-related shifts in cell type abundance, we performed a cellular compositional analysis with age (see **Methods** and **Fig. 3A)**. Consistent with previous findings, we observe a significant increase in the proportion of both CD8+ TEM[,14,19] and NK cells[14,39,40] as well as a decrease in CD8+ T naive[14,19] and B memory[41] cells **(Fig. 3A, Supplementary Table 3)**. Interestingly, all cell types that significantly increase in proportion had most age-DEGs up-regulated in our previous analysis**.** Conversely, cell types that significantly decrease in proportion with age have most age-DEGs down-regulated **(Fig. 1A)**. Consistent with this, the magnitude of cell type composition changes with age is strongly correlated with the proportion of up-regulated age-DEGs (Spearman R = 0.82, p-value = $9 \cdot 10^{-5}$) **(Fig. 3B)**. We also replicated this correlation in [32]

(Spearman R = 0.64, p-value = $5 \cdot 10^{-3}$) **(Fig. S2H, Supplementary Table 3)**. These results suggest that changes in cell type abundance and gene expression with age are strongly associated.

We hypothesized that these age-related changes in cell type proportions in the circulating immune system could result from altered proliferation[42–46] or accumulation of senescent or quiescent cells[47]. To explore this, we identified proliferating, senescent, or quiescent cells using established gene set markers[48–50] and machine learning-based methods[51,52] (see **Methods**). While circulating immune cells do not typically proliferate under normal conditions, homeostatic proliferation can occur during aging or lymphopenia[42–46]. As a proxy for proliferation, we assigned cell cycle phase to each cell and conducted a compositional analysis with age (see **Methods** and **Fig. S3A,B, Supplementary Table 3)**. We then correlated the cell cycle composition age-estimates for each phase with our age-estimates of cell type composition and found a significant but modest negative correlation between the proportion of cells in the G1/G0 phase and changes in cell type abundance (Spearman R = -0.51, p-value = 0.04) **(Fig. 3C)**; Spearman R = -0.67, p-value = 0.006 **(Fig. S3C)).** Cells in G1/G0 phases can be in cell cycle arrest, which is indicative of quiescence (temporary arrest) or senescence (permanent arrest)[53]. Consistent with previous literature[54], T cells have the most significant increase in senescent cell enrichment with age (see **Methods** and **Fig. S3D-F)**. However, changes in the cell type abundance with age do not correlate with changes in the relative proportions of either senescent or quiescent cells with age **(Fig S3G-I, Supplementary Table 3).**

Taken together, our results suggest that cellular abundance changes with age are associated with shifts in gene expression throughput in the circulating immune compartment and cannot be explained by the accumulation of senescent or quiescent cells.

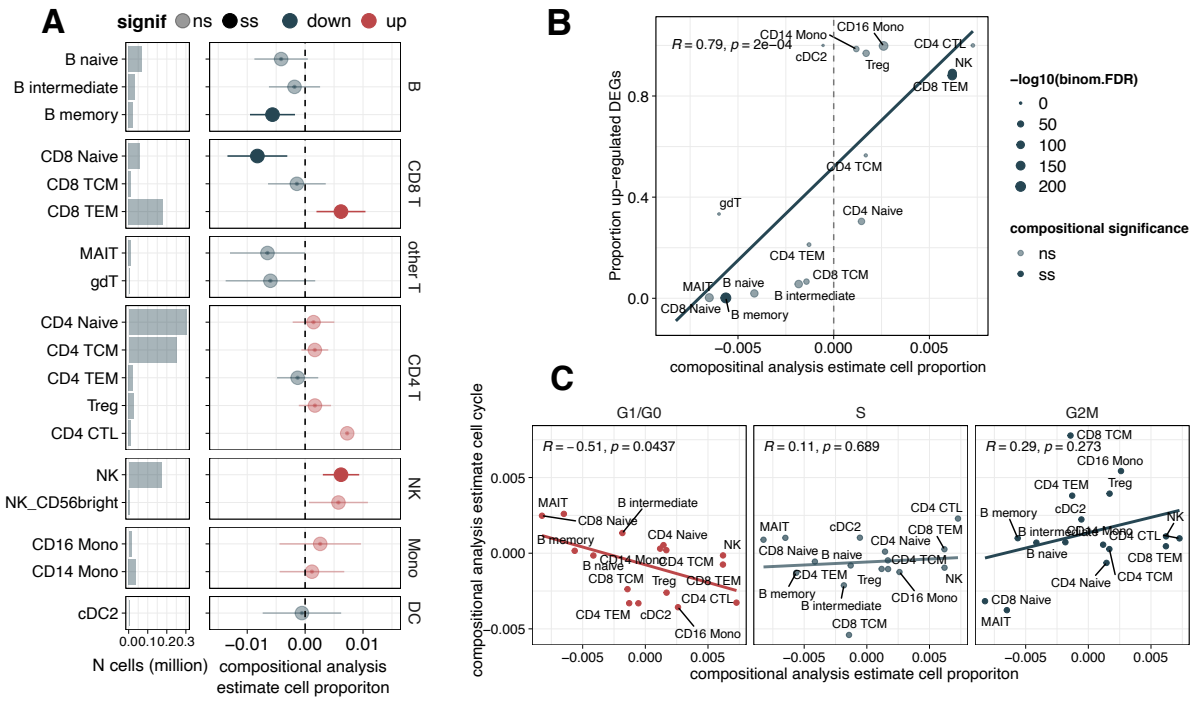

Fig. 3. Cellular abundance changes with age are always associated with expression shifts in the same direction. A. Left. Total number of cells analyzed per cell type. Right. Estimates of the cell type composition analysis with age across cell types. Bar length indicates the interval of confidence. **B.** Correlation between the proportion of age-up-regulated genes (y-axis) versus the estimate of the cell type composition analysis with age (x-axis) in each cell type. In A and B, significant (ss, FDR < 0.05) and non-significant (ns, FDR > 0.05) age-estimates of the cell type composition analysis are shown in dark or light colors, respectively. **C.** Correlation of the estimate of the compositional analysis of cell

cycle phase (G1/G0, S and G2M; assigned using gene markers from [5530,3156] (see Methods)) with age (y-axis) versus the estimate of the cell type composition analysis with age (x-axis) in each cell type.

Immune effector cells have sex-biased transcriptional changes with age

Males and females are known to differ in their innate and adaptive immune responses[22,23,57]. However, differences between males and females in immune aging still remain largely unexplored[24]. To determine whether our observed age-related changes in gene expression and composition differ between males and females, we repeated our differential gene expression and cell type composition analyses now stratifying by sex (see **Methods** and **Fig. 4A,B, Fig. S1D-H).** We obtained similar transcriptional and cellular compositional results when testing for the interaction between age and sex, but with substantially lower statistical power (see **Methods** and **Fig. S4A,B, Supplementary Table 4)**. Therefore, all subsequent analyses were based on the sex-stratified results. Most cell types had a similar number of up- and down-regulated genes with age between males and females. Likewise, most changes in cell type abundances were similar between sexes (**Fig. 4A,B, Fig. S4C-I, Supplementary Table 4)**. However, the bias in the number of up- versus down age-associated DEGs was more pronounced in one of the sexes in three immune effector cell types. Specifically, CD8+ TEM and NK cells showed a bias towards up-regulation in females (Binomial test, CD8+ TEM: females-FDR = $1.70 \cdot 10^{-176}$, males-FDR = $4.20 \cdot 10^{-07}$, NK: test, females-FDR = $8.62 \cdot 10^{-130}$, males-FDR = $3.19 \cdot 10^{-31}$) and B memory cells towards down-regulation in males (Binomial test, CD8+ TEM: females-FDR = $1.50 \cdot 10^{-09}$, males-FDR = $7.35 \cdot 10^{-50}$ **(Fig. 4A, Fig. S4J)**. In addition, CD8+ TEM cells increased in abundance with age only in females (females-FDR = 0.03, males-FDR = 0.64). Similarly, CD8+ naive T cells showed a decrease with age primarily in females (females-FDR = 0.03, males-FDR = 0.2), although a similar trend was observed in males at nominal significance (females-p-value = $9.5 \cdot 10^{-4}$, males-p-value = 0.01) **(Fig. 4B, Fig. S4K, Supplementary Table 4)**. Lastly, the positive correlation between cell type abundance and the proportion of up-regulated genes with age was more pronounced in females than in males (females: Spearman R= 0.81, p-value = $2 \cdot 10^{-9}$; males: Spearman R = 0.65, p-value = $2 \cdot 10^{-4}$) (**Fig. 4C).** To further explore the age-and-sex-dependent transcriptional responses, we performed a functional enrichment analysis on female-specific age-up-regulated genes (i.e., age-up-DEGs only found in females) in CD8+ TEM and NK cells, as well as on male-specific age-down-regulated genes (i.e., age-down-DEGs only found in males) in B memory cells. Genes exclusively up-regulated in female CD8+TEM cells were mostly related to mitochondrial function (30% of total GO enriched terms, FDR < 0.05) **(Fig. 4D, Supplementary Table 4**), which plays an essential role in the cytotoxic function of T lymphocytes[58]. Additionally, sex-specific age-DEGs in NK and B memory cells were enriched in ribosomal genes **(Fig. S4L,M, Supplementary Table 4)**.

Sex is a well-known factor influencing disease susceptibility, driven by both hormonal and genetic factors[22]. However, how these sex differences may influence age-related disease susceptibility remains understudied. To explore this, we performed a functional enrichment analysis of sex-specific age-DEGs (i.e., age-DEGs only found in one of the two sexes) against a disease ontology database[59] (see **Methods** and **Fig. 4E, Supplementary Table 4)**. In females, age-up-regulated DEGs in NK were enriched in pathways related to autoimmune and cardiovascular diseases, consistent with the higher susceptibility of females to autoimmune diseases[23] and increased risk of heart disease in postmenopausal females[60]. Although NK cells are not the primary drivers of these conditions, they play a significant role in their development[61–66]. In contrast, in males, we detected enrichments for age-down-regulated DEGs in B memory cells in pathways related to anemia mostly driven by translation-related genes. Given that autoimmune diseases have a strong heritability component[67] and disease susceptibility is sex-biased[68,69], we explored whether genetic variants associated with autoimmune diseases in the Genome Wide Association Studies (GWAS) catalog[70] are enriched in age-DEGs from our sex-stratified differential gene expression analysis (see **Methods**). Age-DEGs in naive CD4+ T female cells were enriched for genes associated with multiple sclerosis (OR = 2.2, FDR = $8.81 \cdot 10^{-03}$), while those detected in CD8+ TEM female cells were enriched for Crohn's disease (OR = 2.8, FDR = $1.50 \cdot 10^{-02}$) **(Fig. 4F, Supplementary Table 4).** We did not detect significant enrichments (FDR < 0.05)

when considering sex-specific age-DEGs, and neither when considering all male age-DEGs, even though some genes were differentially expressed in both sexes **(Supplementary Table 4)**.

Together, our analysis highlights sex as an important factor driving cell type-specific changes in immune cell composition and molecular profiles with aging, particularly in CD8+ TEM, NK and B memory cells. Furthermore, our findings suggest potential links between female-specific molecular and compositional changes and both autoimmune and cardiovascular diseases, underscoring the complex relationship between sex, aging, and immune function.

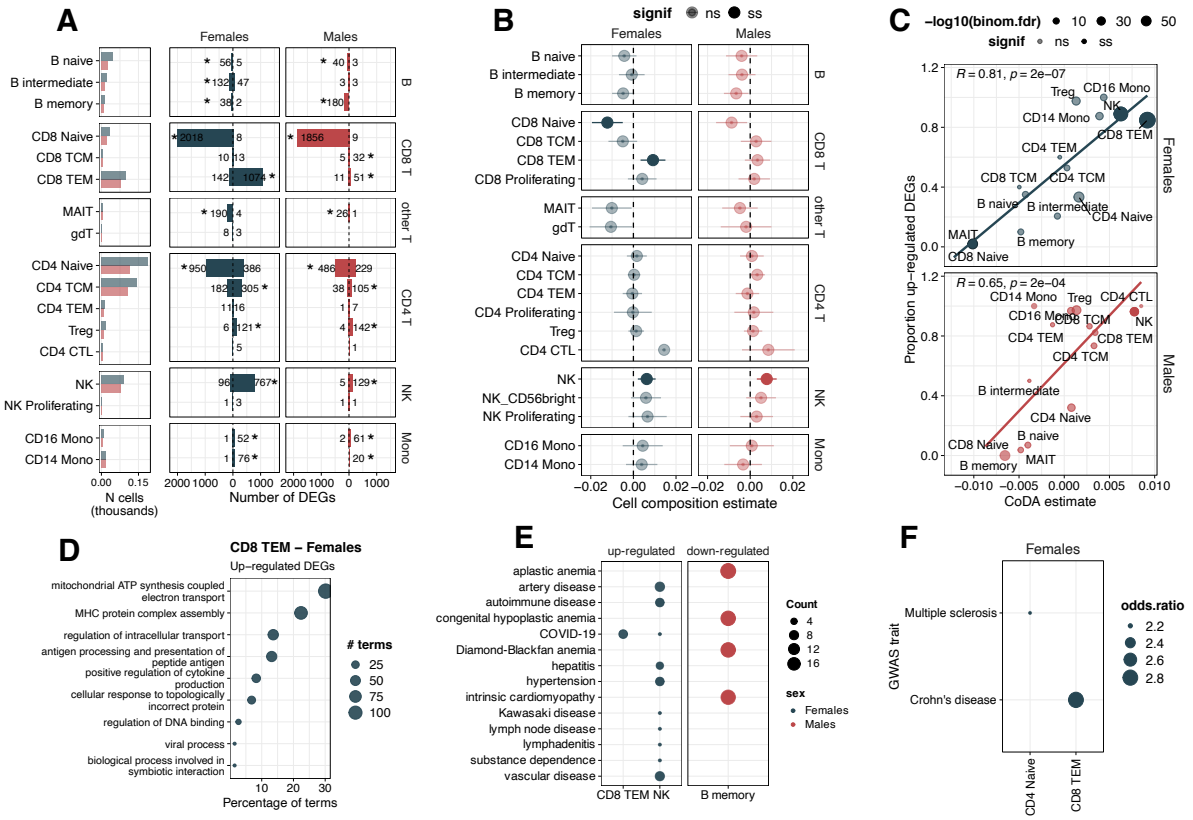

Fig. 4. Genes changing expression with age in CD8+ TEM, CD4+ T naive and NK in females are associated with autoimmune disease. A. Left. Total number of cells analyzed per cell type and sex. Right. Number of age-DEGs (FDR < 0.05) per cell type in females and males. Asterisk indicates a significant bias towards up- or down-regulation (Binomial test; FDR < 0.05). **B.** Estimates of the sex-stratified cell type composition analysis with age across cell types. Bar length indicates the 95% confidence interval. **C.** Correlation between the proportion of age-up-regulated genes (y-axis) versus the estimate of the cell type composition analysis with age in females (top) and males (bottom) (x-axis) in each cell type. In B and C, significant (ss, FDR < 0.05) and non-significant (ns, FDR > 0.05) age-estimates of the cell type composition analysis are shown in dark or light colors, respectively. **D.** Non-redundant GO term set from enriched GO terms (FDR < 0.05) of female-specific up-regulated age-DEGs in CD8+ TEM cells. **E.** Disease Ontology enrichment of sex-specific age-DEGs. **F.** Overlap between age-DEGs from the sex-stratified analysis and GWAS hits for autoimmune traits (FDR < 0.01).

Cell type-agnostic abundance analysis identifies sex-specific increases in cell subpopulations with age

Recent studies have identified specific cell subpopulations that accumulate with age, serving as hallmarks of the aging immune system[13,14,71]. Nonetheless, the extent to which these changes differ between males and females remains unclear. Our previous findings suggest that CD8+ TEM expand

with age specifically in females. This expansion could be driven by particular cell subpopulations yet to be identified. However, relying solely on pre-defined cell type annotation based on discrete clustering strategies limits the detection of novel subpopulations[72–75]. To address this limitation, we used Milo, a framework for differential abundance (DA) testing that does not rely on discrete cell type labels[72,73] . In brief, Milo constructs a k-nearest neighbor (KNN) graph where each node represents a cell, and edges connect each cell to its nearest neighbors based on gene expression similarity. Milo then identifies cellular states as neighborhoods with overlapping cells on this graph, which we used for DA testing with age (see **Methods**). We ran Milo stratifying the data by sex and observed that females had more changes in cell abundance with age than males. In females, 11% of neighborhoods were age-enriched and 8% age-depleted, compared to 4.3% age-enriched and 3% age-depleted neighborhoods in males (**Fig. 5A, Fig. S5A-C, Supplementary Table 5).** To compare the neighborhood-based DA approach, which is cell type-agnostic, to our previous cell type-based composition analysis (**Fig. 4A**), we assigned each neighborhood to a particular pre-annotated cell type (see **Methods**). Both approaches were highly concordant (females: Spearman R = 0.82, p-value = 0.001; males: Spearman R = 0.71, p-value = 0.002) **(Fig. S5D, E)**. Consistent with our sex-stratified compositional analysis **(Fig. 4B)**, females had six times more age-enriched CD8+ TEM neighborhoods than males (25% vs. 4%, respectively) **(Fig. 5A)**. Furthermore, the neighborhood-based approach revealed sex-specific subcellular alterations that were missed in our cell type-based analysis. Specifically, females had an age-related decline in nearly the entire population of MAIT cells (85% of neighborhoods in females vs. 16% in males), along with a female-specific increase in Monocytes (CD14+ Monocytes: 5% of neighborhoods in females vs. 0% in males, and CD16+ Monocytes: 63% of neighborhoods in females vs. 3% in males) **(Fig. 5A, Fig. S5C).** Notably, previous research showed that while females of reproductive age typically have higher MAIT cell levels than males, this difference diminishes with age[76]. This suggests that this female-specific age-related decline may be linked to hormonal changes after menopause. In contrast, B naive and B intermediate cells had a few neighborhoods with the largest increase in abundance with age specifically in males (B naive: 4% of neighborhoods, mean-logFC = 0.06, and B intermediate: 17% of neighborhoods, mean-logFC = 0.07) **(Fig. 5A)**. Indeed, these sex-specific abundance biases were disregarded in our non-sex-stratified analysis, which detected both the male and female abundance changes simultaneously **(Fig. S5F, G)**.

We then wanted to functionally characterize the cell subpopulations that belong to neighborhoods with sex-specific age-related abundance changes. To do this, we first identified gene markers in age-enriched or age-depleted DA neighborhoods within a specific cell type and then performed functional enrichment analyses (see **Methods**). We found biological pathways significantly enriched in age-enriched CD8+ TEM and CD14+ Monocyte neighborhoods in females, and in B naive in males **(Fig 5B, Supplementary Table 5)**. The CD8+ TEM cells increasing in abundance with age in females had marker genes related to cell-killing processes and NK activation, suggesting a cytotoxic phenotype[77]. Consistent with this, the top marker genes we identified have been described previously as a marker signature for a cytotoxic *GZMB*+ CD8+ TEM subpopulation[13,14,71]. Importantly, when we classified all CD8+ TEM cells into *GZMB*+ or *GZMB*- and repeated the sex-stratified compositional and differential gene expression analyses with age (see **Methods** and **Supplementary Table 5**), only the CD8+ TEM *GZMB*+ subpopulation in females increased in proportion with age (females: estimate = 0.02, FDR = $1.1 \cdot 10^{-3}$, males: estimate = 0.007, FDR = 0.3) **(Fig. S5J)** and had a higher number of age-DEGs **(Fig. S5K).** This suggests that *GZMB*+ subpopulation increase with age was driving the observed age-abundance shift in females of the whole CD8+ TEM population **(Fig. 4B)** as well as the higher number of age-DEGs **(Fig. 4A)**. We then wanted to further characterize the CD14+ Monocyte subpopulation that increased in abundance in females with age. Cells from this subpopulation had marker genes mostly enriched in defense responses to viruses and interferon (IFN) type I-mediated signaling pathways **(Fig. 5B, Supplementary Table 5)**. Consistent with this, these cells had a greater induction of both interferon-stimulated genes (estimate = 0.018, FDR = $7.3 \cdot 10^{-03}$) and inflammatory genes (estimate = 0.019, FDR = $9.3 \cdot 10^{-12}$) (see **Methods** and **Fig. 5D, Supplementary Table 5)**. Unlike the pattern observed in CD8+ TEM cells, we did not find a unique age-related transcriptional signature

within this female-specific CD14+ Monocyte subpopulation **(Fig. S5L, Supplementary Table 5)**. Finally, we found an enrichment in terms related to extracellular structure organization in male age-enriched naive B cells **(Fig. 5B, Supplementary Table 5)**. This is consistent with the growing knowledge of the relationship between the extracellular matrix and the immune system in the context of aging[78]. In addition, top gene markers in these neighborhoods showed that this subpopulation belongs to a previously described *CD5*+ B cell population characterized by the high expression of lipid-transport genes (*ABCA6, APOD*)[13,14,71] **(Fig. 5E, Fig. S5M,N, Supplementary Table 5).** The limited number of cells in this subpopulation hampered the identification of subpopulation-specific gene expression changes with age **(Supplementary Table 5)**. Using external datasets with limited sample size[14,32], we were able to replicate the observations that in males *CD5*+ B cell subpopulation increase with age (estimate = 0.03, FDR = 0.003) and in females *GZMB*+ CD8+ TEM increases with age (estimate = 0.05, FDR = 0.07). We did not observe female-specific accumulation of the CD14+ Monocyte subpopulation likely due to lack of power as this was the most subtle abundance change (17 female-age-eriched neighborhoods out of 353 tested neighborhoods (5%), mean-logFC = 0.02) (see **Methods** and **Supplementary Table 5**).

Our findings underscore the importance of cell type-agnostic strategies and identify for the first time three sex-specific immune subpopulations that change abundance with age. Overall, we found that sex plays a major role in how immunosenescence progresses both in terms of gene expression and immune cell populations, which explains, at least in part, why the function of the immune system differs in male or female individuals during aging.

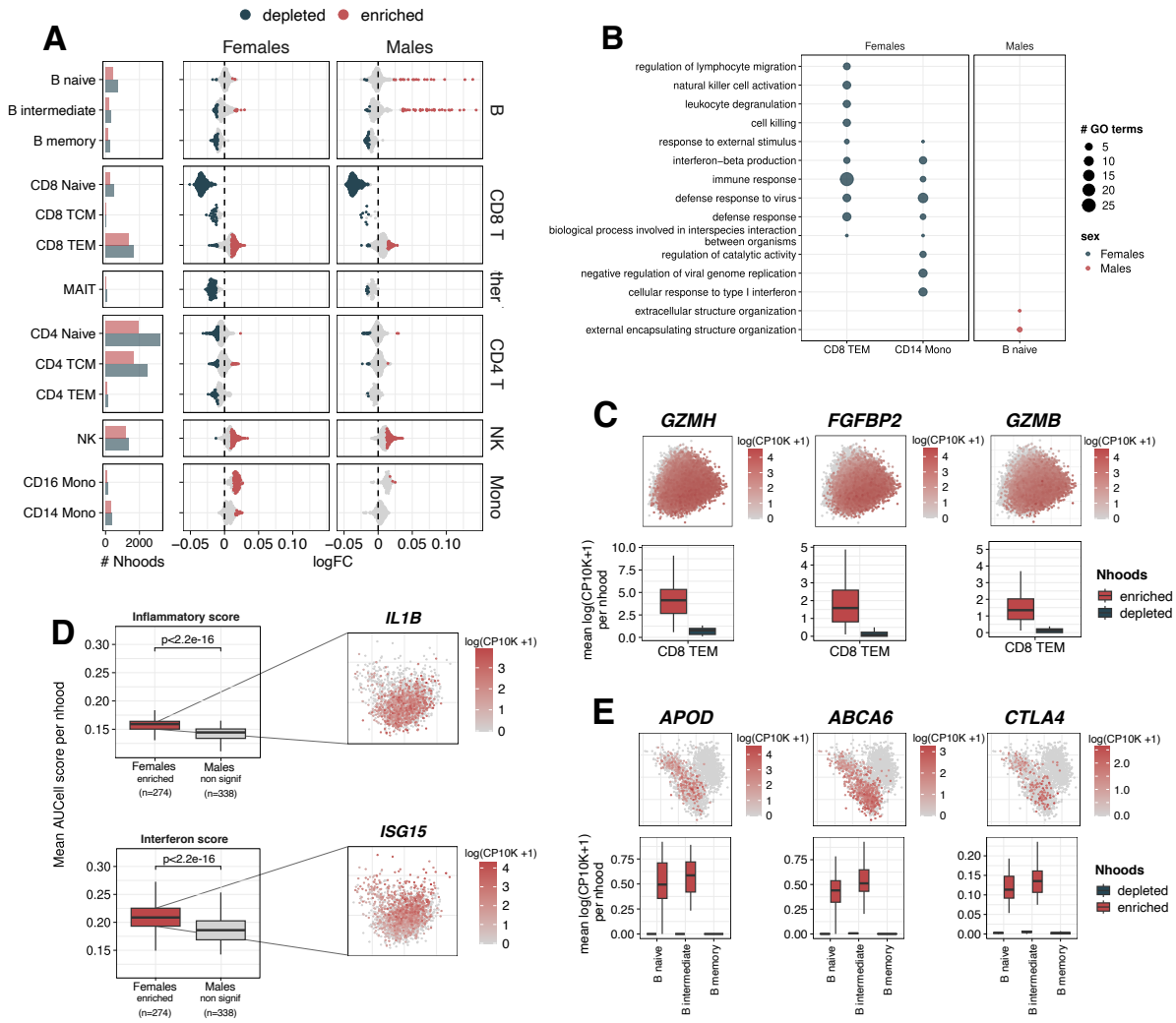

Fig. 5. Cell type-agnostic analysis identifies functionally distinct sex-specific subpopulations accumulating with age. A. Left. Total number of neighborhoods tested per cell type. Right. Beeswarm plot of the distribution of neighborhood logFC with age in females and males. Each dot corresponds to a neighborhood of cells assigned to a cell type (> 80% of cells belonging to that cell type per neighborhood (see **Methods**)). Age-DA neighborhoods (Spatial-FDR < 0.05) are colored. **B.** Non-redundant GO term set from enriched GO terms (FDR < 0.05) of marker genes of age-enriched subpopulations of neighborhoods in females and males. **c.** Top. UMAP of female-age-enriched CD8+ TEM cells colored by the expression of the top marker genes (*GZMH, FGFBP2,* and *GZMB*). Bottom**.** Average log-normalized and scaled expression per neighborhood of the same genes. Only neighborhoods that significantly increase (enriched) or decrease (depleted) with age are shown. **d.** Left. Distribution of inflammatory (top) and interferon scores (bottom) among female-age-enriched CD14+ Monocyte neighborhoods compared to male non-significant CD14+ Monocyte neighborhoods. Right. UMAP of female-age-enriched CD14+ Monocyte cells colored by the expression of a top gene marker related to inflammation (*IL1B,* top) and an interferon-stimulated gene (*ISG15,* bottom) **e.** Top. UMAP of male-age-enriched B naive and intermediate cells colored by the expression of the top marker genes (*APOD, ABCA6,* and *CTLA4*). Bottom. Average log-normalized and scaled expression per neighborhood of the same genes. Only neighborhoods that significantly increase (enriched) or decrease (depleted) with age are shown.

Discussion

Immune aging is characterized both by cell type abundance changes and a decline in cellular function [11–15]. Our single-cell analysis of almost a thousand individuals' cohort[29] provides insights into how these two processes are interconnected in the circulating immune system at high resolution. We show that nearly all immune cell types show expression changes that positively correlate with abundance shifts. While this correlation may not be unique to aging, it is particularly prevalent. Notably, these expression shifts often involve ribosome-related genes, suggesting that cells with a higher number of up-regulated genes may need to increase ribosome levels to translate newly available transcripts and vice versa. One possible explanation for this correlation is that cell types with increased abundance harbor cells that are transitioning between cell states. As cells enter new states, they need to express new regulatory programs leading to transient higher transcriptional throughput with a subsequent adjustment for available ribosomes. For instance, the observed CD8+ TEM increase in abundance and bias towards up-regulation specially for translation genes with age in females, is mostly explained by the increase in abundance of a novel cell subpopulation with a cytotoxic regulatory program. We speculate that this correlation between transcriptional changes, particularly in ribosomal gene expression, and changes in cell abundance may serve as reliable proxies for monitoring immune cell dynamics using single-cell data in other contexts beyond aging. Finally, despite observing changes in cell type abundance with age in specific immune cell types, we could not associate it to proliferation changes or changes in abundance of senescence or quiescence cells. This may be explained because immune cells predominantly proliferate in lymphoid organs. Future studies should examine proliferation dynamics within primary and secondary lymphoid organs, where turnover rates are more pronounced.

Together, this work offers a comprehensive view of human circulating immune cell aging, with important insights into sex-specific immune aging dynamics. Specifically, our cell type-agnostic analysis reveals substantial differences in the aging of the immune system between males and females. In females, we observe an age-related increase of a subpopulation of CD8+ TEM cells, characterized by *GZMB* expression and an NK-like cytotoxic phenotype[77] (**Fig. S5I)**. This finding suggests that this subpopulation may have a sex-specific role in the immunosurveillance of senescent cells, whose accumulation with age likely contributes to the chronic low-grade inflammation characteristic of aging, referred to as inflammaging. Myeloid subpopulations have been reported to remain static across lifespan[14,79]. However, when we stratify our analyses by sex, we identify a female-specific increase in CD16+ Monocyte population as well as a CD14+ Monocyte subpopulation characterized by an

interferon and inflammatory gene signature. The observed increase in these CD8+ TEM and CD14+ Monocyte subpopulations suggests that the pro-inflammatory capacity in females, which may be linked to higher rates of autoimmunity, becomes more pronounced with age. Consistent with this, we find that female, but not male, age-DEGs in CD8+ TEM and CD4+ T naive are enriched in GWAS hits for autoimmune diseases. This provides a strong link between sex-specific transcriptional response with age and autoimmune disease susceptibility. Conversely, in males, we detect a strong age-related enrichment in a B cell subpopulation characterized by a signature of lipid-transport genes (*ABCA6, APOD)*. This subpopulation also exhibits increased B cell receptor clonality with age[14], similar to an *APOE*-expressing B cell subpopulation in mice that accumulate with age in multiple tissues[80,81,82]. Our results suggest that this age-related clonal expansion may be specific to males. Together, our findings evidence the power of cell-type agnostic approaches to identify subpopulations that accumulate with age in a sex-specific manner. These subpopulations have the potential to serve as sex-specific hallmarks of immunosenescence, emphasizing the need to include a sex perspective in all immunological studies.

Nevertheless, our study has certain inherent limitations that should be considered in the design of future research. First, expanding the cohort to add individuals from diverse ancestries and a wider age range, including pediatric and supercentenarian samples, would help capture the full spectrum of aging transcriptional dynamics. Second, while we rely on a population cohort which excludes individuals with active infections, additional information on factors such as cytomegalovirus exposure, smoking, and comorbidities[83–86], would provide valuable context for interpreting immune aging. Lastly, we focus on chronological aging, which may not fully reflect the biological age of an individual [87–89]. Incorporating additional data layers to predict biological age, such as methylation profiles[90–92] or telomere length[93] could offer a more accurate proxy of biological age to further study immunosenescence in the future. Despite these limitations, our results highlight the value of high-resolution, single-cell analyses in revealing the complex and sexual-dimorphic nature of immune aging. This work offers promising avenues for developing sex-tailored interventions that could enhance immune function in elderly individuals and mitigate the impact of age-related immune diseases.

Methods

scRNA-seq dataset

Data overview and pre-processing

We analyzed scRNA-seq data over one million PBMCs (1,272,518 cells) from the OneK1K cohort[29]. This population cohort consists of 982 individuals (416 (42%) male and 565 (58%) female) of Northern European ancestry recruited in Hobart, Australia. The age of participants ranged from 19 to 97, with 73% of the participants being older than 60 years old. Participants with active immune diseases and non-European individuals were excluded from the study, following ancestry inference based on genotype data. The pre-processing of the scRNA-seq data (sample preparation and pooling, sequencing, quantification and alignment, demultiplexing, quality control and normalization) was conducted according to the original publication[29].

Cell type annotation

We performed automated cell type classification using Azimuth to annotate the cells[30]. In detail, we conducted a supervised analysis guided by a reference dataset to enumerate cell types that would be challenging to define with an unsupervised approach. We mapped our scRNA-seq query dataset to a recently published CITE-seq scRNA-seq reference dataset of 162,000 PBMCs measured with 228 antibodies[30]. For this process, we used a set of specific functions from the R package Seurat (v4.0.0)[30,31]. First, we normalized the reference dataset using the Seurat::SCTransform function. Next, we found anchors between reference and query using a precomputed supervised PCA (spca)

transformation with 50 dimensions using the Seurat::FindTransferAnchors function. We then transferred cell type labels from the reference to the query by projecting the query set on the UMAP embeddings of the reference using the Seurat::MapQuery function. To optimize computational resources, these steps were executed independently for each sequencing pool. Subsequently, the resulting cell type annotations were mapped back onto the complete dataset. We ended up with a total of 31 circulating immune cell subpopulations (https://github.com/powellgenomicslab/onek1k_scd). Finally, we filtered the dataset to retain only cell types with at least 5 cells in a minimum of 4 individuals, resulting in 24 distinct cell types **(Supplementary Table 1)**.

Differential gene expression analysis with age

To identify DEGs with age within each of the 24 immune cell subpopulations, we followed the dreamlet pipeline (v1.1.19)[94]. First, we applied a pseudobulk approach, which aggregates gene counts for each donor to account for both zero-inflation and within-sample correlation, using the dreamlet::aggregateToPseudoBulk function with default parameters. Then, we applied a voom-style normalization for pseudobulk counts using the dreamlet::processAssays function with default parameters except for min.prop=0.2 to retain genes located in the sex chromosomes. Precision-weighted linear mixed models enable accounting for repeated measures study designs, high dimensional batch effects, and varying sequencing depth or observed cells per biosample. Lastly, we fit a regression model for each gene to test differential expression across individuals associated with a trait of interest (i.e., age, which was treated as a continuous variable) while controlling for other sources of biological (sex) and technical variation (sequencing pool) using the dreamlet::dreamlet function with default parameters. We corrected for multiple testing using FDR, through the Benjamini–Hochberg method, and extracted the results using the dreamlet::topTable function. The model used was:

pseudobulked normalized expression ~ Age + Sex + (1|Sequencing Pool)

Genes were considered differentially expressed with age when the gene expression change was significant at an FDR < 0.05.

To study age-and-sex dependent gene expression changes, we followed two different approaches:

(1) We performed differential gene expression analysis using the same steps as previously described, but testing for the interaction term between age and sex. The model used was:

pseudobulked normalized expression ~ Age + Sex + Age:Sex + (1|Sequencing pool)

Genes were considered differentially expressed when the gene expression change was significant at a nominal p-value < 0.01.

(2) We stratified the data by sex and conducted the pseudobulk aggregation, pre-processing, age-differential gene expression analysis separately within the male and female subsets following the same steps as previously described, except for: *i)* the min.prop parameter (default, 0.4) in the dreamlet::processAssays function, and *ii)* the model used the in dreamlet::dreamlet function. The model used was:

pseudobulked normalized expression (by sex) ~ Age + (1|Sequencing Pool)

Genes were considered differentially expressed with age when the gene expression change was significant at an FDR < 0.05.

Downsampling analysis for differential gene expression analysis with age

The 24 immune cell subpopulations are detected in different numbers of donors **(Fig. S1E)**. To assess how sample size may influence the detection of age-DEGs, we randomly downsampled the whole

dataset 10 times to the minimum number of donors a cell type was detected in (i.e., 553 donors in MAIT cells), and ran the differential gene expression analysis with age per cell type.

Directionality bias in the differential gene expression analysis with age

To assess whether age-DEGs had directionality bias (i.e., whether the frequency of up-regulated age-DEGs is statistically significantly different from the frequency of down-regulated age-DEGs), we performed a binomial test on the proportion of age up-regulated DEGs to age down-regulated DEGs. The resulting p-values were then adjusted for multiple comparisons across cell types using the Benjamini-Hochberg method to control for FDR. The same procedure was followed in the sex-stratified analysis.

Directionality concordance of age-DEGs across cell types

To analyze the similarities in age-related gene expression changes across different cell types, we looked at each pair of cell types and identified common age-DEGs. Then, we determined the number of genes that exhibited 'consistent directionality' in expression changes with age, (i.e., same logFC sign; either both showing logFC < 0 or logFC > 0). Finally, we calculated the percentage of these 'consistently changing genes with age' relative to the total intersection of age-DEGs between the two cell types. We used these percentages to perform hierarchical clustering using the pheatmap function from the R package pheatmap (v1.0.12)[95], with default parameters, to group cell types based on their similar age-related expression patterns. The same procedure was followed in the sex-stratified analysis.

Hierarchical clustering of highly-shared age-DEGs

We selected genes that were differentially expressed with age in more than a specified number of cell types (between n>1 and n>9) and used their age-related expression changes (logFC) for hierarchical clustering. Both genes and cell types were clustered using the hclust function from the R package stats (v4.3.0) with default parameters, applying the "average" method for linkage. The same procedure was followed in the sex-stratified analysis.

Comparison of hierarchical clustering dendrogram structures

To compare trees (dendograms) of hierarchical clustering (i.e., trees generated by selecting age-DEGs shared across varying numbers of cell types, or trees generated by selecting age-DEGs shared in ≥ 2 cell types when using Oelen 2022 data[32]), we used the cor.dendlist function from the R package dendextend (v1.17.1)[96] with default parameters. This analysis provides a correlation matrix between a list of trees with the 'cophenetic' correlation coefficient, which quantifies the similarity between the different clustering dendrograms ranging from 0 to 1. The same procedure was followed in the sex-stratified analysis.

Cell-type agnostic DA analysis with age

We used the R package miloR (v1.99.16, devel branch)[73], a method designed to detect DA cell populations between conditions by modeling cell counts in neighborhoods of a KNN graph. Milo implements generalized linear mixed model (GLMM) for DA testing, which comes with computational and memory constraints. Due to these limitations, we could not run this analysis on the entire dataset. Instead, we randomly downsampled our data to 25% (316,939 cells out of a total of 1,272,489 cells) and repeated the analysis three times **(Fig. S6A,B, Supplementary Table 6)**. For the sex-stratified analysis, we split the dataset into males and females, and randomly downsampled at 50% within each sex group (369,888 out of 739,776 in females, and 255,507 out of 511,014 in males).

Identification of neighborhoods of cells

To begin, we applied the batch-correction R package harmony (v1.0.3)[97] to remove batch effects (i.e., sequencing pool) in our dataset. We used the function harmony::RunHarmony with default parameters and group.by.vars="Sequencing Pool". We then used the harmony-corrected PCA embedding to construct the KNN graph with the miloR::bulidGraph function, setting parameters k=100 and d=30. This function calculates the Euclidean distance between each cell and its KNN in the principal component space. We set a *k* parameter (*k*=100) that reduces the overall number of neighborhoods while keeping a relative small neighborhood size to achieve high resolution **(Fig. S6C,D)**. We also checked the count distributions for each neighborhood based on our variable of interest (age) using miloR::checkSeparation function with default parameters at several values of *k* **(Fig. S6E)**. This function determines whether neighborhoods exhibit perfect separation into all-zero versus non-zero counts for age, which can cause model failure in GLMM. The function outputs a logical classification ('TRUE' or 'FALSE') for each neighborhood, indicating the presence or absence of perfect separation, respectively. Identifying such neighborhoods is crucial for informing decisions on neighborhood subsetting or the recomputation of the nearest-neighbor graph to refine neighborhood definitions and ensure robust GLMM analysis. The distribution was similar among all *k* parameters. Next, we used the miloR::makeNhoods function to create neighborhoods of cells with parameters prop=0.05, k=100, and d=30, along with the graph-based refinement method (refinement_scheme="graph"), which has been demonstrated to speed up this procedure[73]. Finally, we applied the miloR::countCells and miloR::calcNhoodExpression functions with default parameters to count the number of cells and their gene expression within each neighborhood. The same procedure was followed in the sex-stratified data (**Fig. S6F-H)**.

Testing for age-DA

We performed DA analysis with age using the miloR::testNhoods function with the following parameters: fdr.weighting="graph-overlap", glmm.solver="HE-NNLS" and norm.method="TMM". The model for the entire dataset was the following:

Neighborhood counts ~ Age + Sex + (1|Sequencing pool) + (1|Donor)

The model for the sex-stratified data was as follows:

Neighborhood counts (by sex) ~ Age + (1|Sequencing pool) + (1|Donor)

To control for multiple testing, we used an adapted version of a weighted FDR procedure[98] for a KNN graph. In detail, this procedure accounts for the spatial overlap of neighborhoods, building upon an approach initially introduced in Cydar[99]. We employed an alternative weighting scheme to avoid distance calculations using the number of overlapping cells between each neighborhood[73].

We then ran the function miloR::buildNhoodGraph with default parameters to create an abstracted graph of neighborhoods for visualization and miloR::annotateNhoods with coldata_col="cell_type" to assign a cell type label to each neighborhood (here referred to as 'Nhood cell type label') based on the most frequent label of the neighborhood. For visualization purposes, changes in neighborhood abundance grouping by cell type, we omitted neighborhoods with a frequency of the major cell type label < 70%.

Identification of gene markers for age-DA enriched or depleted neighborhoods in males and females

To identify gene markers of neighborhoods significantly enriched or depleted with age in our sex-stratified DA within a specific cell type, we first extracted the expression data for all cells within each neighborhood of that cell type. In this case, we used the cell type annotation corresponding to the 'Nhood cell type label' obtained using the Milo framework as explained in a previous section. Then, we extracted the top 2,000 highly variable genes with Seurat::FindVariableFeatures function with default parameters. We used this set of genes to run Seurat::FindMarkers to identify significantly age-enriched or -depleted neighborhoods markers. Specifically, to find gene markers of age-DA enriched

neighborhoods, we set *idents.1* to include all cells from neighborhoods that were significantly enriched with age (age-logFC > 0 & spatial-FDR < 0.05) and *idents.2* to include all cells from depleted neighborhoods with age, including both significant and non-significant ones (age-logFC < 0). Conversely, to identify gene markers of age-DA depleted neighborhoods, we set *idents.1* to include all cells from significantly depleted neighborhoods (age-logFC < 0 & spatial-FDR < 0.05), and *idents.2* to include all cells from enriched neighborhoods, including both significant and non-significant ones (age-logFC > 0). Any cells that were present in both *idents.1* and *idents.2* categories, due to the overlapping nature of the Milo graph, were excluded from the analysis. Genes were considered markers if they met the criteria of FDR < 0.05 and logFC > 1.

Cell cycle phase assignment

We assigned a cell cycle phase to each individual cell using two different approaches as an orthogonal validation: *(1)* using the CellCycleScoring function from the R package Seurat (v4.0.0)[30,31] and *(2)* using the project_cycle_space function from the R package tricycle (v1.12.0)[51].

> *(1)* The Seurat approach uses a set of pre-computed cell cycle markers for G2M and S phase[55] to score and assign cells to a specific discrete cell cycle phase (G2M/G1/S). G2M and S marker sets should be anticorrelated in their expression levels, and cells expressing neither are likely not cycling and in G1 phase. Since Seurat cannot distinguish between G1 and G0 phases[56], we relabeled the G1 cells as G1/G0.
>
> *(2)* The tricycle approach infers cell cycle position by projecting cells into a pre-learned cell cycle space derived from a reference dataset. The estimated cell cycle position is represented as an angle between 0 and 2π. Based on this, we estimated cell cycle phases: 0.5π marks the start of the S phase, π indicates the start of the G2M phase, 1.5π corresponds to the middle of the M phase, and the range from 1.75π to 0.25π represents the G1/G0 phase.

Next, for each independent approach (*1* and *2*) and cell type, we modeled how the proportion of cells in each cell cycle phase changes with age using compositional data analysis as described in the previous section. We applied the Benjamini–Hochberg method to correct for multiple testing and controlled the FDR. We considered cell cycle phases significantly changing with age at FDR < 0.05. The LMM used was:

$$\text{CLR cell cycle score} \sim \text{Age} + \text{Sex} + (1|\text{Sequencing Pool})$$

Only cell types with age-DEGs (18 cell types shown in **Fig. 1A**) were tested.

Functional enrichment analysis

Over-representation analysis (ORA)

We used the R package clusterProfiler (v4.12.0)[100] to perform all ORA throughout the manuscript. We ran the analysis against the GO database[101,102] using clusterProfiler::enrichGO function with default parameters but setting ont="GO" in all cases except for the enrichments of cell type-specific DEGs that ont="ALL". For cell type-specific ORA, we used the set of all expressed genes in each cell type as the background. When analysing shared age-DEGs across cell types, the background was the intersection of expressed genes across the tested cell types. Multiple testing correction was applied using the Benjamini–Hochberg method to control the FDR, with significant terms identified at FDR < 0.05. Finally, in some cases **(Fig. 2D, Fig. 4D, Fig. 5B, Fig. S4L, M)**, we obtained a non-redundant gene set of terms for visualization purposes. To do so, we used the R package rrvgo (v1.16.0)[103], which simplify the redundancy of GO sets by grouping similar terms based on their semantic similarity. We applied the function rrvgo::calculateSimMatrix to get the similarity matrix between terms, and the rrvgo::reduceSimMatrix function to group terms, based on the similarity matrix, with a similarity score below a specific threshold, and selected the term with the higher score (i.e., score reflecting how unique

the term is) within the group as the group representative. In the case of multiple cell types, we merged the all enrichment terms and performed the reduction across cell types.

Additionally, we performed an ORA of the sex-specific age-DEGs (i.e., age-DEGs only found in one of the two sexes) using a Disease Ontology database[59] via the R package DOSE (v3.30.1)[104]. We used the same background selection and p-value correction strategies as mentioned above.

AUCell Gene Set Enrichment Analysis (GSEA) using senescent, quiescent, interferon, and inflammatory gene sets

We performed GSEA using the R package escape (v1.99.1)[105]. This package is developed to incorporate multiple methods of GSEA that are particularly advantageous for scRNA-seq data. We used the function escape::runEscape with default parameters except for min.size=0, groups=5000, method="AUCell". AUCell is a GSEA method developed in [106] that calculates the area under the curve (AUC) to determine whether specific gene sets are enriched among the top 10% expressed genes in each cell. Only cell types with age-DEGs (18 cell types shown in **Fig. 1A**) were tested. The output is an enrichment score for each gene set and cell.

<u>Senescent and quiescent enrichments</u>

To compute the enrichment of our annotated cell types in senescent and quiescent gene sets, we used a senescent gene set from [50] and a quiescent gene set manually curated based on [49]**.**

For each gene set, after computing the enrichment score for each cell as previously explained, we used the function escape::performNormalization to normalize the enrichment scores by accounting for single-cell dropout and producing positive values for downstream differential enrichment analysis [105].

Then, we performed a differential enrichment analysis to explore how the enrichment scores for those gene sets changed with age across different cell types. We fitted a GLMM using a Gaussian distribution and corrected for FDR through the Benjamini–Hochberg method. We considered cell types to be significantly enriched with age in a particular gene set at FDR < 0.05. The GLMM used was:

$$\text{AUCell scores} \sim \text{Age} + \text{Sex} + (1|\text{Sequencing Pool}) + (1|\text{Donor})$$

<u>Interferon and inflammatory enrichments of age-enriched CD14+ Monocytes in females</u>

To compute the enrichment of CD14+ Monocyte neighborhoods in interferon-stimulated and inflammatory gene sets, we followed the same approach as in[107]. Briefly, we defined our interferon-stimulated gene set (n=182) by selecting the genes included in the union of GSEA hallmark database (https://www.gsea-msigdb.org/gsea/msigdb/genesets.jsp?collection=H) IFN-α response ('HALLMARK_INTERFERON_ALPHA_RESPONSE', n=97) and IFN-γ response ('HALLMARK_INTERFERON_GAMMA_RESPONSE', n=200) gene sets, but excluded those from the inflammatory response set ('HALLMARK_INFLAMMATORY_RESPONSE', n=200). Similarly, we defined our inflammatory gene set (n=158) by selecting the genes included in the inflammatory response set, but excluded those from the IFN-α and IFN-γ response gene sets **(Supplementary Table S5).**

For each gene set, after computing the enrichment score for each cell as previously explained, we used the function escape::performNormalization to normalize the enrichment scores by accounting for single-cell dropout and producing positive values for downstream differential enrichment analysis[105].

Then, we performed a differential enrichment analysis to explore how the enrichment scores for those gene sets changed between the two groups of CD14+ Monocyte neighborhoods female age-enriched DA neighborhoods and male age-depleted neighborhoods. We fitted a GLMM using a Gaussian distribution and corrected for FDR through the Benjamini–Hochberg method. We considered CD14+

Monocyte neighborhoods to be significantly enriched with sex in a particular gene set at FDR < 0.05. The GLMM used was:

$$\text{AUCell scores} \sim \text{Sex} + \text{Age} + (1|\text{Nhood ID}) + (1|\text{Sequencing Pool}) + (1|\text{Donor})$$

In parallel, for visualization purposes **(Fig. 5D),** we computed the mean AUCell score among neighborhoods with at least 15 cells. Then, we performed a Wilcoxon test to compare the means of the two groups of CD14+ Monocyte neighborhoods (female age-enriched DA neighborhoods and male age-depleted neighborhoods).

Senescent cell identification with SenCID

We used the Python package SenCID (v1.0.0)[52], a machine learning-based method for identifying senescent cells from scRNA-seq data. We use this approach as an orthogonal validation to the AUCell enrichment score for the senescence gene set. SenCID categorizes cells into six major senescence identities (SIDs) and determines the most appropriate identity for each cell. For our analysis, we used the score corresponding to each individual cell's best-matched identity and performed differential enrichment analysis. We modeled how the best senescence score varies with age using the following GLMM with Gaussian distribution :

$$\text{SenCID scores} \sim \text{Age} + \text{Sex} + (1|\text{Sequencing Pool}) + (1|\text{Donor})$$

We applied the Benjamin–Hochberg method to adjust for FDR, considering a cell type as significantly changing in senescence scoring with age at FDR < 0.05.

Only cell types with age-DEGs (18 cell types shown in **Fig. 1A**) were tested.

Overlap of sex-stratified age-DEGs with GWAS hits

The GWAS data was obtained from the GWAS Catalog (https://www.ebi.ac.uk/gwas/) and filtered to include only traits related to autoimmune diseases ("autoimmune diseases") (https://www.ebi.ac.uk/gwas/efotraits/EFO_0005140). We identified genes associated with the strongest SNPs using the "MAPPED GENE(S)" column, including those located within or near these SNPs. Then, per each cell type, we assessed whether sex-stratfied age-DEGs were enriched for these autoimmune-associated genes using the fisher.test function from the R package stats (v4.3.0) with default parameters. We applied the Benjamini–Hochberg method to adjust for FDR, considering a cell type to be significantly enriched with age in a particular gene set at FDR < 0.05.

Sex-stratified cellular compositional and differential gene expression analyses with age on manually curated subsets of CD8+ TEM, CD14+ Monocytes, and B naive cells

To further characterize the sex-specific CD8+ TEM, CD14+ Monocytes and B naive subpopulations (neighborhoods) changing with age identified in our cell-type agnostic age-DA, we classified these cell types based on the expression of top marker genes of these subpopulations. The methodology to identify the gene markers for sex-specific age-DA enriched or depleted neighborhoods was explained in a previous section. According to the summary statistics from our gene markers analysis **(Supplementary Table 5)**, together with previous literature[13,14,71,108,109], we used *GZMB*, *ISG15*, and *ABCA6* gene expression to classify the whole set of CD8+ TEM, CD14+ Monocytes, and B naive cells, respectively. Specifically, we classified CD8+ TEM cells into two sub-cell types based on their expression of *GZMB*, categorizing them as *GZMB*+ (≥ 1 count) or *GZMB*- (< 1 count). Then, we re-ran the sex-stratified differential gene expression analysis with age within these sub-cell types, following the same approach as described previously. We also re-ran the sex-stratified cell type-based composition analysis with age, this time separating CD8+ TEM cells into their *GZMB*+ and *GZMB*- subpopulations. Similarly, we classified CD14+ Monocytes into two sub-cell types based on their expression of *ISG15*, categorizing them as *ISG15*+ (≥ 1 count) or *ISG15*- (< 1 count). Then, we re-ran

the sex-stratified differential gene expression analysis with age within these sub-cell types, following the same approach as described previously. Lastly, we classified B naive cells into two sub-cell types based on their expression of *ABCA6*, categorizing them as *ABCA6*+ (≥ 1 count) or *ABCA6*- (< 1 count). Then, we re-ran the sex-stratified differential gene expression analysis with age within these sub-cell types, following the same approach as described previously.

Replication with external datasets

To validate our main findings from the OneK1K cohort[29] (discovery dataset), we relied on two external datasets[32,14] (replication datasets) of healthy individuals spanning a wide age range. While these two datasets provided valuable replication potential, they differed in sample size and sex distribution compared to the discovery cohort. To accommodate these differences, we adjusted statistical significance thresholds to ensure robust inference where necessary.

Oelen et al 2022[32]

We used a subset of previously processed scRNA-seq data[32] on unstimulated PBMCs from 72 healthy LifeLines Deep donors aged 27-77 years (mean 56.83 ± 12.7) (https://www.eqtlgen.org/sc/datasets/1m-scbloodnl-dataset.html). In short, scRNA-seq data was generated using the 10X Genomics Chromium Single Cell 3' v2 chemistry and libraries were 150 bp paired-end sequenced on Illumina's NovaSeq6000. The Cellranger v3.0.2 pipeline was used with default parameters to demultiplex, generate FASTQ files, map reads to the GRCh37 reference genome and generate a unique molecular identifier (UMI)-corrected count matrix per cell. To assign cells to one of the eight individuals in a lane, Demuxlet was used. After quality control, 54,373 cells remained. To enable a direct comparison between the two datasets, we employed the same cell annotation strategy used for the OneK1K cohort. In detail, we performed automated cell type classification using Azimuth to annotate the cells[30]. We mapped our scRNA-seq query dataset to a recently published CITE-seq scRNA-seq reference dataset of 162,000 PBMCs measured with 228 antibodies[30]. For this process, we used a set of specific functions from the R package Seurat (v4.0.0)[30,31]. First, we normalized the reference dataset using the Seurat::SCTransform function. Next, we found anchors between reference and query using a precomputed supervised PCA (spca) transformation with 50 dimensions using the Seurat::FindTransferAnchors function. We then transferred cell type labels from the reference to the query by projecting the query set on the UMAP embeddings of the reference using the Seurat::MapQuery function. We ended up with a total of 31 circulating immune cell subpopulations. Finally, we filtered the dataset to retain only the 24 cell types present in the OneK1K data after applying specific filters (i.e., at least 5 cells in a minimum of 4 individuals, **Supplementary Table 1**). We used this dataset to replicate the differential gene expression and the cellular composition analyses with age **(Supplementary Table 2, 3)**, as well as the sex-specific abundance changes with age **(Supplementary Table 5)**. To do so, we applied the same pipelines as in the OneK1K data.

Terekhova et al 2023[14]

We used previously processed scRNA-seq data[14] on PBMCs from a short-term longitudinal study. In detail, donors were split into five groups according to their age, falling in 10-year intervals. Each age group followed over the course of 3 years with annual blood collection, which results in a total of 317 samples, with 2 technical replicates each one, from 166 healthy individuals aged 25-85 years (mean 49.38 ± 16.86) (https://www.synapse.org/Synapse:syn50542388). Briefly, scRNA-seq data was obtained using the 10x Genomics Chromium Single Cell 5' v2 chemistry and libraries were sequenced on a NovaSeq S4 flow cell (300 cycles), targeting 25,000 read pairs per cell. In total, 107 sequencing outputs were generated from the 317 samples. Each output was then independently demultiplexed using a combined hashtag-based and genotype-based demultiplexing strategy, including Seurat v4.0.5 *HTODemux* and *minimap2* v2.17-r941, *freebayes* v1.3.2, and *souporcell* v2.5. This strategy improved demultiplexing efficiency and accuracy compared to using a single method. After quality control,

1,916,367 cells remained. We used this dataset to replicate the sex-specific abundance changes with age **(Supplementary Table 5)**. To do so, we applied the same pipelines as in the OneK1K data.

Data and code availability

The original public-access data from[29], including OneK1K single-cell gene expression and genotype data are available via Gene Expression Omnibus (GSE196830). The cell by gene data are available at Human Cell Atlas (HCA) (https://cellxgene.cziscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1). All the custom scripts developed for this study have been made openly accessible and can be found on the GitHub repository at https://github.com/mariasr3/2024_SingleCell_ImmuneAging. Processed data files derived from the analysis conducted in this paper are deposited at Zenodo: https://doi.org/10.5281/zenodo.14035325. By visiting this repository, researchers and interested individuals can access and use the custom scripts for their own analysis or to replicate the study's findings.

Acknowledgements

M.S.R was supported by a predoctoral "AGAUR-FI Joan Oró" fellowship from "Secretaria d'Universitats i Recerca del Departament de Recerca i Universitats de la Generaliat de Catalunya i del Fons Europeu Social Plus" (2024 FI-3 0065) . A.R.C was supported by a predoctoral fellowship "Formación Personal Investigador (FPI)" from "el Ministerio de Ciencia, Innovación y Universidades (MCIN)" and "la Agencia Estatal de Investigación (AEI)" (MCIN/AEI FPI with code PRE2019-090193). M.M. was supported by a grant PID2019-107937GA-I00 funded by MCIN/AEI/10.13039/501100011033 and a grant RYC-2017-22249 funded by MCIN/AEI/10.13039/501100011033 and by "ESF Investing in your future". J.E.P is supported by the National Health and Medical Research Council Fellowship (APP1107599) and with the support of the Fok Family in memory of Dr. and Mrs. Wing Kan Fok. L.F. is supported by a grant from the Dutch Research Council (ZonMW-VICI 09150182010019), and a sponsored research collaboration with Biogen, Roche and Takeda. This work is co-financed by Oncode Institute, which is partly funded by the Dutch Cancer Society. This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101057553. M.G.P.W. is supported by a NWO Vidi grant of the Dutch Research Council (VI.Vidi.223.041). We thank Winona Oliveros, José Miguel Ramirez, and Fairlie Reese for useful discussions at the end of the work.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT to enhance the clarity and language of the text and Perplexity to locate relevant references for the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Author contribution

M.S.R. designed and performed all analysis, A.R.C designed, performed and supervised all analysis.
F.O. performed specific analysis (i.e., overlap of sex-stratified age-DEGs with GWAS hits).
S.B. and J.A.H. performed specific analysis (i.e., OneK1K cell type annotation).
R.O. performed specific analysis (i.e., cell type annotation of the subset of data from[32] used for several replication analyses).
M.G.P.W, L.F, and J.P. provided overall feedback.
M.M. conceived the study and supervised all analysis.
M.S.R., A.R.C, and M.M, wrote the manuscript.

All authors read and agreed on the contents of this manuscript.

Declaration of interests

The authors declare no competing interests.

References

1. Deeks, S.G. (2011). HIV infection, inflammation, immunosenescence, and aging. Annu. Rev. Med. *62*, 141–155.
2. Aw, D., Silva, A.B., and Palmer, D.B. (2007). Immunosenescence: emerging challenges for an ageing population. Immunology *120*, 435–446.
3. Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K., et al. (2015). The transcriptional landscape of age in human peripheral blood. Nat. Commun. *6*, 8570.
4. Goronzy, J.J., and Weyand, C.M. (2013). Understanding immunosenescence to improve responses to vaccines. Nat. Immunol. *14*, 428–436.
5. Yousefzadeh, M.J., Flores, R.R., Zhu, Y., Schmiechen, Z.C., Brooks, R.W., Trussoni, C.E., Cui, Y., Angelini, L., Lee, K.-A., McGowan, S.J., et al. (2021). An aged immune system drives senescence and ageing of solid organs. Nature *594*, 100–105.
6. Widjaja, A.A., Lim, W.-W., Viswanathan, S., Chothani, S., Corden, B., Dasan, C.M., Goh, J.W.T., Lim, R., Singh, B.K., Tan, J., et al. (2024). Inhibition of IL-11 signalling extends mammalian healthspan and lifespan. Nature *632*, 157–165.
7. Sayed, N., Huang, Y., Nguyen, K., Krejciova-Rajaniemi, Z., Grawe, A.P., Gao, T., Tibshirani, R., Hastie, T., Alpert, A., Cui, L., et al. (2021). An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging. Nat. Aging *1*, 598–615.
8. Martínez de Toda, I., Maté, I., Vida, C., Cruces, J., and De la Fuente, M. (2016). Immune function parameters as markers of biological age and predictors of longevity. Aging *8*, 3110–3119.
9. Tran Van Hoi, E., De Glas, N.A., Portielje, J.E.A., Van Heemst, D., Van Den Bos, F., Jochems, S.P., and Mooijaart, S.P. (2023). Biomarkers of the ageing immune system and their association with frailty - A systematic review. Exp. Gerontol. *176*, 112163.
10. Borgoni, S., Kudryashova, K.S., Burka, K., and de Magalhães, J.P. (2021). Targeting immune dysfunction in aging. Ageing Res. Rev. *70*, 101410.
11. Huang, Z., Chen, B., Liu, X., Li, H., Xie, L., Gao, Y., Duan, R., Li, Z., Zhang, J., Zheng, Y., et al. (2021). Effects of sex and aging on the immune cell landscape as assessed by single-cell transcriptomic analysis. Proc. Natl. Acad. Sci. U. S. A. *118*. https://doi.org/10.1073/pnas.2023216118.
12. Luo, O.J., Lei, W., Zhu, G., Ren, Z., Xu, Y., Xiao, C., Zhang, H., Cai, J., Luo, Z., Gao, L., et al. (2022). Multidimensional single-cell analysis of human peripheral blood reveals characteristic features of the immune system landscape in aging and frailty. Nat Aging *2*, 348–364.
13. Mogilenko, D.A., Shpynov, O., Andhey, P.S., Arthur, L., Swain, A., Esaulova, E., Brioschi, S., Shchukina, I., Kerndl, M., Bambouskova, M., et al. (2021). Comprehensive Profiling of an Aging Immune System Reveals Clonal GZMK+ CD8+ T Cells as Conserved Hallmark of Inflammaging. Immunity *54*, 99–115.e12.
14. Terekhova, M., Swain, A., Bohacova, P., Aladyeva, E., Arthur, L., Laha, A., Mogilenko, D.A., Burdess, S., Sukhov, V., Kleverov, D., et al. (2023). Single-cell atlas of healthy human blood unveils age-related loss of NKG2C+GZMB-CD8+ memory T cells and accumulation of type 2 memory T cells. Immunity *56*, 2836–2854.e9.
15. Carr, E.J., Dooley, J., Garcia-Perez, J.E., Lagou, V., Lee, J.C., Wouters, C., Meyts, I., Goris, A., Boeckxstaens, G., Linterman, M.A., et al. (2016). The cellular composition of the human immune system is shaped by age and cohabitation. Nat. Immunol. *17*, 461–468.
16. Gong, Q., Sharma, M., Kuan, E.L., Glass, M.C., Chander, A., Singh, M., Graybuck, L.T., Thomson, Z.J., LaFrance, C.M., Zaim, S.R., et al. (2024). Longitudinal Multi-omic Immune Profiling Reveals Age-Related Immune Cell Dynamics in Healthy Adults. Immunology.
17. Alpert, A., Pickman, Y., Leipold, M., Rosenberg-Hasson, Y., Ji, X., Gaujoux, R., Rabani, H., Starosvetsky, E., Kveler, K., Schaffert, S., et al. (2019). A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. Nat. Med. *25*, 487–495.
18. Wertheimer, A.M., Bennett, M.S., Park, B., Uhrlaub, J.L., Martinez, C., Pulko, V., Currier, N.L., Nikolich-Žugich, D., Kaye, J., and Nikolich-Žugich, J. (2014). Aging and cytomegalovirus infection differentially and jointly affect distinct circulating T cell subsets in humans. J. Immunol. *192*, 2143–2155.
19. Zheng, Y., Liu, X., Le, W., Xie, L., Li, H., Wen, W., Wang, S., Ma, S., Huang, Z., Ye, J., et al. (2020). A human circulating immune cell landscape in aging and COVID-19. Protein Cell *11*, 740–770.
20. Zhu, H., Chen, J., Liu, K., Gao, L., Wu, H., Ma, L., Zhou, J., Liu, Z., and Han, J.-D.J. (2023). Human PBMC scRNA-seq-based aging clocks reveal ribosome to inflammation balance as a single-cell aging hallmark and

super longevity. Sci Adv *9*, eabq7599.

21. Holdcroft, A. (2007). Gender bias in research: how does it affect evidence based medicine? J. R. Soc. Med. *100*, 2–3.
22. Klein, S.L., and Flanagan, K.L. (2016). Sex differences in immune responses. Nat. Rev. Immunol. *16*, 626–638.
23. Márquez, E.J., Chung, C.-H., Marches, R., Rossi, R.J., Nehar-Belaid, D., Eroglu, A., Mellert, D.J., Kuchel, G.A., Banchereau, J., and Ucar, D. (2020). Sexual-dimorphism in human immune system aging. Nat. Commun. *11*, 751.
24. Hägg, S., and Jylhävä, J. (2021). Sex differences in biological aging with a focus on human studies. Elife *10*. https://doi.org/10.7554/eLife.63425.
25. Furman, D., Hejblum, B.P., Simon, N., Jojic, V., Dekker, C.L., Thiébaut, R., Tibshirani, R.J., and Davis, M.M. (2014). Systems analysis of sex differences reveals an immunosuppressive role for testosterone in the response to influenza vaccination. Proc. Natl. Acad. Sci. U. S. A. *111*, 869–874.
26. Fischinger, S., Boudreau, C.M., Butler, A.L., Streeck, H., and Alter, G. (2019). Sex differences in vaccine-induced humoral immunity. Semin. Immunopathol. *41*, 239–249.
27. Jacobson, D.L., Gange, S.J., Rose, N.R., and Graham, N.M. (1997). Epidemiology and estimated population burden of selected autoimmune diseases in the United States. Clin. Immunol. Immunopathol. *84*, 223–243.
28. Connolly, E., Pan, T., Aluru, M., Chockalingam, S., Dhere, V., and Gibson, G. (2024). Loss of immune cell identity with age inferred from large atlases of single cell transcriptomes. Aging Cell, e14306.
29. Yazar, S., Alquicira-Hernandez, J., Wing, K., Senabouth, A., Gordon, M.G., Andersen, S., Lu, Q., Rowson, A., Taylor, T.R.P., Clarke, L., et al. (2022). Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease. Science *376*, eabf3041.
30. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29.
31. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502.
32. Oelen, R., de Vries, D.H., Brugge, H., Gordon, M.G., Vochteloo, M., single-cell eQTLGen consortium, BIOS Consortium, Ye, C.J., Westra, H.-J., Franke, L., et al. (2022). Single-cell RNA-sequencing of peripheral blood mononuclear cells reveals widespread, context-specific gene expression regulation upon pathogenic exposure. Nat. Commun. *13*, 3267.
33. Metcalf, T.U., Wilkinson, P.A., Cameron, M.J., Ghneim, K., Chiang, C., Wertheimer, A.M., Hiscott, J.B., Nikolich-Zugich, J., and Haddad, E.K. (2017). Human Monocyte Subsets Are Transcriptionally and Functionally Altered in Aging in Response to Pattern Recognition Receptor Agonists. J. Immunol. *199*, 1405–1417.
34. Gismondi, A., and Santoni, A. (2006). Migration of NK cells. In Lymphocyte Trafficking in Health and Disease, R. Badolato and S. Sozzani, eds. (Birkhäuser Basel), pp. 95–112.
35. Beli, E., Clinthorne, J.F., Duriancik, D.M., Hwang, I., Kim, S., and Gardner, E.M. (2011). Natural killer cell function is altered during the primary response of aged mice to influenza infection. Mech. Ageing Dev. *132*, 503–510.
36. Fang, M., Roscoe, F., and Sigal, L.J. (2010). Age-dependent susceptibility to a viral disease due to decreased natural killer cell numbers and trafficking. J. Exp. Med. *207*, 2369–2381.
37. Frasca, D., and Blomberg, B.B. (2011). Aging affects human B cell responses. J. Clin. Immunol. *31*, 430–435.
38. Mogilenko, D.A., Shchukina, I., and Artyomov, M.N. (2022). Immune ageing at single-cell resolution. Nat. Rev. Immunol. *22*, 484–498.
39. Sansoni, P., Cossarizza, A., Brianti, V., Fagnoni, F., Snelli, G., Monti, D., Marcato, A., Passeri, G., Ortolani, C., and Forti, E. (1993). Lymphocyte subsets and natural killer cell activity in healthy old people and centenarians [see comments]. Blood *82*, 2767–2773.
40. Guo, C., Wu, M., Huang, B., Zhao, R., Jin, L., Fu, B., Wang, P., Wang, D., Zheng, M., Fang, J., et al. (2022). Single-cell transcriptomics reveal a unique memory-like NK cell subset that accumulates with ageing and correlates with disease severity in COVID-19. Genome Med. *14*, 46.
41. Frasca, D., and Blomberg, B.B. (2020). Aging induces B cell defects and decreased antibody responses to influenza infection and vaccination. Immun. Ageing *17*, 37.
42. Marrack, P., and Kappler, J. (2004). Control of T cell viability. Annu. Rev. Immunol. *22*, 765–787.
43. Jameson, S.C. (2002). Maintaining the norm: T-cell homeostasis. Nat. Rev. Immunol. *2*, 547–556.
44. Ma, A., Koka, R., and Burkett, P. (2006). Diverse functions of IL-2, IL-15, and IL-7 in lymphoid homeostasis.

Annu. Rev. Immunol. *24*, 657–679.

45. Surh, C.D., and Sprent, J. (2008). Homeostasis of naive and memory T cells. Immunity *29*, 848–862.
46. Sun, J.C., Beilke, J.N., Bezman, N.A., and Lanier, L.L. (2011). Homeostatic proliferation generates long-lived natural killer cells that respond against viral infection. J. Exp. Med. *208*, 357–368.
47. Goronzy, J.J., Fang, F., Cavanagh, M.M., Qi, Q., and Weyand, C.M. (2015). Naive T cell maintenance and function in human aging. J. Immunol. *194*, 4073–4080.
48. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., et al. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat. Biotechnol. *42*, 293–304.
49. Mitra, M., Ho, L.D., and Coller, H.A. (2018). An In Vitro Model of Cellular Quiescence in Primary Human Dermal Fibroblasts. Methods Mol. Biol. *1686*, 27–47.
50. Saul, D., Kosinsky, R.L., Atkinson, E.J., Doolittle, M.L., Zhang, X., LeBrasseur, N.K., Pignolo, R.J., Robbins, P.D., Niedernhofer, L.J., Ikeno, Y., et al. (2022). A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. Nat. Commun. *13*, 4827.
51. Zheng, S.C., Stein-O'Brien, G., Augustin, J.J., Slosberg, J., Carosso, G.A., Winer, B., Shin, G., Bjornsson, H.T., Goff, L.A., and Hansen, K.D. (2022). Universal prediction of cell-cycle position using transfer learning. Genome Biol. *23*, 41.
52. 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.
53. Huang, W., Hickson, L.J., Eirin, A., Kirkland, J.L., and Lerman, L.O. (2022). Cellular senescence: the good, the bad and the unknown. Nat. Rev. Nephrol. *18*, 611–627.
54. Mittelbrunn, M., and Kroemer, G. (2021). Hallmarks of T cell aging. Nat. Immunol. *22*, 687–698.
55. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196.
56. seurat (Github).
57. Giefing-Kröll, C., Berger, P., Lepperdinger, G., and Grubeck-Loebenstein, B. (2015). How sex and age affect immune responses, susceptibility to infections, and response to vaccination. Aging Cell *14*, 309–321.
58. Lisci, M., Barton, P.R., Randzavola, L.O., Ma, C.Y., Marchingo, J.M., Cantrell, D.A., Paupe, V., Prudent, J., Stinchcombe, J.C., and Griffiths, G.M. (2021). Mitochondrial translation is required for sustained killing by cytotoxic T cells. Science *374*, eabe9977.
59. Schriml, L.M., Arze, C., Nadendla, S., Chang, Y.-W.W., Mazaitis, M., Felix, V., Feng, G., and Kibbe, W.A. (2012). Disease Ontology: a backbone for disease semantic integration. Nucleic Acids Res. *40*, D940–D946.
60. Aggarwal, N.R., Patel, H.N., Mehta, L.S., Sanghani, R.M., Lundberg, G.P., Lewis, S.J., Mendelson, M.A., Wood, M.J., Volgman, A.S., and Mieres, J.H. (2018). Sex Differences in Ischemic Heart Disease. Circ. Cardiovasc. Qual. Outcomes *11*, e004437.
61. Segerberg, F., Lundtoft, C., Reid, S., Hjorton, K., Leonard, D., Nordmark, G., Carlsten, M., and Hagberg, N. (2019). Autoantibodies to Killer Cell Immunoglobulin-Like Receptors in Patients With Systemic Lupus Erythematosus Induce Natural Killer Cell Hyporesponsiveness. Front. Immunol. *10*, 2164.
62. Poggi, A., and Zocchi, M.R. (2014). NK cell autoreactivity and autoimmune diseases. Front. Immunol. *5*, 27.
63. Hudspeth, K., Pontarini, E., Tentorio, P., Cimino, M., Donadon, M., Torzilli, G., Lugli, E., Della Bella, S., Gershwin, M.E., and Mavilio, D. (2013). The role of natural killer cells in autoimmune liver disease: a comprehensive review. J. Autoimmun. *46*, 55–65.
64. Schleinitz, N., Vély, F., Harlé, J.-R., and Vivier, E. (2010). Natural killer cells in human autoimmune diseases. Immunology *131*, 451–458.
65. Selathurai, A., Deswaerte, V., Kanellakis, P., Tipping, P., Toh, B.-H., Bobik, A., and Kyaw, T. (2014). Natural killer (NK) cells augment atherosclerosis by cytotoxic-dependent mechanisms. Cardiovasc. Res. *102*, 128–137.
66. Whitman, S.C., Rateri, D.L., Szilvassy, S.J., Yokoyama, W., and Daugherty, A. (2004). Depletion of natural killer cell function decreases atherosclerosis in low-density lipoprotein receptor null mice. Arterioscler. Thromb. Vasc. Biol. *24*, 1049–1054.
67. Seldin, M.F. (2015). The genetics of human autoimmune disease: A perspective on progress in the field and future directions. J. Autoimmun. *64*, 1–12.
68. Bove, R. (2013). Autoimmune diseases and reproductive aging. Clin. Immunol. *149*, 251–264.
69. Desai, M.K., and Brinton, R.D. (2019). Autoimmune disease in women: Endocrine transition and risk across

the lifespan. Front. Endocrinol. *10*, 265.

70. Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., et al. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. *51*, D977–D985.
71. Park, H.-J., Shin, M.S., Kim, M., Bilsborrow, J.B., Mohanty, S., Montgomery, R.R., Shaw, A.C., You, S., and Kang, I. (2019). Transcriptomic analysis of human IL-7 receptor alpha low and high effector memory CD8+ T cells reveals an age-associated signature linked to influenza vaccine response in older adults. Aging Cell *18*, e12960.
72. Dann, E., Henderson, N.C., Teichmann, S.A., Morgan, M.D., and Marioni, J.C. (2022). Differential abundance testing on single-cell data using k-nearest neighbor graphs. Nat. Biotechnol. *40*, 245–253.
73. Alice, K., Marioni, J.C., and Morgan, M.D. (2023). Milo2.0 unlocks population genetic analyses of cell state abundance using a count-based mixed model. bioRxiv, 2023.11.08.566176. https://doi.org/10.1101/2023.11.08.566176.
74. Zhao, J., Jaffe, A., Li, H., Lindenbaum, O., Sefik, E., Jackson, R., Cheng, X., Flavell, R.A., and Kluger, Y. (2021). Detection of differentially abundant cell subpopulations in scRNA-seq data. Proc. Natl. Acad. Sci. U. S. A. *118*. https://doi.org/10.1073/pnas.2100293118.
75. Burkhardt, D.B., Stanley, J.S., 3rd, Tong, A., Perdigoto, A.L., Gigante, S.A., Herold, K.C., Wolf, G., Giraldez, A.J., van Dijk, D., and Krishnaswamy, S. (2021). Quantifying the effect of experimental perturbations at single-cell resolution. Nat. Biotechnol. *39*, 619–629.
76. Novak, J., Dobrovolny, J., Novakova, L., and Kozak, T. (2014). The decrease in number and change in phenotype of mucosal-associated invariant T cells in the elderly and differences in men and women of reproductive age. Scand. J. Immunol. *80*, 271–275.
77. Vojdani, A., Koksoy, S., Vojdani, E., Engelman, M., Benzvi, C., and Lerner, A. (2024). Natural Killer Cells and Cytotoxic T Cells: Complementary Partners against Microorganisms and Cancer. Microorganisms *12*. https://doi.org/10.3390/microorganisms12010230.
78. Sutherland, T.E., Dyer, D.P., and Allen, J.E. (2023). The extracellular matrix and the immune system: A mutually dependent relationship. Science *379*, eabp8964.
79. Shchukina, I., Bagaitkar, J., Shpynov, O., Loginicheva, E., Porter, S., Mogilenko, D.A., Wolin, E., Collins, P., Demidov, G., Artomov, M., et al. (2021). Enhanced epigenetic profiling of classical human monocytes reveals a specific signature of healthy aging in the DNA methylome. Nat. Aging *1*, 124–141.
80. Rubtsova, K., Rubtsov, A.V., Cancro, M.P., and Marrack, P. (2015). Age-associated B cells: A T-bet-dependent effector with roles in protective and pathogenic immunity. J. Immunol. *195*, 1933–1937.
81. Kimmel, J.C., Penland, L., Rubinstein, N.D., Hendrickson, D.G., Kelley, D.R., and Rosenthal, A.Z. (2019). Murine single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging. Genome Res. *29*, 2088–2103.
82. Russell Knode, L.M., Naradikian, M.S., Myles, A., Scholz, J.L., Hao, Y., Liu, D., Ford, M.L., Tobias, J.W., Cancro, M.P., and Gearhart, P.J. (2017). Age-associated B cells express a diverse repertoire of VH and Vκ genes with somatic hypermutation. J. Immunol. *198*, 1921–1927.
83. Andreu-Sánchez, S., Ripoll-Cladellas, A., Culinscaia, A., Bulut, O., Bourgonje, A.R., Netea, M.G., Lansdorp, P., Aubert, G., Bonder, M.J., Franke, L., et al. (2024). Antibody signatures against viruses and microbiome reflect past and chronic exposures and associate with aging and inflammation. iScience *27*, 109981.
84. Ramirez, J.M., Ribeiro, R., Soldatkina, O., Moraes, A., García-Pérez, R., Ferreira, P.G., and Melé, M. (2024). The molecular impact of cigarette smoking resembles aging across tissues. bioRxiv, 2024.03.14.585016. https://doi.org/10.1101/2024.03.14.585016.
85. Sánchez-Valle, J., and Valencia, A. (2023). Molecular bases of comorbidities: present and future perspectives. Trends Genet. *39*, 773–786.
86. Sánchez-Valle, J., Tejero, H., Fernández, J.M., Juan, D., Urda-García, B., Capella-Gutiérrez, S., Al-Shahrour, F., Tabarés-Seisdedos, R., Baudot, A., Pancaldi, V., et al. (2020). Interpreting molecular similarity between patients as a determinant of disease comorbidity relationships. Nat. Commun. *11*, 2854.
87. Levine, M.E., Lu, A.T., Quach, A., Chen, B.H., Assimes, T.L., Bandinelli, S., Hou, L., Baccarelli, A.A., Stewart, J.D., Li, Y., et al. (2018). An epigenetic biomarker of aging for lifespan and healthspan. Aging *10*, 573–591.
88. Lu, A.T., Quach, A., Wilson, J.G., Reiner, A.P., Aviv, A., Raj, K., Hou, L., Baccarelli, A.A., Li, Y., Stewart, J.D., et al. (2019). DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging *11*, 303–327.
89. Belsky, D.W., Caspi, A., Arseneault, L., Baccarelli, A., Corcoran, D.L., Gao, X., Hannon, E., Harrington, H.L., Rasmussen, L.J., Houts, R., et al. (2020). Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. Elife *9*. https://doi.org/10.7554/eLife.54870.

90. Mi, T., Soerens, A.G., Alli, S., Kang, T.G., Vasandan, A.B., Wang, Z., Vezys, V., Kimura, S., Iacobucci, I., Baylin, S.B., et al. (2024). Conserved epigenetic hallmarks of T cell aging during immunity and malignancy. Nat Aging *4*, 1053–1063.

91. Zhang, Z., Reynolds, S.R., Stolrow, H.G., Chen, J.-Q., Christensen, B.C., and Salas, L.A. (2024). Deciphering the role of immune cell composition in epigenetic age acceleration: Insights from cell-type deconvolution applied to human blood epigenetic clocks. Aging Cell *23*, e14071.

92. Bonder, M.J., Clark, S.J., Krueger, F., Luo, S., Agostinho de Sousa, J., Hashtroud, A.M., Stubbs, T.M., Stark, A.-K., Rulands, S., Stegle, O., et al. (2024). scEpiAge: an age predictor highlighting single-cell ageing heterogeneity in mouse blood. Nat. Commun. *15*, 7567.

93. Andreu-Sánchez, S., Aubert, G., Ripoll-Cladellas, A., Henkelman, S., Zhernakova, D.V., Sinha, T., Kurilshikov, A., Cenit, M.C., Jan Bonder, M., Franke, L., et al. (2022). Genetic, parental and lifestyle factors influence telomere length. Commun Biol *5*, 565.

94. Hoffman, G.E., Lee, D., Bendl, J., Fnu, P., Hong, A., Casey, C., Alvia, M., Shao, Z., Argyriou, S., Therrien, K., et al. (2023). Efficient differential expression analysis of large-scale single cell transcriptomics data using dreamlet. bioRxiv. https://doi.org/10.1101/2023.03.17.533005.

95. Kolde, R. (2010). pheatmap: Pretty Heatmaps. (The R Foundation). https://doi.org/10.32614/cran.package.pheatmap https://doi.org/10.32614/cran.package.pheatmap.

96. Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics *31*, 3718–3720.

97. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296.

98. Benjamini, Y., and Hochberg, Y. (1997). Multiple hypotheses testing with weights. Scand. Stat. Theory Appl. *24*, 407–418.

99. Lun, A.T.L., Richard, A.C., and Marioni, J.C. (2017). Testing for differential abundance in mass cytometry data. Nat. Methods *14*, 707–709.

100. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.

101. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

102. Gene Ontology Consortium, Aleksander, S.A., Balhoff, J., Carbon, S., Cherry, J.M., Drabkin, H.J., Ebert, D., Feuermann, M., Gaudet, P., Harris, N.L., et al. (2023). The Gene Ontology knowledgebase in 2023. Genetics *224*. https://doi.org/10.1093/genetics/iyad031.

103. Sayols, S. (2023). rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms. MicroPubl Biol *2023*. https://doi.org/10.17912/micropub.biology.000811.

104. Yu, G., Wang, L.-G., Yan, G.-R., and He, Q.-Y. (2015). DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics *31*, 608–609.

105. Borcherding, N., Vishwakarma, A., Voigt, A.P., Bellizzi, A., Kaplan, J., Nepple, K., Salem, A.K., Jenkins, R.W., Zakharia, Y., and Zhang, W. (2021). Mapping the immune environment in clear cell renal carcinoma by single-cell genomics. Commun Biol *4*, 122.

106. 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.

107. Aquino, Y., Bisiaux, A., Li, Z., O'Neill, M., Mendoza-Revilla, J., Merkling, S.H., Kerner, G., Hasan, M., Libri, V., Bondet, V., et al. (2023). Dissecting human population variation in single-cell responses to SARS-CoV-2. Nature *621*, 120–128.

108. Perng, Y.-C., and Lenschow, D.J. (2018). ISG15 in antiviral immunity and beyond. Nat. Rev. Microbiol. *16*, 423–439.

109. Der, S.D., Zhou, A., Williams, B.R., and Silverman, R.H. (1998). Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. Proc. Natl. Acad. Sci. U. S. A. *95*, 15623–15628.

110. Sopena-Rios, M., Ripoll-Cladellas, A., Omid, F., Ballouz, S., Alquicira-Hernández, J., Oelen, R., Franke, L., van der Wijst, M., Powell, J.E., and Melé, M. (2024). Single-cell atlas of the human immune system reveals sex-specific dynamics of immunosenescence. Preprint at Zenodo, https://doi.org/10.5281/ZENODO.14035324 https://doi.org/10.5281/ZENODO.14035324.

Supplementary files

Document S1. Figures S1–S7

Table S1. Cell type names, related to Supplementary Figure S1

Table S2. Transcriptional changes with age in human circulating immune cells, related to Figures 1 and 2. A. Genes differentially expressed with age across cell types. **B.** Genes differentially expressed with age across cell types when downsampling to the same number of donors **C.** Cell type sharing of genes differentially expressed with age. Gene ontology enrichment of highly shared genes (DE > 6 cell types) in **D.** cluster related to translation pathways and **E.** cluster related to immunological processes. **F.** Gene ontology enrichment of cell type specific differentially expressed genes **H.** Genes differentially expressed with age across cell types in a replication dataset. **I.** Gene ontology enrichment of shared genes (DE >= 2 cell types) in the replication dataset.

Table S3. Cell type composition changes in human circulating immune cells with age, related to Figure 3. A. Cellular compositional analysis with age across cell types. **B.** Cellular compositional analysis with age across cell types in a replication dataset. Cell cycle compositional analysis with age inferred from **D.** Seurat and **E.** Tricycle. Differential enrichment analysis results in **F.** a quiesence gene set, **G.** a senescence gene set and **H.** senescence cells inferred from SenCID.

Table S4. Sex-stratified transcriptional changes in human circulating immune cells with age, related to Figure 4. Genes differentially expressed with age across cell types in **A.** females, **B.** males and **C.** sex:age interaction. Cellular compositional analysis with age across cell types in **D.** females, **E.** males and **F.** interaction age:sex. Gene ontology enrichment of female-specific up-regulated age-DEGs in **G.** CD8+ TEM and **H.** NK and **I.** male-specific down-regulated age-DEGs. **J.** Disease ontology enrichment of sex-specific age-DEGs. GWAS overlap with **K.** all age-DEGs **I.** sex-specific age-DEGs.

Table S5. Cell type-agnostic differential abundance changes in human circulating immune cells with age, related to Figure 5. A. Sex-stratified differential abundance results with age. Marker genes of age differentially expressed neighborhoods **B.** age-enriched in females, **C.** age-enriched in males, **D.** age-depleted in females and **E.** age-depleted in males. **E.** Genes differentially expressed with age in CD8+ TEM *GZMB*+ and CD8+ TEM *GZMB*- cells . **F.** Cellular compositional analysis with age in CD8+ TEM *GZMB*+ and CD8+ TEM *GZMB*- cells. **G.** Gene ontology enrichment of differentially expressed genes in CD8+ TEM *GZMB*+ and CD8+ TEM *GZMB*- cells **H.** Cellular enrichments in inflammatory and interferon gene sets. Genes differentially expressed with age in **I.** ISG- CD14 Monocytes and **J.** *CD5*+ B cells. **K.** Differential abundance results with age. Marker genes of age differentially expressed neighborhoods. **L-M**. Compositional analysis with age in cellular subpopulations using external datasets to replicate the differential abundance results

Table S6. Replication of cell type-agnostic differential abundance changes with age, related to Supplementary Figure S6. Differential abundance results in several data subsamplings (25% of the data).

Supplementary Figures

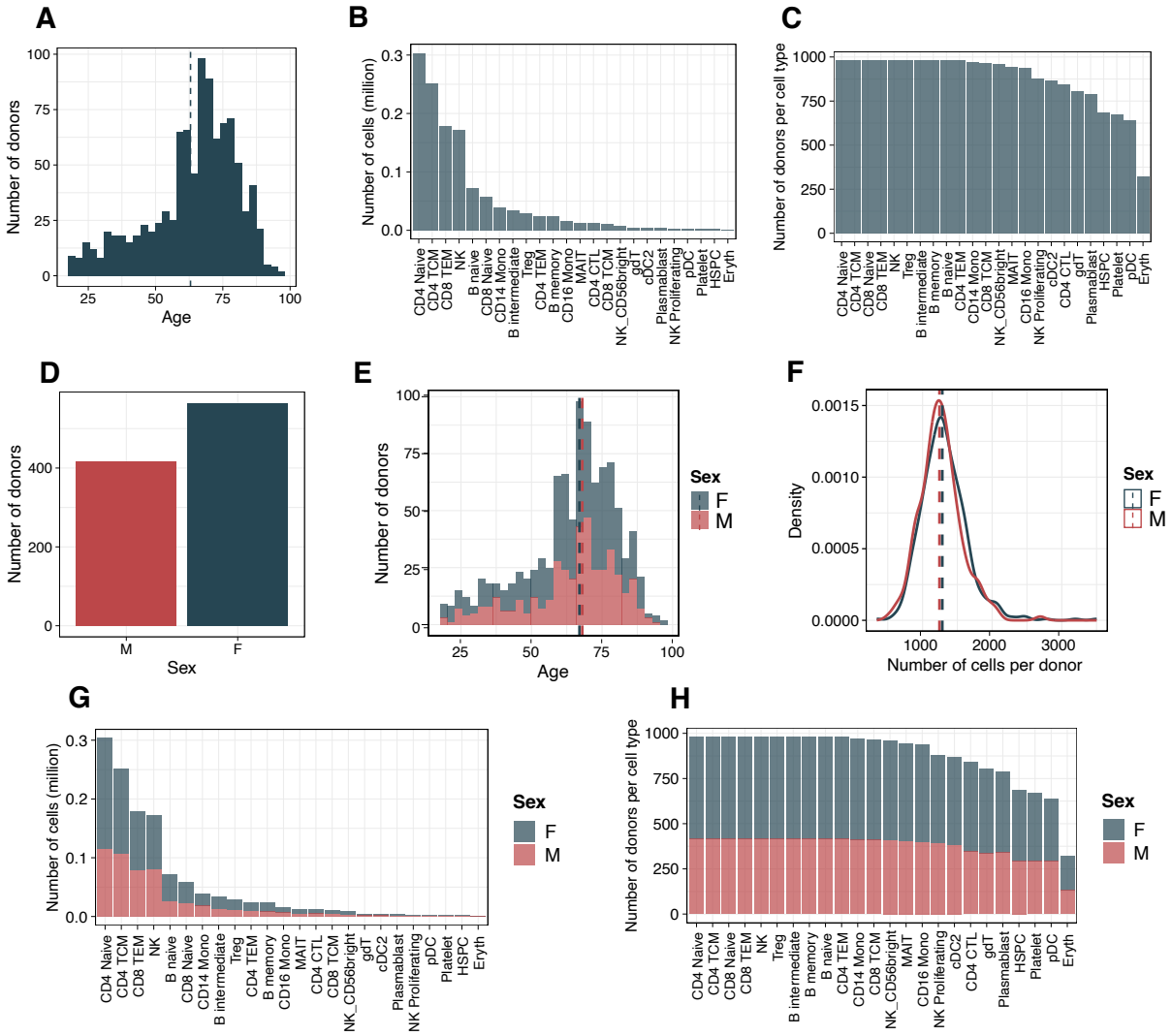

Fig S1. Data overview. a. Distribution number of donors across the ages. **b.** Number of cells per cell type. **c.** Number of donors that have a particular cell type **d.** Number of donors stratified by sex **e.** Distribution number of donors across the ages per sex. **f.** Distribution of number of cells per donor. The dashed line indicates the median (females= 1248, males = 1288) **g.** Number of cells per cell type and sex. **h.** Number of donors that have a particular cell type stratified by sex.

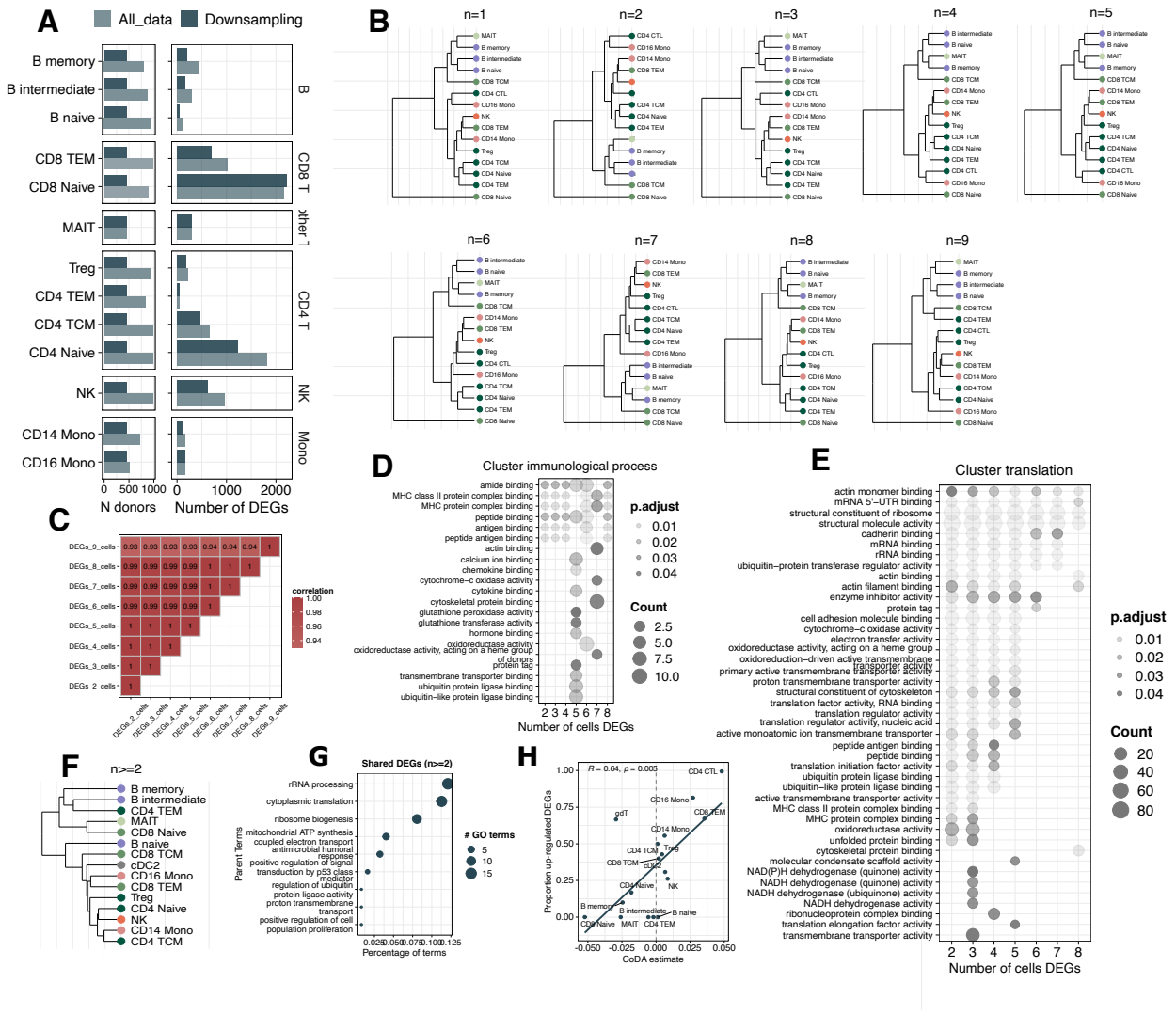

Fig S2. Differential expression analysis with age. **a.** Left. Total number of donors analyzed per cell type. Number of age-DEGs per cell type. Light colors correspond to the results on running the analysis to all data. Dark color correspond to tDark color correspond to tmean number of age-DEGs obtained from ten downsampling analysis. In this analysis, the dataset was randomly downsampled 10 times to match the minimum number of donors detected in MAIT (i.e., 553 donors). **b.** Dendrograms showing the hierarchical clustering of age-DEGs shared across different numbers of cell types. **c.** Heatmap displaying the correlation values between pairs of dendrograms generated in panel c. **d.** Functional enrichment analyses of cluster of age-DEGs shared across varying numbers of cell types related to translation **e.** same as e but for cluster related to immunological process **f.** Dendrograms generated by selecting age-DEGs shared in 2 or more cell types in the replication dataset **g.** Functional enrichment analyses of age-DEGs share in 2 or more cell types in the replication dataset. **h.** Correlation between proportion of age-up-regulated genes in each cell type (y-axis) versus the estimate of the cell type composition analysis with age (x-axis).

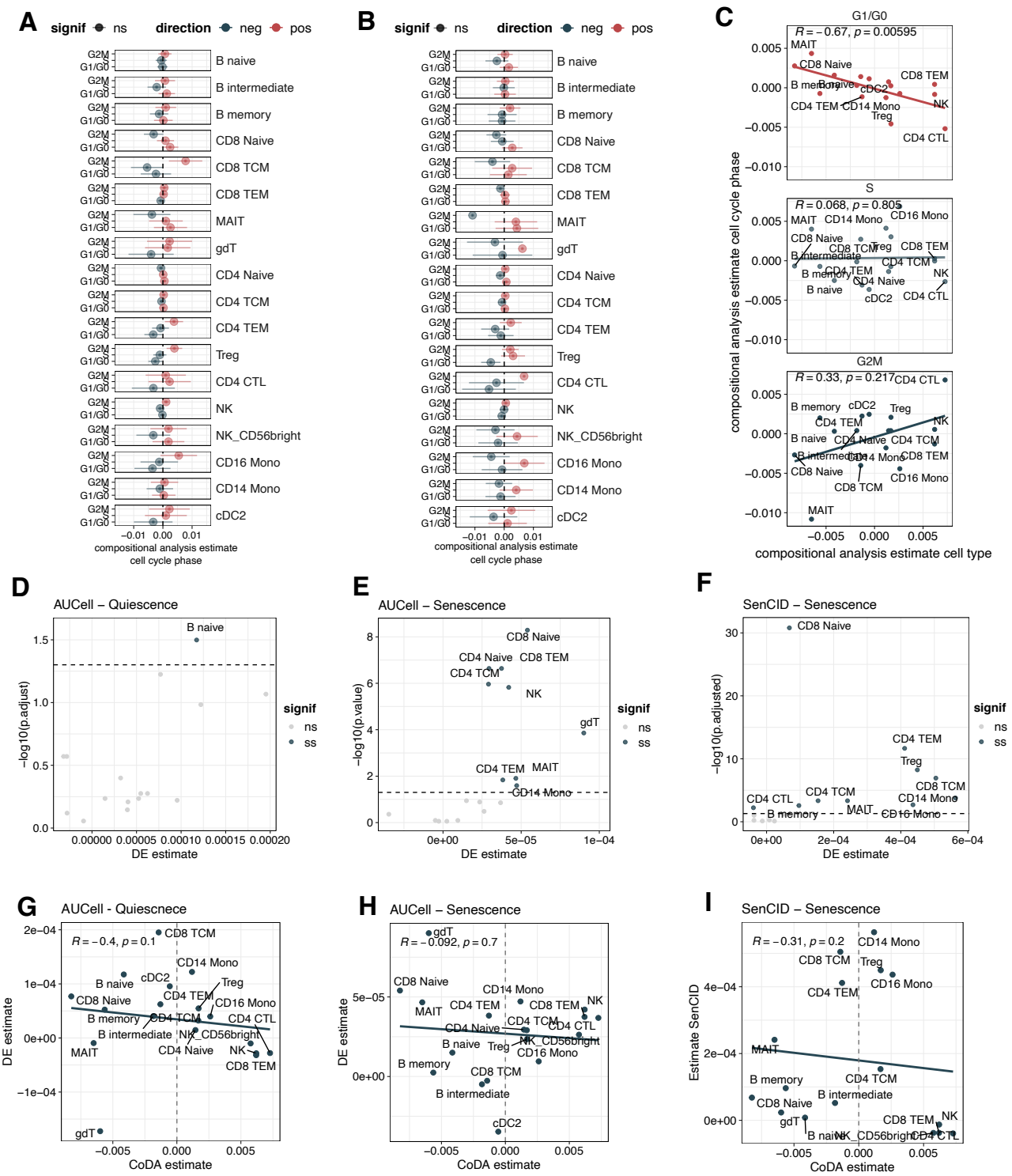

Fig S3. Cell type composition changes with age. Estimates of the cell cycle phase compositional analysis with age across cell types. Bar length indicates the interval of confidence. Cell cycle phase estimates computed with **a.** Seurat (ref) and **b.** Tricycle (ref) **c.** Correlation of age-compositional estimate of each cell cycle phase (G1/G0, S and G2M) (y-axis) versus the age-compositional estimate of cell type proportions (x-axis). Cell cycle estimates computed with Tricycle (ref). Volcano plot of the differential enrichment results for **d.** quiescent gene set **e.** senescent gene set **f.** senescent cells' prediction score with SenCID. Correlation between the differential enrichment estimate in each cell type (y-axis) versus the corresponding cellular CoDA estimate with age (x-axis) for **g.** quiescent gene set **h.** senescent gene set **i.** senescent cells' prediction score with SenCID

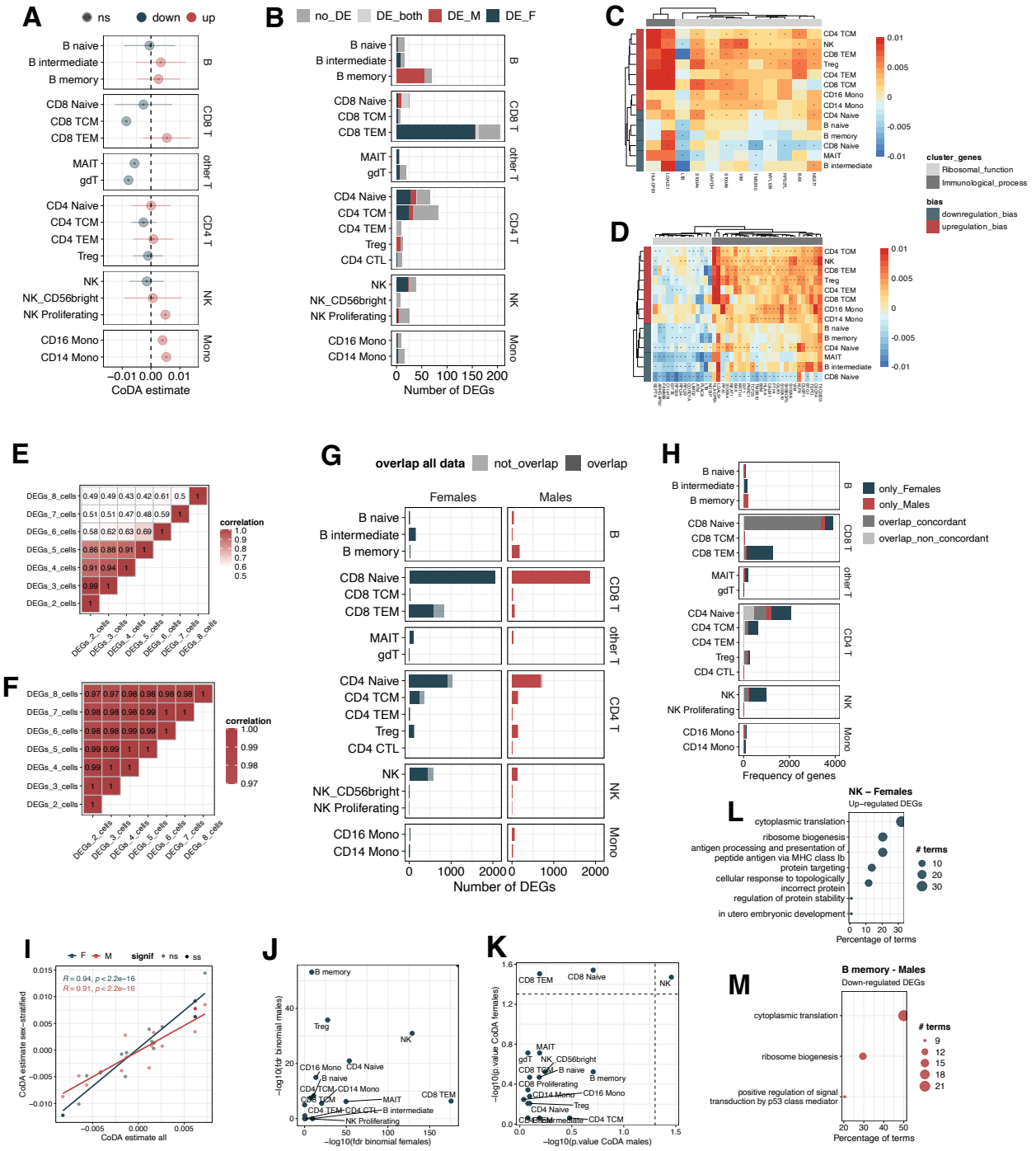

Fig S4. Sex-stratified changes in gene expression and cell type composition with age a. Cellular CoDA estimates for the interaction term Age:Sex across cell types. Bar length indicates the interval of confidence. **b.** Classification of the DEGs (p-value < 0.01) from modelling the interaction term Age:Sex based on the sex-stratified age-DEA analysis. Age-logFC per cell type of the highly shared age-DEGs (age-DEGs > 5 cell types) in **c.** males **d.** females. Heatmap of correlation values between pairs of dendrograms generated by selecting age-DEGs shared across varying numbers of cell types in **e.** males **f.** females **g.** Overlap between the sex-stratified age-DEA analysis and the same analysis performed with all the data **h.** Overlap of the age-DEGs in each sex-stratified analysis **i.** Correlation between CoDA estimates with age in males and females (y-axis) versus the CoDA estimates with all the data (x-axis). **j.** FDR values of the binomial test of directionality in age-DEGs in males (y-axis) versus the same values in females **k.** p.values of the CoDA analysis with age in males (y-axis) versus the same values in females **l.** Reduction of enriched GO term of female-specific up-regulated age-DEGs in NK cells **m.** same plot but for male-specific down-regulated age-DEGs in B memory cells

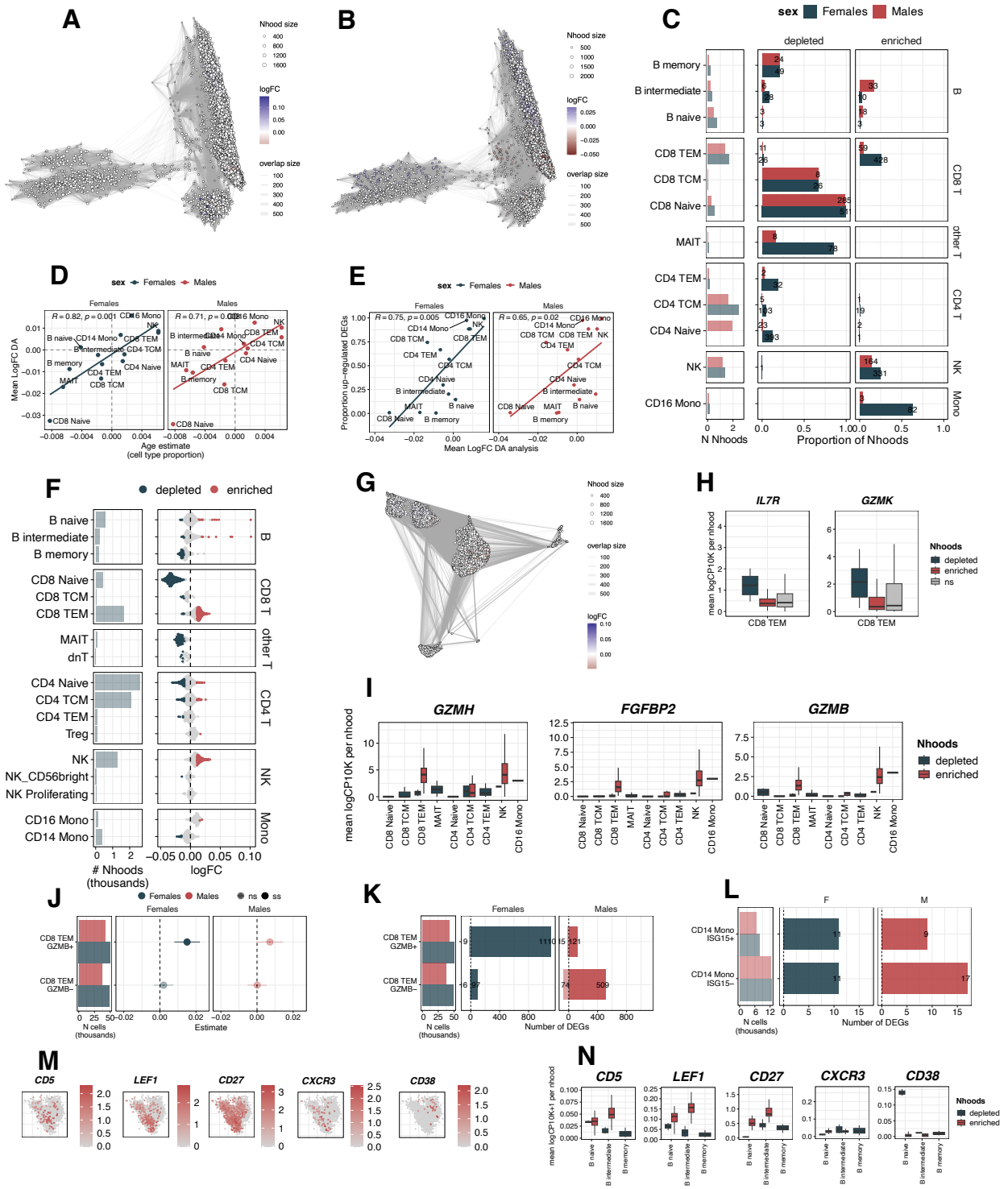

Figure S5. Sex-stratified differential abundance analysis Neighborhood graph of the results from differential abundance testing with age. Nodes are neighborhoods colored when they are significant (Spatial-FDR < 0.05), and edges correspond to the number of cells shared between neighborhoods. Colored neighborhoods colored by their log fold change across ages. Nhoods = neighborhoods for **a.** males **b.** females **c.** Left. Number of neighborhoods across cell types. Right. Percentage of age-enriched and age-depleted neighborhoods. Labels indicate the numbers **d.** Correlation of mean logFC per cell-type across neighborhoods (y-axis) versus the age-compositional estimate of cell type proportions (x-axis). **e.** Correlation between proportion of up-regulated genes in each cell type (y-axis) versus mean logFC per cell-type across neighborhoods (x-axis). **f.** Left. Number of neighborhoods per cell type. Right. Beeswarm plot of the distribution of neighborhood logFC with age computed with both sexes together. Each dot corresponds to a neighborhood of cells assigned to a cell type (> 80% of cells belonging to that cell type per neighborhood (see Methods)). Age-DA neighborhoods (Spatial-FDR < 0.05) are colored. **g.** Neighborhood graph of the

results from differential abundance testing with age. Nodes are neighborhoods colored when they are significant (Spatial-FDR < 0.05), and edges correspond to the number of cells shared between neighborhoods. Colored neighborhoods colored by their log fold change across ages. Nhoods = neighborhoods **h.** Average log-normalized and scaled expression per neighborhood of top down-regulated genes of age-enriched *GZMB* + CD8+ TEM subpopulation **i.** Bottom**.** Average log-normalized and scaled expression per neighborhood of *GZMH, FGFBP2, GZMB* genes across all cells. Colors refer to significant neighborhoods that increase (enriched) and decrease (depleted) with age **j.** Left. Total number of cells analyzed per CD8+ TEM subpopulation and sex. Center. Cell type composition analysis estimates with age from CD8+ TEM subpopulations in females Right. Cell type composition analysis estimates with age from CD8+ TEM subpopulations in males **k.**Left. Total number of cells analyzed per CD8+ TEM subpopulation and sex. Center. Number of age-DEGs per CD8+ TEM subpopulation in females. Right. Number of age-DEGs per CD8+ TEM subpopulation in males. **l.** Left. Total number of cells analyzed per CD14+ monocyte subpopulation and sex. Center. Number of age-DEGs per CD14+ monocyte subpopulation in females. Right. Number of age-DEGs per CD14+ monocyte subpopulation in males. **m.** UMAP of male-age-enriched B naive and intermediate cells colored by the expression of the top marker genes. **n.** Average log-normalized and scaled expression per neighborhood in males of the same marker genes as in l.

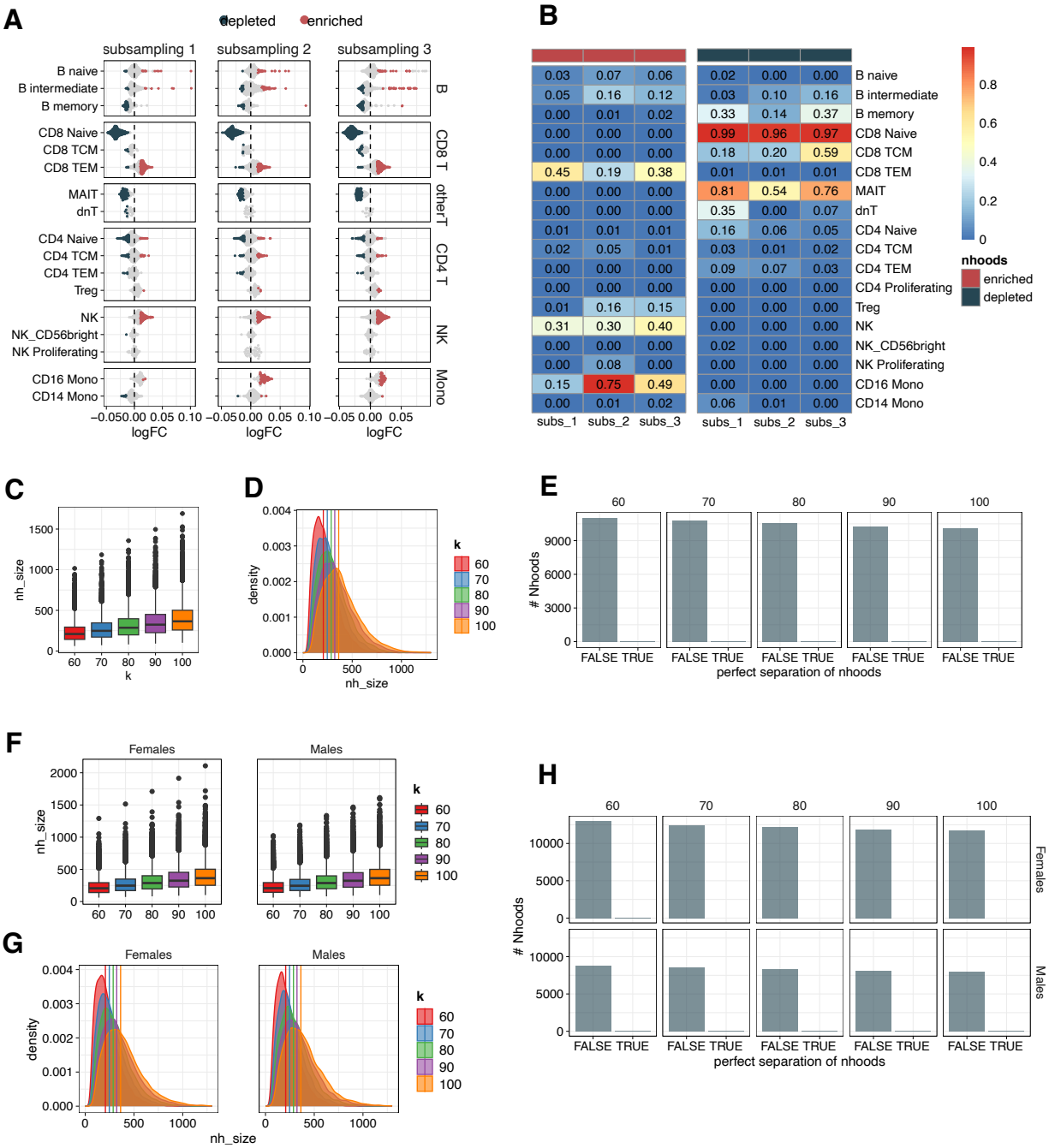

Figure S6. Differential abundance analysis: replication using data subsamplings and fine-tuning the k parameter for the KNN graph a. Beeswarm plot of the distribution of neighborhood logFC with age for each data subsampling (25% of the total cells; 316,939 cells out of a total of 1,272,489 cells). Each dot corresponds to a neighborhood of cells assigned to a cell type (> 80% of cells belonging to that cell type per neighborhood (see Methods)). Age-DA neighborhoods (Spatial-FDR < 0.05) are colored. **b.** Percentage of age-DA enriched (left) and depleted (right) neighborhoods in each data subsampling. **c.** and **d.** To identify the proper value of nearest-neighbors to consider for the graph building (*k*), we set different *k* values (60, 70, 80, 90 and, 100), and compared the **c.** distribution of the number of neighborhoods for each neighborhood size and **d.** distribution of the neighborhood size across different values of *k*. Vertical lines in **c** correspond to the mean neighborhood size. **e.** Number of neighborhoods that have perfect data separation based on age categories (TRUE) versus the ones that do not (FALSE) at different values of *k*. Age categories have been defined as: young (=<40 years), middle-aged (41–59 years), and old (>=60 years). **f-h.** Same plots as **c,d,e** but stratifying the data per sex and performing the analysis within each subset of the data.