Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Sex-specific trajectories of nonlinear immune aging at single cell level

Jacques Behmoaras

jacquesb@nus.edu.sg

Duke NUS Singapore https://orcid.org/0000-0002-5170-2606

Harry Park

Duke NUS Singapore

Nina Le Bert

Duke NUS Singapore

Antonio Bertoletti

DUKE-NUS Medical School https://orcid.org/0000-0002-2942-0485

Nicholas Tolwinski

Duke-NUS Medical School Singapore https://orcid.org/0000-0002-8507-2737

Jan Gruber

NUS

Article

Keywords:

Posted Date: March 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8944546/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Sex-specific trajectories of nonlinear immune aging at single cell level

Harry Park[1,2], Nina Le Bert[3], Antonio Bertoletti[3], Nicholas Tolwinski[4], Jan Gruber[5,6,7] and Jacques Behmoaras[1,2]

[1]Centre for Biomedical Data Science, Duke-NUS Medical School, Singapore, Singapore
[2]Program in Cardiovascular and Metabolic Disorders, Duke-NUS Medical School, Singapore
[3]Program in Emerging Infectious Disease, Duke-NUS Medical School, Singapore, Singapore
[4]Program in Cancer and Stem Cell Biology, Duke-NUS Medical School, Singapore, Singapore
[5]Healthy Longevity Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[6]Center for Healthy Longevity, National University Health System, Singapore, Singapore
[7]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

Correspondence: Jacquesb@duke-nus.edu.sg

Abstract

Despite the female predominance in age- and immune-associated diseases, the pathways underlying sex difference in healthy immune aging remain incompletely understood. Here, we present a multi-ethnic single-cell transcriptomic atlas of healthy human immune aging (35% Asian), comprising 3.8 million peripheral blood mononuclear cells (PBMCs) from 1,828 individuals aged 19–97 years. Prominent peaks of differential gene expression around 40 (mainly CD4 T cells) and after 60 (mainly CD8 T cells) years of age were accompanied by age-dependent decline in RNA/protein homeostasis and inflammatory PBMC polarization with sex-dependent kinetics. While females displayed sustained CD8 T cell immunometabolism and late-life senescence signatures in CD4 T, NK, and B cells, males exhibited early-life fluctuations in CD4 T cell immunometabolism associated with hypomethylation of SSH3 at chromosome 11q13. Deep learning-based biological age clocks stratified for sex outperformed sex-combined models, learning from transcriptional immune trajectories. We thus unravel a nonlinear PBMC aging whereby targeting sex-specific pathways can allow precision geromedicine.

Introduction

Recent advances in systems biology have revealed that aging does not progress linearly. Multi-omics profiling of human blood has uncovered distinct temporal waves of molecular changes during the human lifespan[1–3], suggesting critical transition periods that may represent windows of heightened vulnerability or therapeutic opportunity. While recent single-cell RNA sequencing (scRNA-seq) studies have begun to catalog immune changes across the lifespan[4–6], most analyses aggregate data across age groups or model aging as a linear process, potentially missing critical temporal dynamics and individual-level heterogeneity.

Emerging evidence indicates that biological aging exhibits substantial sexual dimorphism[7–9]. Females live longer than males yet experience greater burden from age-related diseases. Females also mount stronger innate and adaptive immune responses, contributing to enhanced pathogen clearance but also increased susceptibility to autoimmune diseases[10–12]. Despite these well-established immunological sex differences, how immune aging trajectories diverge between males and females at the transcriptional level—and whether these differences accelerate or diminish over the lifespan—remains largely unexplored. Recent work suggests sex-specific epigenetic aging patterns[3] and differential immune responses in older adults[13,14], yet comprehensive single-cell analyses integrating large cohorts spanning both sexes and the full age range are lacking. Importantly, whether the nonlinear aging dynamics manifest at single-cell resolution in peripheral immune cells, and the extent to which sex contributes to these processes remain unknown.

Here, we present a multi-ethnic single-cell transcriptomic atlas of peripheral blood immune aging, integrating 3.8 million cells from 1,828 healthy individuals aged 19–97 years with balanced sex representation. We reveal nonlinear aging dynamics with prominent peaks in the 40s and 60s, recapitulating and extending previous findings[1–3] to single cell resolution in the peripheral blood. Functional analyses highlight coordinated decline in protein and RNA homeostasis pathways across myeloid and lymphoid cells in both sexes, accompanied by a sex-dependent immune cell activation. Sex-stratification uncovers distinct temporal patterns: female immune cells display sustained CD8 T immunometabolism and late-life senescence signatures involving CD4 T, NK, and B cells, while males exhibit early, transient changes in CD4 T cell immunometabolism. Parallel DNA methylation analysis implicates contribution of epigenetic regulation in chromosome 11q13 and the MHC locus (6p21) in these sex-dependent immune trajectories. Finally, we develop deep learning-based scRNA-seq biological age clocks to demonstrate that sex-stratified models efficiently learn from specific immune kinetics and outperform sex-combined models.

Results

Aging human PBMCs show nonlinear transcriptional dynamics

We integrated and analyzed four large-scale scRNA-seq datasets of peripheral blood mononuclear cells (PBMCs) from healthy donors with age and sex information[5,15–17] (Fig. 1a-c). After quality control (see Methods), we obtained approximately 3.8 million immune cells from 1,021 females and 807 males aged 19-97 years, representing 12 self-reported ethnicities (34.8% Asian) (Fig. 1a-b, Extended Data Fig. 1a). Cell type proportion analyses revealed canonical aging features including CD8 naïve T cell loss and MAIT cell decline in both sexes, aligning with established immunological aging features[3–6,18–20] (Table 1).

Leveraging this integrated dataset, we explored comprehensive aging patterns across immune cell types in males and females. To investigate whether nonlinear aging dynamics previously observed in plasma transcriptomics and proteomics[1,2] are conserved at single-cell resolution, we applied Differential Expression-Sliding Window Analysis (DE-SWAN)[2] to pseudobulk expression profiles of five major PBMC cell types: CD4 T cells, CD8 T cells, B cells, NK cells, and monocytes (see Methods).

We performed parallel analyses on female-only, male-only, and combined-sex cohorts to define sex-aware aging features (SAFs)—cell type–gene pairs exhibiting significant differential expression ($q < 0.05$, window size 20) across age windows. Among 7,014 features tested, we identified 6,322 both-sex, 5,204 female, and 4,346 male SAFs. While 55% overlapped across all three sets, 23% were exclusive to female and both-sex analyses (Fig. 1d), indicating more pronounced female-specific transcriptional changes during aging.

To quantify global changes regardless of cell types, we visualized these features across the age range and confirmed two distinct "aging peaks": a primary peak of differential expression occurring at ~40 years and a secondary peak after 60 years (Fig. 1e). These peaks remained after applying different sliding-window sizes and statistical thresholds (Extended Data Fig. 2a–b) but were no longer detected upon random age permutation (Extended Data Fig. 2c), confirming true biological signal. Individual datasets captured either the first or second peak, but not both, reflecting their limited sample size, age range, and biased sex distribution (Extended Data Fig. 1b-d). These results underscore the power of data integration for revealing aging trajectories otherwise not fully detected.

Both peaks manifested within individual cell types, but with cell type- and sex-specific patterns (Fig. 1f, Extended Data Fig. 2d). Females consistently exhibited higher peak amplitudes than males, concordant with their SAF numbers. CD4 T cells harbored the most SAFs (n = 4,137, 3,321, 2,901 for both-sex, female, and male, respectively), followed by CD8 T cells (n = 1,203, 1,077, 1,019), while monocytes showed the fewest (n = 201, 182, 67). CD4 T cells dominated the first peak, while CD8 T cells were prominent in the second one. B cells, male CD8 T cells, and female NK cells exhibited balanced bimodal peaks while female CD8 T cells exclusively showed merged pattern of the two peaks. Female monocytes were characterized by the first peak only, whereas NK cells and monocytes in males did not exhibit clear peaks.

Sex-dependent transcriptional trajectories in immune aging

To further dissect sex differences in aging transcriptome, we performed age trajectory clustering on each SAF set based on pseudobulk temporal kinetics, identifying seven distinct patterns (Fig. 2a, Extended Data Fig. 3b, Supplementary Figure 1, see Methods). Three were

conserved across sexes (early increase, early decrease, and continuous decrease), while four were sex-specific: early fluctuation (male-specific), inverted U-shape (female-specific), continuous increase (female-specific), and late increase (female-specific).

The majority of SAFs exhibited early-changing dynamics. Early increase and early decrease clusters comprised 49.1% of both-sex, 50.1% of female, and 71.8% of male SAFs (Fig. 2a), consistent with the prominent first DE-SWAN peak (Fig. 1e-f). At the cell type level, many populations showed >50% early-changing features, including CD4 T cells, male CD8 T cells, and male monocytes (Extended Data Fig. 3a). T cells, especially CD4 T cells, dominated early-changing clusters—representing 76.1%, 73.4%, and 66.6% of both-sex, female, and male SAFs, respectively (Fig. 2a)—explaining the pronounced first peak observed in CD4 T cells (Fig. 1f). These observations indicate that CD4 T cells exhibit the broadest age-related transcriptional changes, particularly in early-to-mid adulthood.

Continuous-changing clusters represented the next most prominent immune aging trajectory, especially in both-sex and female SAFs. 33.0% and 29.9% of both-sex and female SAFs, respectively, exhibited continuous changes (increases and decreases combined), while 10.7% of male SAFs showed continuous decrease (Fig. 2a). CD8 T cells were overrepresented in the continuous increase clusters in both-sex and female SAFs (38.8% and 51.0%, respectively) (Fig. 2a, Extended Data Fig. 3a). This pattern likely reflects the merged DE-SWAN peak observed in female CD8 T cells (Fig. 1f).

The male-specific early fluctuation cluster predominantly comprised CD4 T cell features (82.0% and 80.1% in both-sex and male SAFs, respectively) (Fig. 2a). Similarly, the female-specific inverted U-shape cluster was strongly enriched for CD4 T cell features (86.7%), whereas the late increase cluster showed more balanced representation across the five major cell types rather than T cell dominance (Fig. 2a). This distribution potentially underlies the higher second peaks observed across cell types in females compared to males (Fig. 1e-f).

Aging PBMCs decline in mRNA/protein homeostasis regardless of sex

To functionally annotate age-associated changes across immune cell types, we employed two complementary approaches: (1) gene set enrichment analysis (GSEA) based on Spearman correlations of all female and male SAFs, and (2) cluster-wise pathway enrichment using Gene Ontology Biological Process (GOBP) terms (see Methods). The first approach identified the strongest upregulated and downregulated processes through correlation-based GSEA ranking, while the second enabled detailed characterization of each trajectory cluster.

GSEA revealed robust and consistent downregulation of translation and RNA metabolism pathways across sexes and cell types (Fig. 2b-c; $q < 0.01$, $NES < 0$), with the exception of male NK cells. Downregulated pathways included rRNA processing, translation elongation, nonsense-mediated mRNA decay, and mRNA processing. This coordinated decline in transcription-translation machinery aligns with recently reported proteostasis failure characterized by mRNA-protein decoupling across multiple aging human tissues[21].

Cluster-wise pathway enrichment characterized the two main decrease clusters (early decrease and continuous decrease) with pervasive enrichment in DNA/RNA metabolism and translation/proteostasis modules across sexes and cell types, reinforcing the coordinated decline in RNA and protein homeostasis (Extended Data Fig. 4a). These clusters were also

enriched for OXPHOS/energy metabolism modules, indicating a generalized PBMC energetics decline.

In contrast, upregulated pathways were more cell type-specific and enriched for established age-associated signatures (Fig. 2b-c; $q < 0.01$, NES > 0). Adaptive immune system, RHO GTPase signaling, and platelet activation pathways were upregulated especially in CD8 T cells in both sexes, consistent with prior aging tissue proteomics[21] and blood transcriptomics[1]. Interleukin-4 and interleukin-13 signaling pathway was increased in male CD4 T cells.

Cluster-wise analysis similarly revealed that two main increase clusters (early increase in males and continuous increase in females) were enriched for immune activation modules (Extended Data Fig. 4a). These included immune response (NF-kB, interferon), immune cell differentiation/activation, and antigen-specific activation, especially in CD8 T cells. Monocytes and CD4 / CD8 T cells in these clusters showed enrichment for actin fiber organization-potentially reflecting cytoskeletal remodeling during immune activation[22]-with upregulation of WAS, WASF2, and ARPC2. CD4 / CD8 T cells also exhibited elevated apoptotic signaling pathways (BAX, CASP8).

Functional annotations reveal coordinated decline in RNA-protein homeostasis across cell types, while increasing trajectories reflect progressive immune cell activation and inflammatory polarization ("inflammaging") from early-to-mid life. This pattern suggests that reduced biosynthetic capacity coincides with heightened cellular stress and immune activation, consistent with transcriptome-proteome decoupling across aging tissues[21].

Sustained CD8 T cell immunometabolism and late-life senescence in females

We next focused on sex-specific clusters of immune aging trajectories. The continuous increase cluster was unique to female SAFs. When mapped to male clustering profiles, these features were distributed across multiple trajectories, with 175 female-only SAFs not detected as significantly changing with age in males (Fig. 3a).

Majority of the female continuous increase cluster (a subcluster of 531 features) mapped to the early increase cluster in males (Fisher's exact test, $p < 2.2 \times 10^{-16}$), indicating sex-specific kinetics (Fig. 3b-c). While increasing with age in both sexes, overall Spearman correlation of this subcluster across the full age range was significantly higher in females than in males (Fig. 3e). Intriguingly, correlation of these features with age in male donors <60 years was significantly higher than in male donors >60 years, whereas the difference was not significant in females (Fig. 3f-g). These results indicate a more continuous increasing trend in females while males display a pronounced early rise that plateaus later in life—a trajectory pattern reflected in locally estimated scatterplot smoothing (LOESS) plots (Fig. 3b-c). This sex-divergent kinetics was also reflected in features with the top positive correlation with age, where they were enriched in continuous increase clusters in females but in early increase clusters in males (Extended Data Fig. 3c).

This subcluster was dominated by CD8 T cell features (62.3%; Fisher's exact test, $p < 2.2 \times 10^{-16}$; Fig. 3d) related to immune and metabolic activation (i.e. immunometabolism). Pathway enrichment analysis ($q < 0.01$, GOBP) characterized these features with immune activation, RNA/protein metabolism, and mitochondrial ATP synthesis (Fig. 3h, Extended Data Fig. 5a). Enrichment of interferon/interleukin signaling and T/NK cell-mediated cytotoxicity pathways further linked these features to age-associated immune activation and inflammation (Extended

Data Fig. 5a). Representative genes include SIGIRR, CD74, MYO1G, and ITGB2 related to effector / cytotoxic T cell function[23–26], and NDUFA2 and IDH3B related to energy metabolism[27] in CD8 T cells, showing sustained late-life increases in females compared to males (Extended Data Fig. 5b).

The female-specific late increase cluster distributed across male SAF trajectories, with more than half not present in male SAFs (Fig. 3i; n = 137, 53.3%; Fisher's exact test, $p < 2.2 \times 10^{-16}$). This cluster (all features, n = 257) showed wide cell type representation with CD4 T, NK and B cells features (Fig. 2a). Top pathways included a broad set of translation-related ones, accompanied by p53 and apoptotic signaling pathways in NK and CD4 T cells (Fig. 3j, Extended Data Fig. 5c). Key contributing genes included TPT1, a regulator of cellular growth and proliferation, and RACK1, a scaffold protein interacting with cyclins, in both CD4 T and NK cells (Extended Data Fig. 5d). LAPTM5, implicated in B cell apoptosis[28] and suppression of excessive T cell activation[29], also showed a female-specific late-life increase in CD4 T cells (Extended Data Fig. 5d). Notably, FTL and FTH1, encoding the ferritin light and heavy chains that store intracellular iron, exhibited a late-increase pattern in NK, CD4 T, CD8 T, and B cells in females - a trend not observed in males (Fig. 3k, Extended Data Fig. 5e). Consistent with this, the iron accumulation signature (IAS)[30] was significantly overrepresented among features in the late increase cluster (Fisher's exact test, $p = 4.874 \times 10^{-6}$), including FTL, TPT1, and EEF1A1. This kinetic pattern of ferritin and IAS-related transcripts aligns with the association between menopause and increased systemic iron burden. Menopause has been shown to substantially alter iron homeostasis in women[31], and serum ferritin levels in females from the NHANES dataset exhibit a significant increase around the menopausal age (~50) (Extended Data Fig. 5f). Notably, excess iron can promote cellular senescence[30]. Taken together, these findings suggest that the female-specific late-increase cluster reflects a senescence-associated cellular state.

The female-specific inverted U-shape cluster was mainly distributed across two male SAF trajectories (early increase and early fluctuation clusters) with more than a third not selected as male SAFs (Fig. 3l; n = 270; 34.6%; Fisher's exact test, $p = 1.78 \times 10^{-7}$). This cluster was dominated by CD4 T cell features (86.7%; Fisher's exact test, $p < 2.2 \times 10^{-16}$; Fig. 2a) enriched in intracellular trafficking pathways including vesicle organization and ER organization (Fig. 3m, Extended Data Fig. 5g). Representative genes were ATP5AP2 and CLN5 located in lysosomes, and TRAM1 and AP1S2 found in ER and Golgi apparatus, respectively (Extended Data Fig. 5h). ANP32E, a suppressor of genome wide chromatin accessibility[32], and ARRDC3, a gene previously reported as a sex-sensitive regulator of insulin receptor signaling[33], also exhibited this kinetics (Extended Data Fig. 5h). MHC-II genes (HLA-DM and HLA-DR) in B cells were also included in this cluster, exhibiting a female-specific late-life decline (Fig. 3n). Interestingly, B cell MHC-II expression has recently been mechanistically linked to CD4 T cell senescence[34].

In summary, female-specific trajectories are characterized by a sustained, life-long increase in CD8 T cell activation and immunometabolism and late-life senescence in CD4 T, NK, and B cells.

DNA methylation and sex-dependent PBMC aging trajectories

The male-specific early fluctuation cluster distributed across six female SAF trajectories, mapping predominantly to the early increase and inverted U-shape clusters, with 187 features absent from female SAFs (24.7%; Fisher's exact test, $p < 2.2 \times 10^{-16}$; Fig. 4a). This cluster

was dominated by CD4 T cell features (80.1%; Fisher's exact test, $p < 2.2 \times 10^{-16}$; Fig. 2a). Pathway enrichment analysis revealed overrepresentation of OXPHOS, mitochondrial gene expression, and actin/vesicle organization pathways (Extended Data Fig. 6a-b, Extended Data Fig. 4).

This cluster was distinguished by a uniquely significant genomic position enrichment signal (MSigDB C1; Fig. 4b). Among all female and male SAF clusters, only this cluster exhibited a highly significant enrichment, driven specifically by CD4 T cell features mapping to chromosome 11q13 (21 overlapping features; $q = 4.50 \times 10^{-4}$; Fig. 4b, Supplementary Table 1). The overlapping genes included key regulators such as DPF2, implicated in myeloid and lymphoid activation[35–37], and OTUB1, involved in immune signaling[38–40], suggesting a regulatory architecture capable of driving the early fluctuation dynamics of this cluster.

Given the exclusivity and strength of this positional enrichment, we tested whether this locus is subject to age- and sex-dependent epigenetic regulation. We analyzed DNA methylation data from whole blood samples isolated from 729 healthy donors[41] (GSE87571; Fig. 4c, Extended Data Fig. 6c). Samples were stratified into young (<40), middle-aged (40–60), and old (>60) groups, and differentially methylated region (DMR) analysis was performed to test for age-sex interactions (see Methods, Fig. 4c).

This analysis identified 49 and 302 DMRs with significant age-sex interactions in the young vs. middle-aged and young vs. old comparisons, respectively (Fig. 4g-h). While most mapped to sex chromosomes, 2 and 10 autosomal DMRs were identified in the respective comparisons (Supplementary Table 2). Notably, a DMR at chromosome 11q13 was detected in the young vs. old comparison (Fig. 4d-f), spanning three CpG sites in the gene body and 5' UTR of SSH3, a regulator of actin dynamics and cell motility. This locus exhibited significant age-associated hypomethylation in males but not in females (Fig. 4e-f), paralleling the male-specific transcriptional enrichment at this genomic position.

Strikingly, four out of 12 autosomal DMRs localized to chromosome 6p21 (Fig. 4g-h). This region encompasses the major histocompatibility complex (MHC) locus and has previously been implicated in age-associated hypomethylation in naïve CD4 T cells[42], accelerated biological aging and mortality in chronic HIV infection[43], and strong gene-sex interactions in systemic lupus erythematosus (SLE)—an autoimmune disease with female predominance[44]. These convergent observations prompted further investigation of sex-specific epigenetic remodelling during aging at this locus.

The 6p21 DMR identified in the young vs. middle-aged comparison spanned 27 CpG sites across the gene body of MPIG6B (c6orf25), a critical regulator of platelet activation and megakaryocyte function (Extended Data Fig. 6d). This region showed age-associated hypomethylation in both sexes but with delayed kinetics - occurring between middle-aged and old groups in females, versus young and middle-aged groups in males (Extended Data Fig. 6e-f). Similarly, three 6p21 DMRs in the young vs. old comparison demonstrated weaker and/or delayed hypomethylation in females. These regions spanned: (1) seven CpGs across the gene body, 5' UTR, and transcription start site (TSS) of NOTCH4, a key regulator of T cell activation, especially regulatory T cells (Tregs)[45,46] (Fig. 4i-m); (2) nine CpGs in the gene body of DAXX, a histone chaperone and transcriptional repressor (Extended Data Fig. 6g-h); and (3) four CpGs in the TSS1500 region of DDAH2 and the gene body of CLIC1 (Fig. 4g-h).

In our integrated scRNA-seq PBMC aging atlas, CLIC1 expression in CD4 and CD8 T cells exhibited a continuous increase in females and early increase in males (Fig. 4j), mirroring the sex-dependent shift in methylation kinetics. The female continuous–male early increase transcriptional trajectories, which include CLIC1 and are enriched in features related to CD8 T cell activation and differentiation (Fig. 3a–h), align with epigenetic timing differences at the MHC locus encompassing key antigen-presentation and immune response-related genes. This alignment suggests that sex-specific DNA methylation dynamics may be linked to the differential kinetics in CD8 T cell-mediated immune signature observed in male and female transcriptome.

Together, these data demonstrate that sex-differential DNA methylation remodeling - particularly at the 11q13 and 6p21 loci - parallels and likely contributes to sex-divergent transcriptional aging trajectories. This provides a mechanistic epigenetic basis for sex-specific immune aging dynamics and implicates DNA methylation as a driver of divergent temporal aging trajectories at the transcriptional level.

Sex-stratified biological age clocks learn from SAF trajectories

Based on the sex differences observed in aging PBMC transcriptional trends, we next tested whether sex stratification improves biological age prediction and enables attribution of trajectory-specific features. We constructed three models: a sex-combined clock trained on both-sex SAFs and two sex-specific clocks trained separately in females and males (Fig. 5a; see Methods). Each model was trained to predict chronological age from SAF expression profiles. The final input matrices comprised 1,828 donors × 6,322 features (sex-combined), 1,021 donors × 5,204 features (female-specific), and 807 donors × 4,350 features (male-specific), respectively.

We first compared elastic net regression (ENET), eXtreme Gradient Boosting (XGBoost), and multilayer perceptron (MLP) approaches using five-fold cross-validation across multiple performance metrics and independent external datasets. Internal validation showed that ENET and MLP significantly outperformed XGBoost, yielding lower root mean squared error (RMSE) and higher Pearson correlations (Extended Data Fig. 7a-b). External validation further demonstrated MLP superiority over ENET, with improved RMSE and correlation, indicating reduced overfitting and enhanced generalizability (Extended Data Fig. 7c-d). Based on these criteria, MLP models were selected for downstream analyses.

Both internal and external validations showed performance gains for sex-specific clocks relative to the sex-combined model (Extended Data Fig. 7e–h). Cross-testing (evaluating male clocks on female data and vice versa) produced significant and consistent performance drops across validation datasets, confirming that sex-specific models capture distinct, sex-dependent aging features (Extended Data Fig. 7e–h). Notably, external validations exhibited larger performance improvements for sex-specific models than internal validations, with significant Pearson correlation improvement for the female-specific clock in female donors and significant RMSE reductions for the male-specific clock in male donors (Extended Data Fig. 7g–h). This pattern suggests that internal datasets may approach a performance ceiling due to shared cohort structure and technical homogeneity, whereas sex-specific clocks show superior generalizability when applied to novel and heterogeneous external populations, highlighting their translational robustness.

From each five-fold MLP set, we selected the best-performing models based on external validation metrics (RMSE, MAE, Pearson *r*) (Table 2). Final models exhibited strong correlations between chronological and predicted age (>0.85 in internal validation; >0.7 in healthy external cohorts; >0.6 disease external cohorts) (Fig. 5b-c, Extended Fig. 8a-b). Overall, sex-stratified models consistently outperformed the sex-combined model across datasets—most prominently the male-specific model in internal validation, the female-specific model in healthy external cohorts, and both sex-specific models in disease and pooled external validations—supporting their robustness. Notably, performance was lower in disease cohorts than in healthy cohorts (Extended Data Fig. 7i), likely reflecting disease-associated transcriptomic changes, which may partially decouple chronological aging signals and introduce additional variance beyond physiological aging trajectories.

We next analyzed SHAP (Shapley Additive exPlanations) values to assess feature contributions. Principal component analysis (PCA) of SHAP values revealed high dataset dependence, indicating MLP models dynamically adjusted feature weighting across cohorts (Supplementary Figure 2). This adaptive reweighting is not supported by linear models such as ENET and likely contributes to the superior external validation performance of MLP models. To isolate age-related biological contributions from dataset effects, SHAP values were corrected by base mean (Supplementary Figure 2, see Methods).

Using corrected SHAP values, we quantified the net contribution of each trajectory cluster to age prediction by computing summed SHAP scores per cluster. The continuous increase cluster showed the highest summed absolute SHAP values in the female model, while the early increase cluster dominated in the male model (Extended Data Fig. 8c-d), mirroring the top age-correlated features in the respective SAFs (Extended Data Fig. 3c). This indicates that the models preferentially rely on biologically dominant, sex-specific aging programs when predicting age.

Summed SHAP values further revealed kinetically aligned contributions of trajectory-specific features (Fig. 5d-e). Upon stratification of donors into young (<40), middle-aged (40-60), and old (>60) groups, early-changing trajectories showed the largest SHAP differences between young and middle-aged groups, whereas continuous-changing trajectories exhibited progressive SHAP increases across age groups, with the strongest effects between middle-aged and old donors in both sexes. The female-specific late increase cluster also showed a pronounced SHAP shift between middle-aged and old females, demonstrating that model attribution faithfully recapitulates the temporal structure of the underlying trajectories.

Collectively, these analyses demonstrate that sex-specific clocks learn from sex-specific temporal aging dynamics that likely underlie their improved performance, underscoring their relevance for precision geromedicine.

Discussion

Our atlas reveals two prominent peaks of differential gene expression in peripheral immune cells—at approximately 40s and 60s—extending observations made in aging human blood[1–3] to transcriptionally defined PBMCs at single cell level. The nonlinear aging peaks vary by immune cell type, with CD4 T cells exhibiting the strongest early peak (40s) across sexes, consistent with their role as central coordinators of adaptive immunity and high sensitivity to systemic aging cues. Consistent with previous studies[4–6], CD8 T cells show the greatest transcriptional magnitude across lifespan, and are the dominant cells underlying the second peak after 60s. These temporal hubs align with known physiological transitions during aging: the fourth decade coincides with declining reproductive capacity and early hormonal changes, while the sixth decade encompasses perimenopause/menopause in females and continued androgen decline in males[47,48]. The nonlinear nature of these changes suggests that immune aging is not a gradual, monotonic process but rather proceeds through punctuated transitions that may represent critical windows for rejuvenation-aimed interventions. Importantly, individual aging PBMC datasets captured only one peak or the other, highlighting the importance of integrative, multi-ethnic, wide age-range sampling and large cohort sizes, for unravelling subtle immune aging dynamics and their underlying pathways.

Age-associated immune dysfunction contributes to increased susceptibility to infection, reduced vaccine efficacy, and elevated risk of autoimmune and neurodegenerative diseases[49], making it a major pillar of aging biology[50,51]. The immune system undergoes profound remodeling with age - a process termed immunosenescence - characterized by declining naïve T cell pools, accumulation of memory and exhausted T cells, and chronic low-grade inflammation ("inflammaging")[52–55]. Our atlas reports proportional changes in PBMC types in line with these immunosenescence features: a marked decline in CD8 naïve T cells[3,5,6,18] accompanied by an age-associated increase in the CD4/CD8 ratio[4,5,19] and increased NK cell proportions[56,57]. Greater age-associated expansion of $CD8^+$ central memory T cells in males and broadly similar age-related changes in innate immune cell populations between sexes are consistent with previous age-by-sex interaction studies[3,20].

One of the most striking findings is the widespread downregulation of transcription and translation-related pathways with age across PBMC types and sexes. Genes encoding ribosomal proteins, translation initiation factors, and RNA processing machinery decline progressively, particularly in early decrease and continuous decrease trajectory clusters. The biological consequences of this decline are likely to be profound. Reduced protein synthesis capacity impairs cellular stress responses, limits proliferative potential, and may promote accumulation of misfolded proteins[58,59], consistent with increased ER stress and unfolded protein response genes (XBP1, ATF6, and EIF2S1) in early increase clusters. This dual trend of declining translational capacity alongside heightened proteotoxic stress responses defines the "failed proteostasis" characteristic of aging cells[40,41], which has increasingly been recognized as a central and evolutionarily conserved hallmark of aging[62,21,63,64].

This coordinated proteostasis decline across all major PBMC types suggests it may be a core aging mechanism, with cell-type-specific responses shaping subsequent trajectories. The cell type-specific immune activation and inflammation pathways (i.e. inflammaging) raise a chicken-and-egg question: whether proteostasis failure precedes abnormal immune activation, or vice versa. The widespread proteostasis decline suggests it may be a shared primary event with cell type-specific downstream consequences. Indeed, proteostasis defects promote

protein aggregates like amyloid depositions in multiple human organs, further triggering immunoglobulin and complement activation[21]. In aging T cells - where defects in proteostasis and increased iron pools have been reported[65] - proteotoxic stress has been mechanistically linked to T cell exhaustion that characterizes age-related pathologies like cancer[66]. Collectively, these findings highlight protein quality control pathways as potential therapeutic targets for ameliorating age-related immune dysfunction[67,68].

Another key finding is a female-specific continuous-increase trajectory, characterized by sustained upregulation of effector function and cytotoxicity (SIGIRR, CD74, GZMA, MYO1G), metabolism (OXPHOS) and inflammatory signaling pathways (interleukin, interferon). This sustained immune activation in aging females may reflect both adaptive and maladaptive processes. Maintained immune activation capacity could contribute to superior vaccine responses and pathogen clearance observed in older women[10]. However, chronic activation carries costs: persistent interferon and NF-κB signaling drive inflammaging[52], promote T cell exhaustion[69], and may increase susceptibility to age-related diseases and autoimmune conditions that disproportionately affect women[12].

The other female-specific trajectories, including the late increase and inverted U-shape patterns, also reflect complex interplay between adaptive and maladaptive polarization processes during aging. Menopause represents a major mid-life physiological transition accompanied by profound systemic, metabolic, and inflammatory changes, including changes in iron homeostasis[9,31]. The female late increase cluster is characterized by elevated ferritin transcripts (FTL and FTH1) alongside markers of senescence and apoptosis signatures, particularly in CD4 T cells and NK cells. The female-specific inverted U-shaped trajectory includes MHC class II genes in B cells, the ablation of which has recently been shown to prevent CD4 T cell senescence[34]. In this context, late-life decrease of these genes in B cells may indicate a potential compensatory immune adaptation to female-specific late-life senescence.

Our discovery that male-specific early fluctuation features localize preferentially to chromosome 11q13 prompted investigation of DNA methylation dynamics. Analysis of age-stratified whole-blood methylation revealed marked sex differences: age-associated hypomethylation occurred either specifically in males (SSH3, 11q13) or with different kinetics (usually greater in magnitude and earlier in males; chr6p21 encompassing the MHC) throughout lifespan. Notably, chromosome 11q13 has been previously characterized as a risk locus for autoimmune and allergic diseases containing an enhancer of $CD4^+$ regulatory T cells[70]. Also, SSH3 belongs to slingshot family of genes which includes SSH1, recently implicated in early TCR activation during CD4 T cell activation[71]. The enrichment of age–sex interaction DMRs at chromosome 6p21 also implicates the MHC locus as a central epigenetic hub in sexually dimorphic immune aging. Building on previous work showing the region's association with aging[42], infection[43,72], and autoimmune disease[44], our findings suggest its contribution to female aging and/or age-related chronic disease.

Our aging clock results should be viewed in the wider context of the aging clock literature. Biological aging clocks have transformed geroscience over the past decade. DNA methylation–based clocks pioneered the field, beginning from Horvath's epigenetic clocks[73], and established highly accurate age prediction. This breakthrough catalyzed the development of clocks across multiple modalities and second-generation clocks that move beyond chronological age to predict all-cause mortality and age-related phenotypes. These include

estimators of mortality risk (GrimAge)[74,75] and intrinsic capacity[76] from DNA methylation data, inflammatory aging clock from blood immunome (iAge)[77], principal component-based aging clock from clinical data (PCAge)[78], and organ-specific aging clocks from plasma proteomics[79]. Recently, high-resolution aging clocks based on single-cell transcriptomics have emerged, enabling cell type-resolved dissection of molecular aging programs[18,80,81]. Within this framework, our deep learning–based scRNA-seq clocks not only achieve strong predictive performance ($R > 0.87$) but establish sex stratification as a biologically informative modeling principle. We show that sex-stratified models consistently outperform sex-combined models and learn from distinct temporal trajectories that are obscured in merged analyses, positioning sex-aware aging clocks as interpretable mechanistic models of aging kinetics. These findings highlight their translational relevance for precision geromedicine, that can be further exploited to other modalities and tissues to enable system-wide, sex-aware aging interventions.

Several limitations warrant consideration. First, experimental validation and replication by longitudinal studies tracking healthy individuals over decades will be pivotal in assessing how nonlinear dynamics of PBMC aging occur individually. Second, pseudobulk analyses of scRNA-seq data collapse cellular heterogeneity; future work employing single-cell trajectory inference may reveal additional granularity and pinpoint previously unappreciated cell state dynamics. An important question related to the latter is the subset of CD4 or CD8 T cells responsible for driving sex-dependent aging trajectories. Third, our focus on peripheral blood leaves open whether other tissues exhibit similar sex-specific aging patterns and whether blood represents the tissue with greatest sexual dimorphism in aging. Also, the sexual dimorphism not only affects transcripts but also proteins and metabolites[21,82,83]. Finally, key mechanistic questions remain unresolved. Sex differences in aging are likely to be driven by multiple interconnected factors. Our study implicates DNA methylation remodeling but does not directly address X chromosome dynamics[84,85], sex hormone effects[9,86], or sex-specific genetic regulatory architectures[87].

Our findings raise the possibility of sex-specific interventions to promote healthy immune aging. From a disease perspective, the results provide a valuable resource for studying female-specific pathways underlying susceptibility to autoimmune disease. As the field moves toward personalized geroscience, integrating single cell omics in healthy aging PBMCs with clinical data can unravel further fundamental mechanisms underlying how males and females age.

Tables

Table 1 Spearman correlation between age and cell type proportion with sex stratification

		Both sex		Female		Male	
Cell type	Type	*rho*	FDR	*rho*	FDR	*rho*	FDR
CD4 T	level 1	0.33	***	0.35	***	0.31	***
CD8 T	level 1	-0.18	***	-0.17	***	-0.20	***
B	level 1	0.076	**	0.067	*	0.094	**
NK	level 1	0.18	***	0.22	***	0.13	***
Mono	level 1	-0.42	***	-0.45	***	-0.38	***
DC	level 1	-0.27	***	-0.32	***	-0.21	***
other T	level 1	-0.40	***	-0.41	***	-0.39	***
other	level 1	-0.23	***	-0.19	***	-0.29	***
CD4 T / CD8 T	ratio	0.30	***	0.30	***	0.31	***
CD4 Naive	level 2	0.092	***	0.072	*	0.12	***
CD4 TCM	level 2	0.34	***	0.38	***	0.29	***
CD4 TEM	level 2	-0.061	**	-0.072	*	-0.042	ns
CD8 Naive	level 2	-0.73	***	-0.72	***	-0.75	***
CD8 TCM	level 2	0.24	***	0.20	***	0.29	***
CD8 TEM	level 2	0.18	***	0.22	***	0.14	***
Treg	level 2	0.060	*	0.069	*	0.051	ns
MAIT	level 2	-0.52	***	-0.52	***	-0.51	***
dnT	level 2	-0.12	***	-0.15	***	-0.081	*
gdT	level 2	-0.22	***	-0.19	***	-0.26	***
B Naive	level 2	0.10	***	0.086	**	0.13	***
B Intermediate	level 2	-0.15	***	-0.15	***	-0.16	***
B Memory	level 2	0.080	***	0.069	*	0.10	**
Plasmablast	level 2	0.18	***	0.16	***	0.22	***
NK	level 2	0.18	***	0.22	***	0.13	***
NK Proliferating	level 2	0.21	***	0.21	***	0.21	***
NK CD56bright	level 2	0.019	ns	0.0023	ns	0.047	ns
CD14 Mono	level 2	-0.46	***	-0.49	***	-0.41	***
CD16 Mono	level 2	-0.21	***	-0.22	***	-0.19	***
cDC1	level 2	-0.20	***	-0.25	***	-0.14	***
cDC2	level 2	-0.22	***	-0.29	***	-0.14	***
pDC	level 2	-0.25	***	-0.29	***	-0.20	***

*FDR < 0.05, ** FDR < 0.01, *** FDR < 0.001, ns not significant

Table 2 Performance metrics of MLP models on internal and external validation sets by fold

		Internal validation set			External validation set		
Model	Fold	RMSE	MAE	Pearson *r*	RMSE	MAE	Pearson *r*
Both Model	1	9.16	7.05	0.86	15.91	11.66	0.60
Both Model	2*	8.96	7.11	0.87	14.92	11.06	0.62
Both Model	3	8.82	6.79	0.88	16.14	12.01	0.64
Both Model	4	8.07	6.16	0.90	15.28	11.36	0.61
Both Model	5	8.37	6.33	0.90	15.89	11.96	0.65
Female Model	1	8.60	6.85	0.89	13.38	9.93	0.68
Female Model	2	9.54	7.29	0.87	14.45	10.95	0.68
Female Model	3	8.22	6.38	0.91	14.37	10.63	0.68
Female Model	4*	9.16	6.77	0.87	13.36	10.04	0.65
Female Model	5	7.69	5.80	0.92	14.53	10.36	0.66
Male Model	1	8.59	6.07	0.88	16.26	12.21	0.66
Male Model	2	9.39	6.93	0.88	15.71	12.05	0.66
Male Model	3*	8.91	6.85	0.88	14.94	11.36	0.68
Male Model	4	8.28	6.47	0.89	16.24	12.03	0.64
Male Model	5	8.23	6.02	0.91	15.62	11.78	0.64

*Selected as final models.

Figures

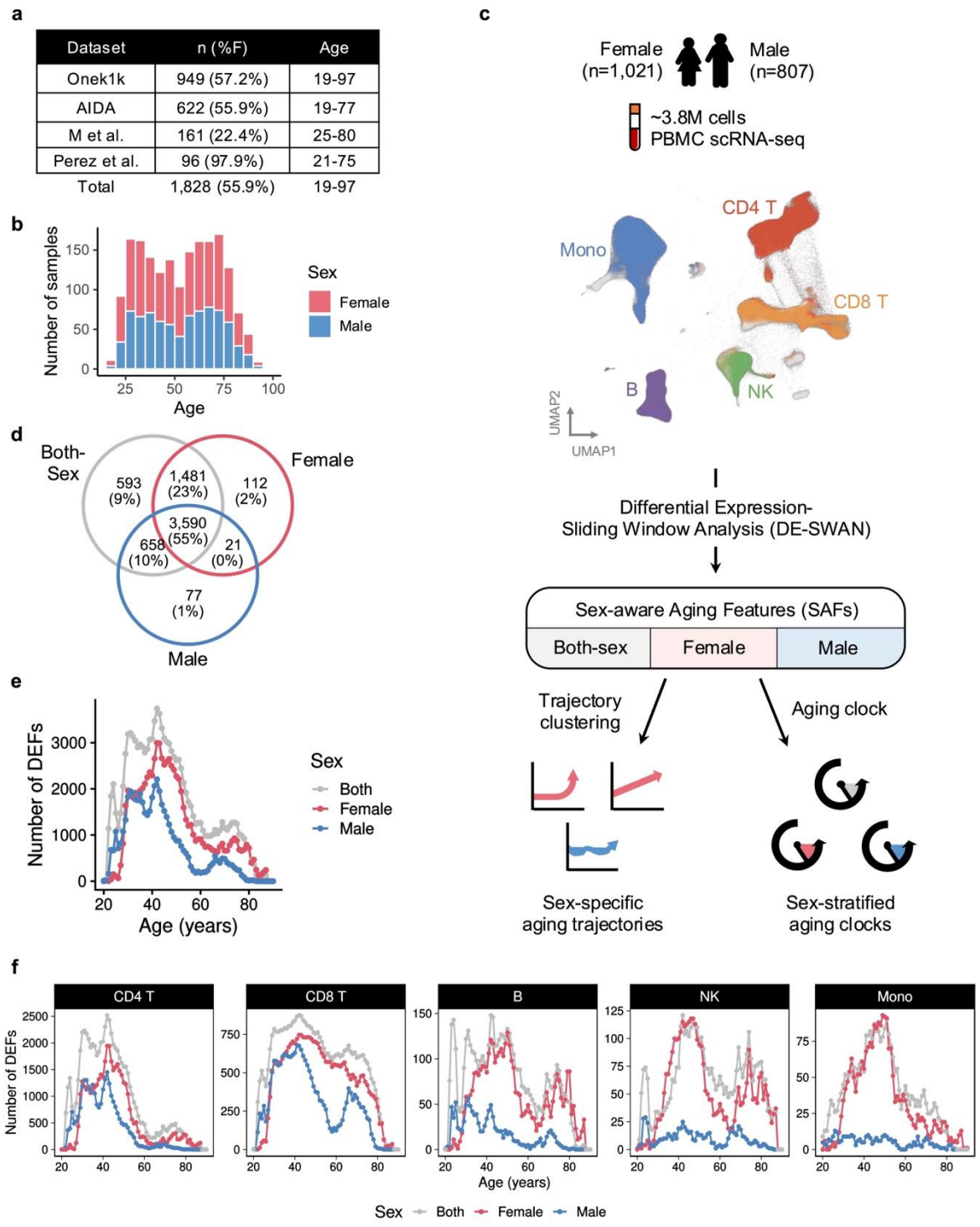


Fig. 1: Nonlinear dynamics of aging captured at single-cell resolution. a, Characteristics of four integrated human PBMC scRNA-seq datasets from healthy donors, including data source, sample size, sex ratio, and age range. **b**, Age and sex distribution in the integrated dataset (n = 1,828 donors). **c**, Uniform manifold approximation and projection (UMAP) embedding of ~3.8 million cells colored by major cell types (CD4 T cells, CD8 T cells, B cells, NK cells, and Monocytes), with schematic diagram illustrating sex-specific, cell type-specific analytical workflow for investigating sex differences in immune aging. **d**, Venn diagram showing overlap of sex-aware aging features (SAFs) identified in both-sex, female-

only, and male-only analyses. **e**-**f**, Number of differentially expressed features (DEFs, q < 0.05) across the age range identified by DE-SWAN algorithm, shown for all major PBMC cell types combined (**e**) and for each cell type separately (**f**). CD4 T, CD4 T cells; CD8 T, CD8 T cells; NK, NK cells; B, B cells; Mono, monocytes.

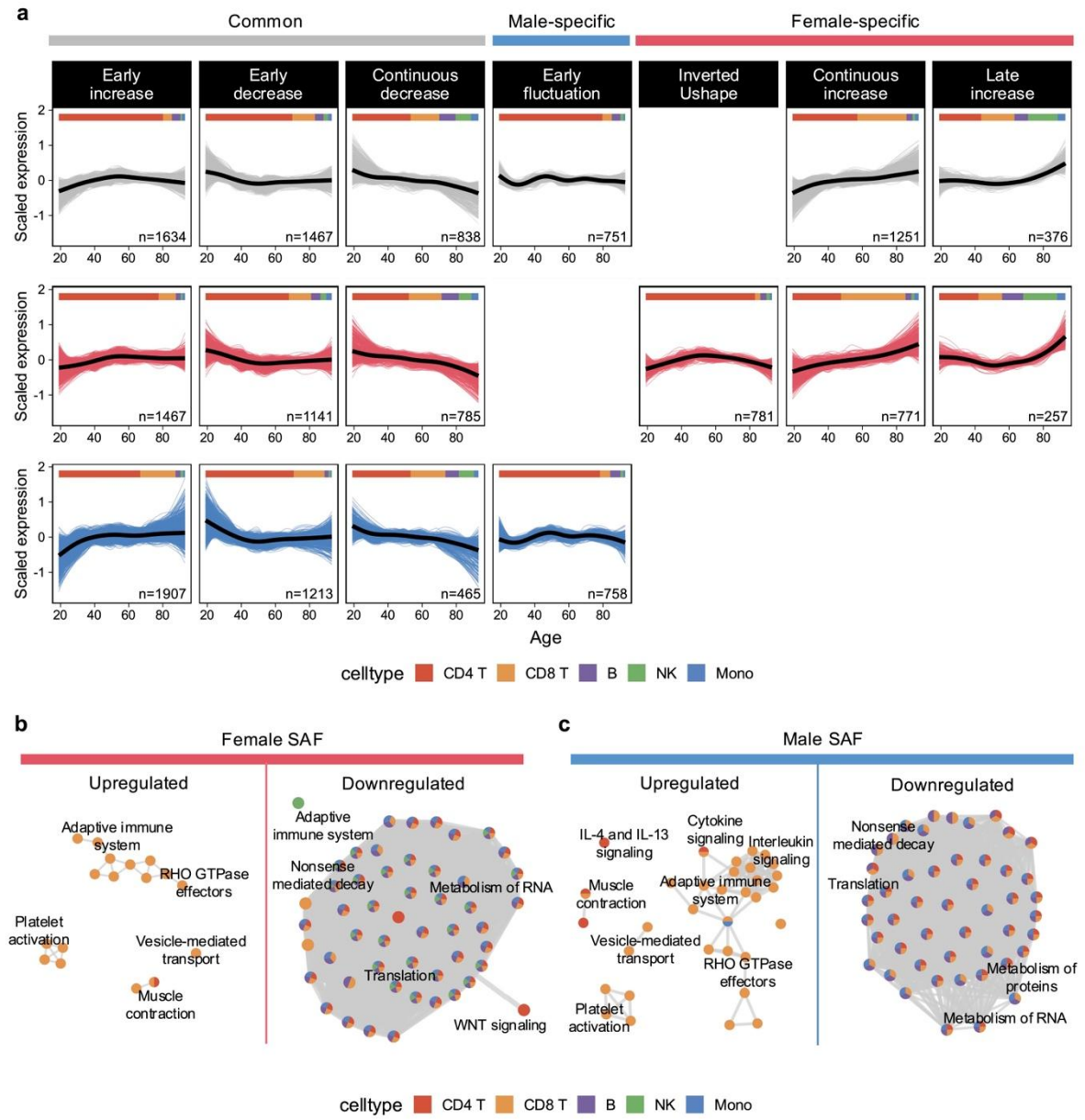


Fig. 2: Sex-aware aging features (SAFs) reveal divergent temporal kinetic trajectories. **a**, Temporal kinetics of SAF clusters across the lifespan. Rows represent clustering results for both-sex SAFs (top), female SAFs (middle), and male SAFs (bottom). Clusters include three patterns common across all SAF sets and four sex-specific patterns. Bar charts (top of each cluster) show the proportion of features from each cell type. **b**-**c**, Network visualization of GSEA results ($q < 0.01$) based on Spearman correlation between pseudobulk expression and age for female SAFs (**b**) and male SAFs (**c**). Nodes represent pathways colored by cell type showing significant enrichment (upregulated: NES > 0; downregulated: NES < 0). Edges connect pathways with pairwise Jaccard index >0.25, indicating gene overlap between terms.

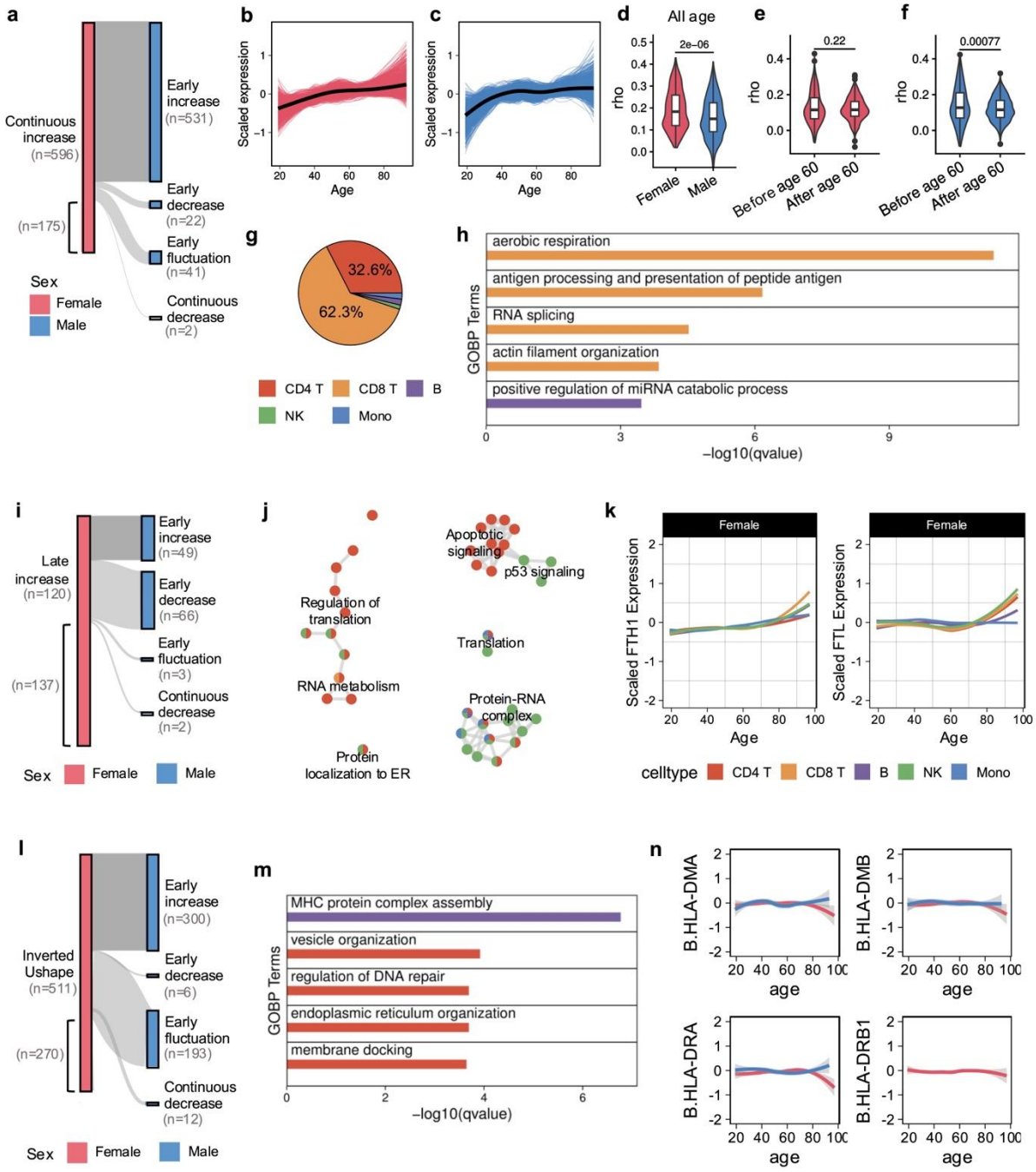


Fig. 3: Female-specific clusters exhibit sustained CD8 T cell activation and late-life senescence. a, Mapping of features in the female-specific continuous increase cluster to male SAF clustering. **b-c**, LOESS trajectories of 531 features assigned to the continuous increase cluster in females and early increase cluster in males, shown for female (**b**) and male donors (**c**). **d-f**, Spearman correlation comparison of the 531 features between female and male donors across all ages (**d**), and stratified by age <60 versus >60 years in females (**e**) and males (**f**). **g**, Cell type composition of the 531 features. **h**, Top GOBP terms from pathway enrichment of the 531 features. **i**, Mapping of features in the female-specific late increase cluster to male SAF clustering. **j**, Network visualization of GOBP pathway enrichment results (q < 0.01) of late increase features. Nodes represent pathways colored by cell type showing significant enrichment. Edges connect pathways with pairwise Jaccard index >0.25. **k**, LOESS trajectories of *FTH1* and *FTL* scaled pseudobulk expression across age in different cell types in female donors. **l**, Mapping of features in the female- specific

inverted U-shape cluster to male SAF clustering. **m**, Top GOBP terms from pathway enrichment of inverted U- shape features. **n**, LOESS trajectories of HLA-DMA, HLA-DMB, HLA-DRA, and HLA-DRB1 scaled pseudobulk expression across age in B cells in female and male donors.

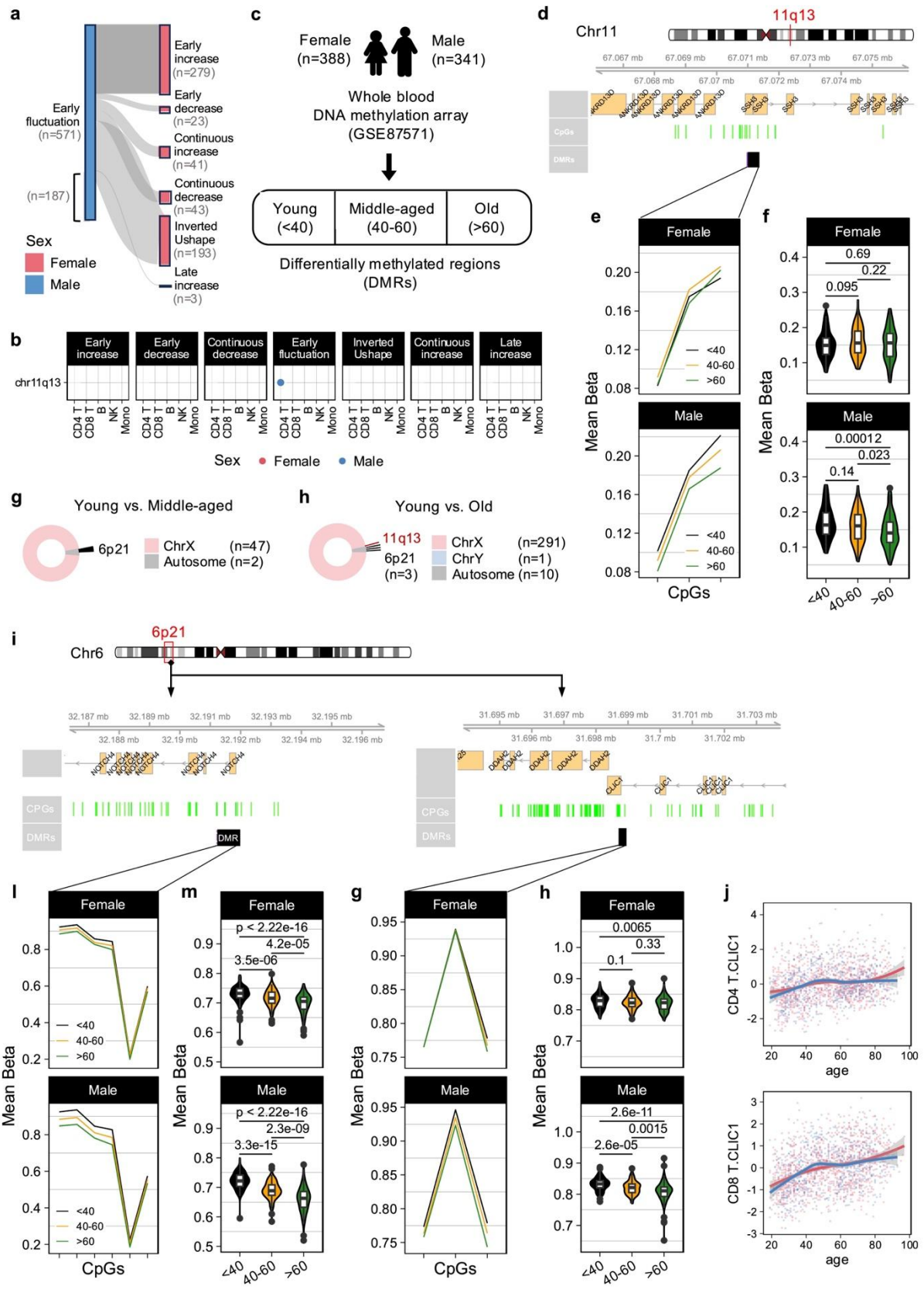
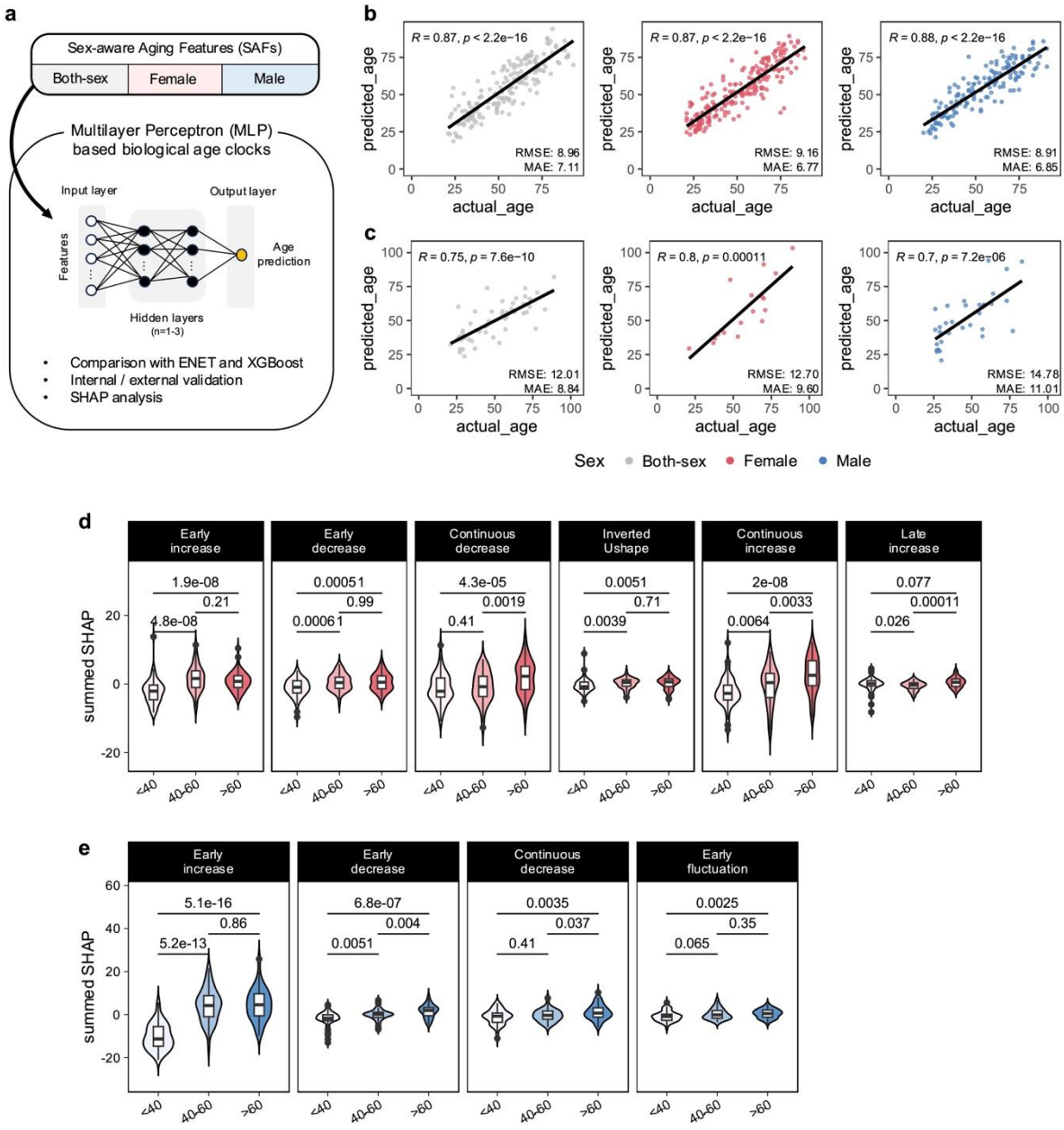


Fig. 4: Male-specific trajectory implicates sex-differential DNA methylation dynamics. **a**, Mapping of features in the male-specific early fluctuation cluster to female SAF clustering. **b**, Positional gene set enrichment results (q < 0.01, MSigDB C1) colored and positioned by sex and cell type. **c**, Schematic diagram of analysis pipeline for identifying differentially

methylated regions (DMRs) with significant age-sex interactions using publicly available whole blood 450K methylation array data. **d**, Chromosomal position of a DMR at chromosome 11q13 from young vs. old comparison. **e**, Mean beta values grouped by age group and stratified by sex for each CpG in the corresponding DMR. **f**, Mean beta values grouped by age group and stratified by sex for all CpGs in the corresponding DMR. **g**-**h**, Chromosomal distribution of DMRs in young vs. middle-aged (**g**) and young vs. old (**h**) comparisons. **i**, Chromosomal positions of two DMRs from the chr6p21 region (young vs. old comparison). **l**, **g**, Mean beta values grouped by age group and stratified by sex for each CpGs in the corresponding DMR. **g**, **h**, Mean beta values grouped by age group and stratified by sex for all CpGs in the corresponding DMR. **j**, Scatter plot and LOESS trajectories of scaled pseudobulk expression of *CLIC1* in CD4 T and CD8 T cells across age, colored by sex.

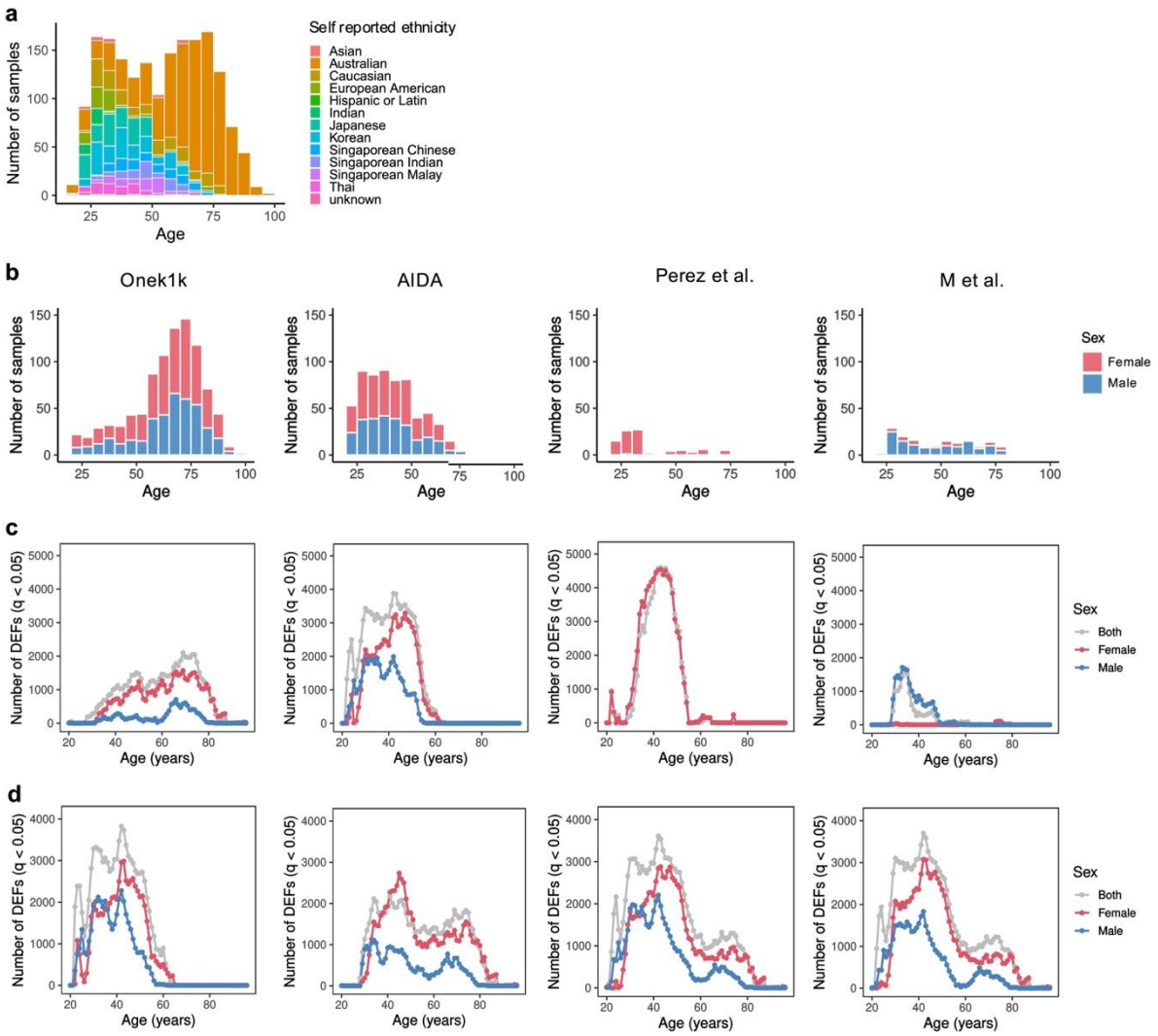


551

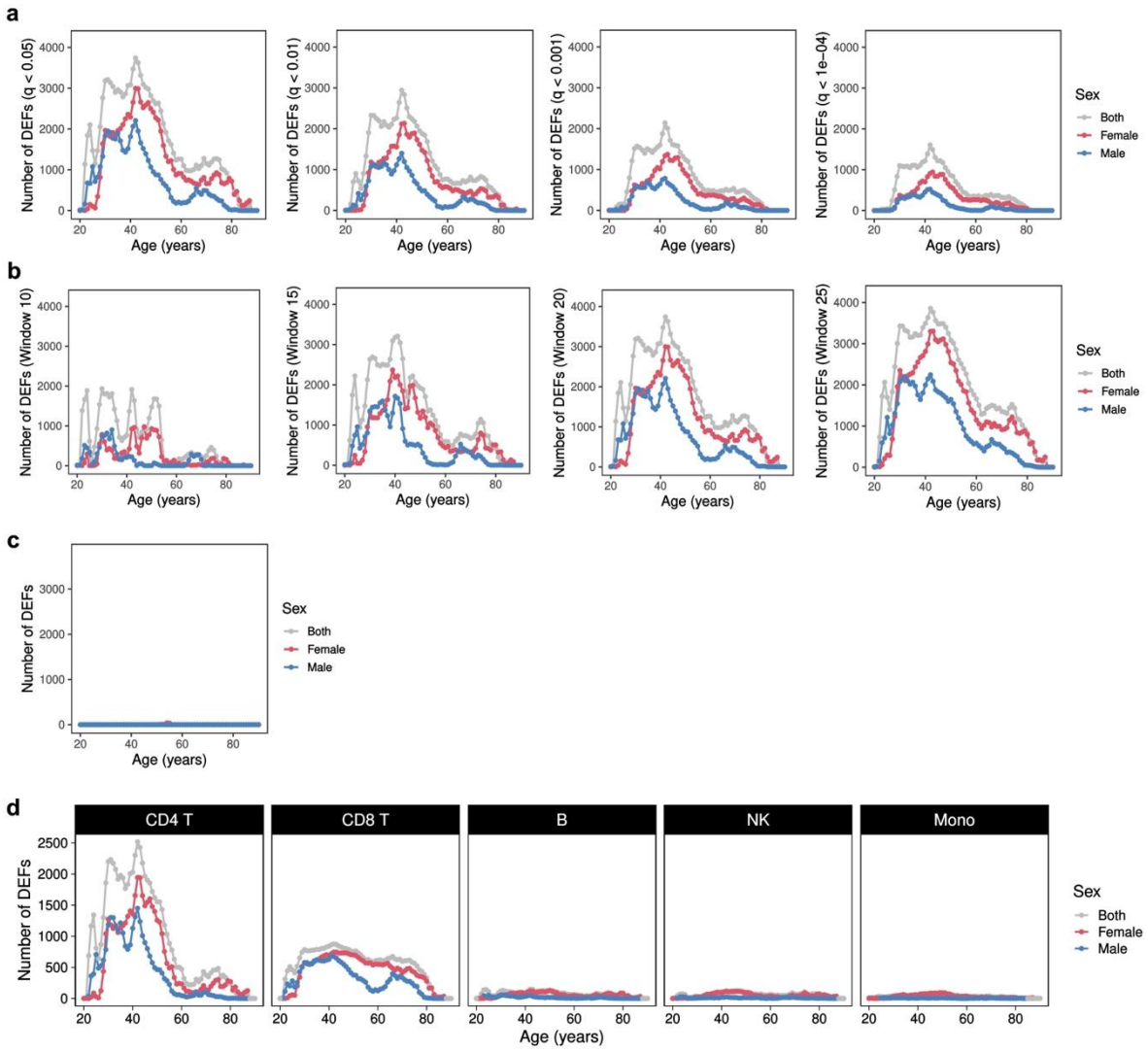
552 **Fig. 5: Sex-stratified biological age clocks outperform sex-combined models and learn**
553 **trajectory-consistent features. a**, Schematic diagram of sex-stratified multilayer perceptron
554 (MLP)-based biological age clocks. **b-c**, Chronological age versus predicted age by sex-
555 combined and sex-stratified models in internal cross-validation test sets (**b**) and healthy
556 external cohorts (**c**). **d-e**, Summed SHAP values by trajectory cluster across age groups in
557 female (**d**) and male donors (**e**).

558

Extended Data



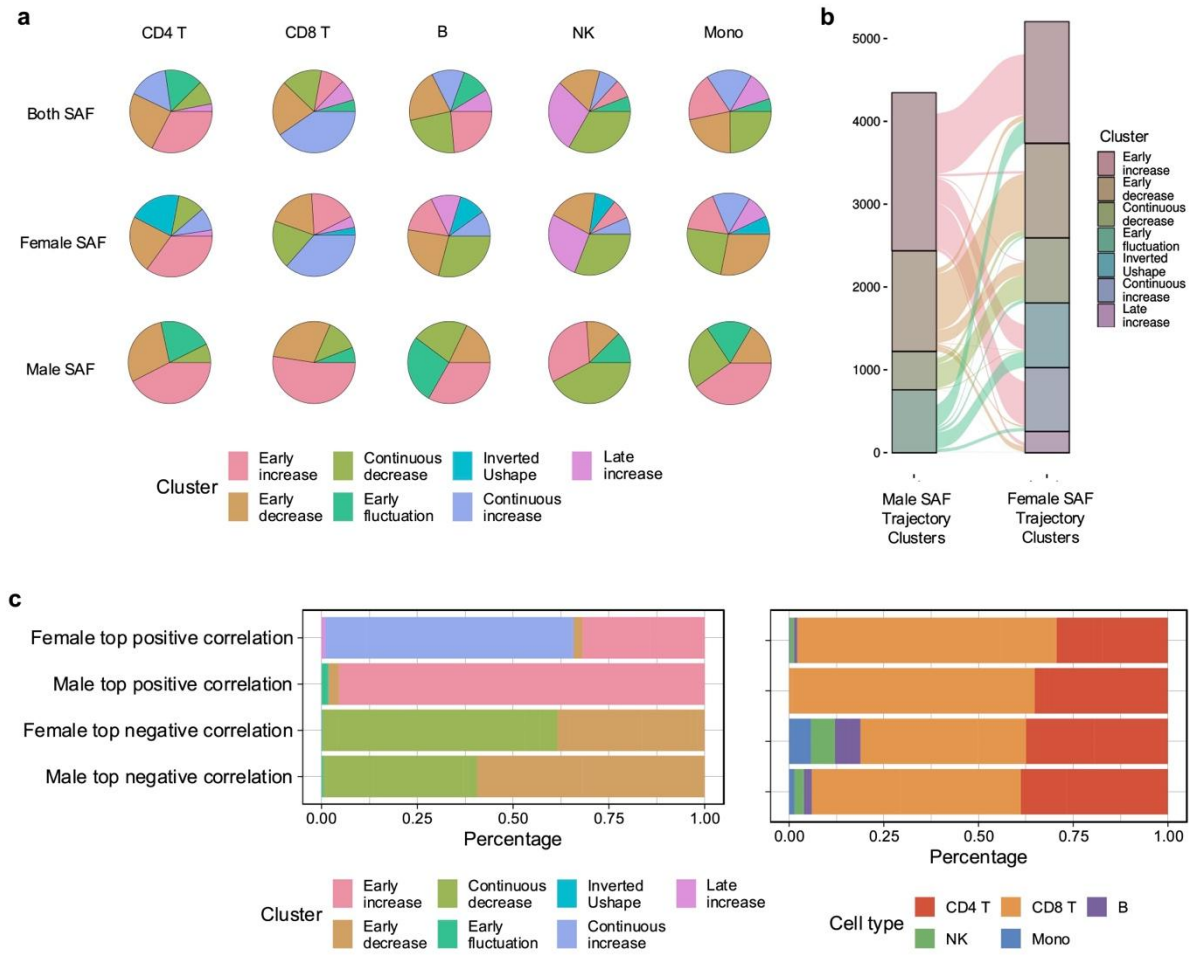
Extended Data Fig. 1: Integrated dataset characteristics and individual dataset DE-SWAN results. a, Age and self-reported ethnicity distribution in the integrated dataset. **b**, Age and sex distribution in individual datasets. **c-d**, Number of differentially expressed features (DEFs, q < 0.05) across age identified by DE-SWAN when using each dataset individually (**c**) or using leave-one-out analysis (**d**).



567

568 **Extended Data Fig. 2: DE-SWAN robustness analysis across parameters and cell**
569 **types. a-c,** Number of differentially expressed features (DEFs) across age identified by DE-
570 SWAN using different q-value cutoffs (**a**), different window sizes (**b**), and randomly shuffled
571 ages (**c**). **d**, Number of DEFs (q < 0.05) across age by each cell type with consistent y-axis.

572



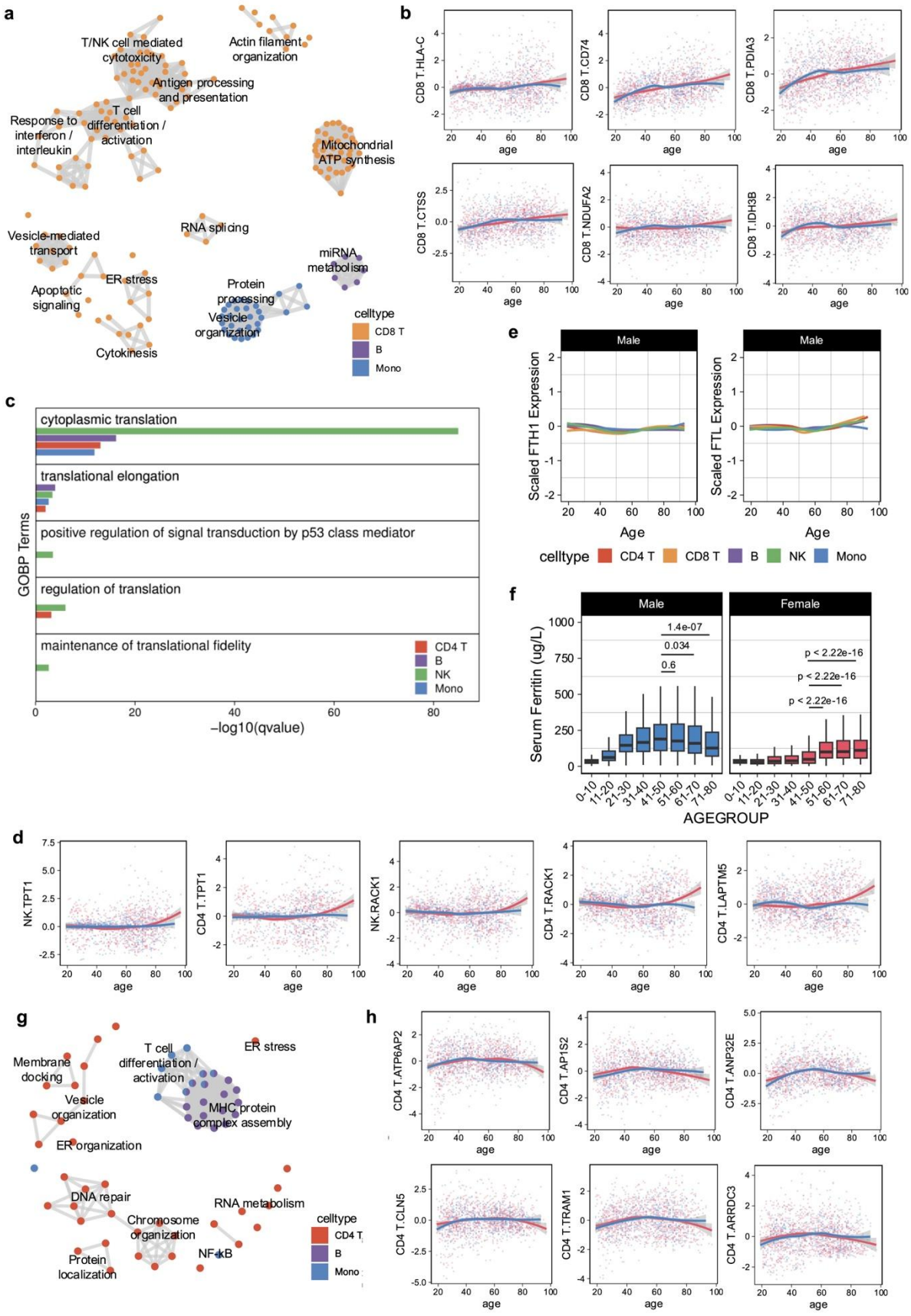
Extended Data Fig. 3: Cell type composition and correlation profiles of trajectory clusters. a, Cell type composition and distribution across temporal kinetics clusters for both-sex, female, and male SAFs. **b**, Assignment flow chart of features across male and female SAF trajectory clusters. **c**, Features with top positive and negative Spearman correlation in females and males (top percentile), colored by cluster assignment and cell type.



580

581 **Extended Data Fig. 4: Functional annotation of sex-stratified trajectory clusters. a,**
582 Pathway enrichment results (q < 0.01, GOBP terms) separated by sex and cell type, ordered
583 by functional module based on manual annotation.

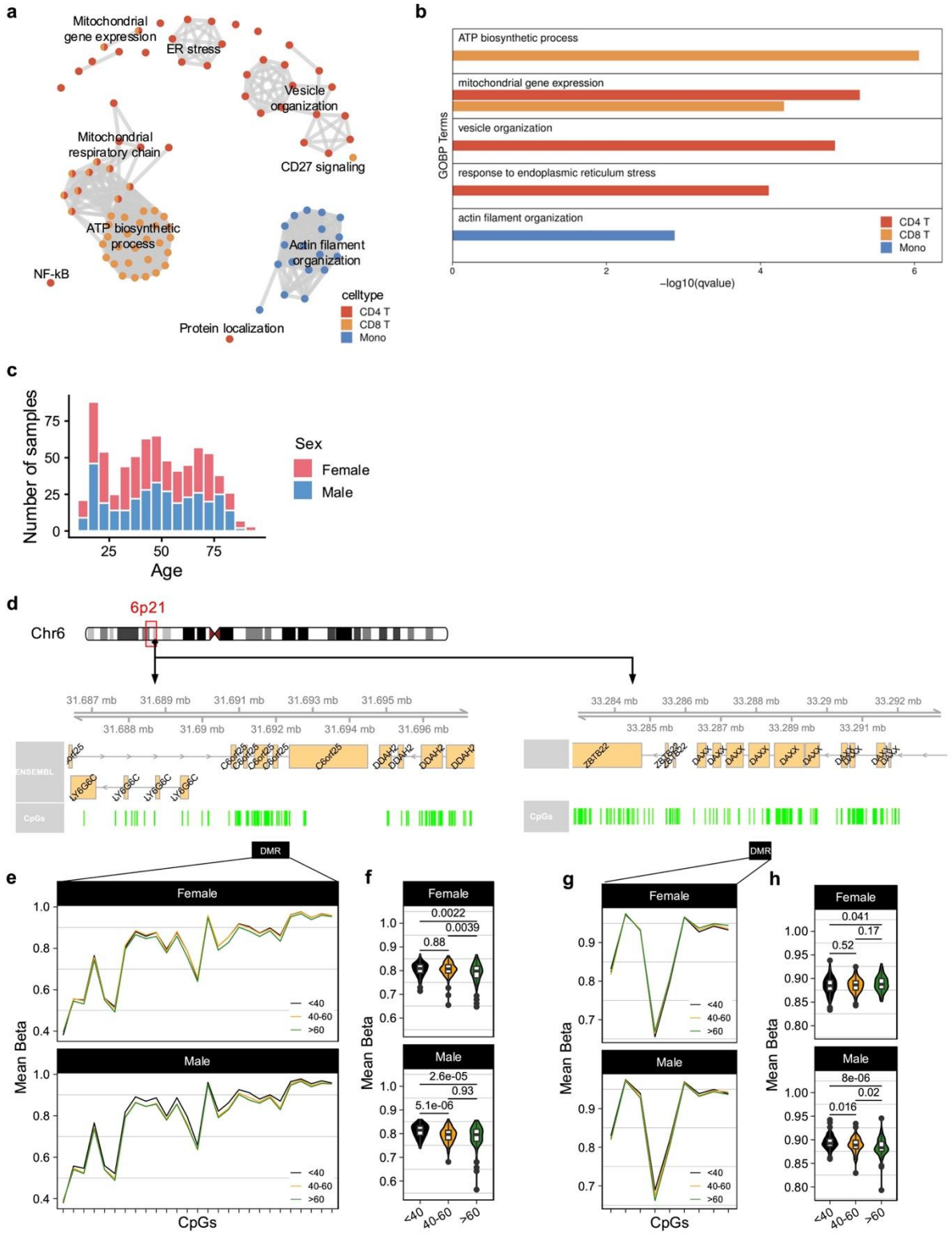
584



585

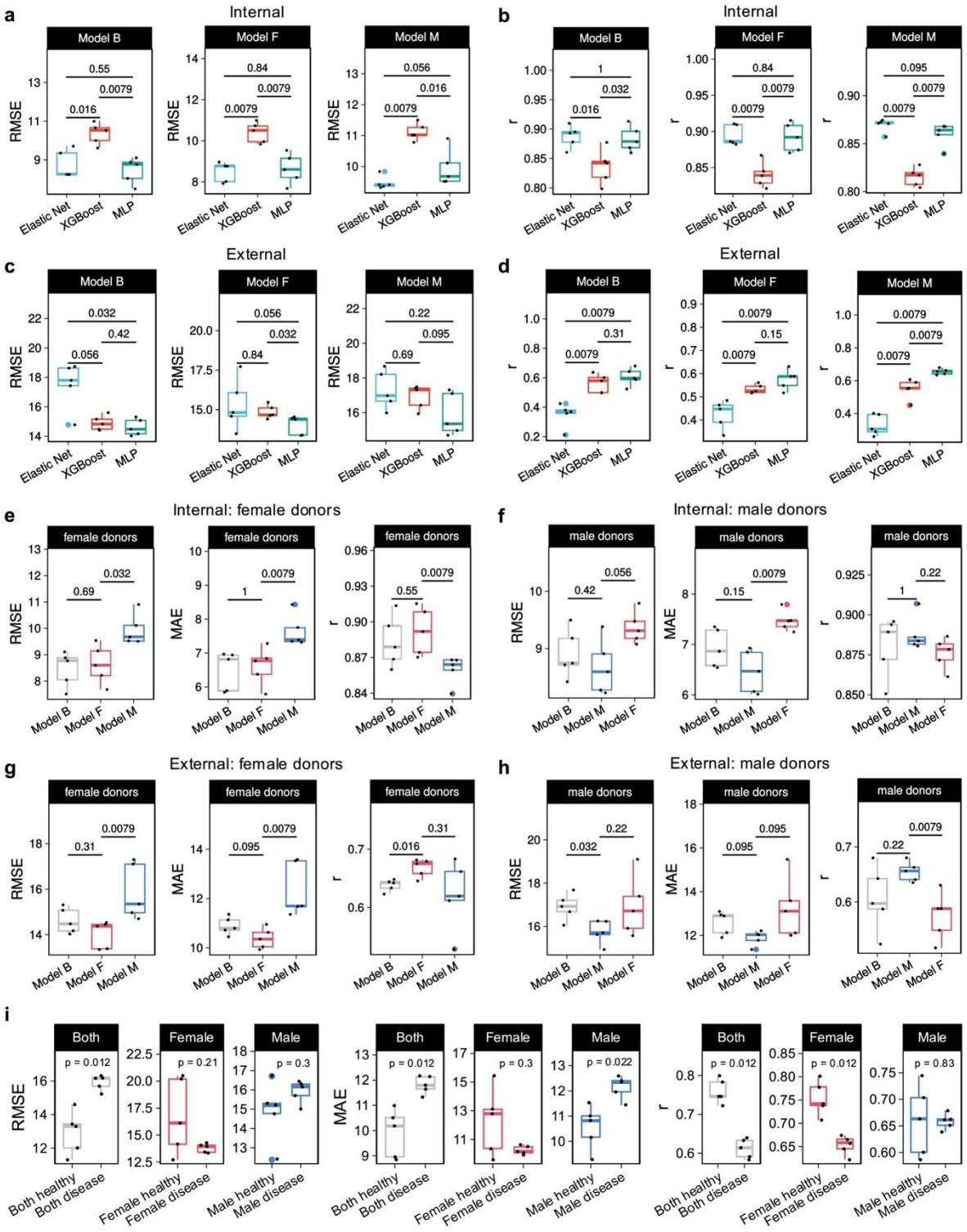
586 **Extended Data Fig. 5: Functional characterization of female-specific trajectory**
587 **clusters. a**, Network visualization of GOBP pathway enrichment results (q < 0.01) for 531

features assigned to the continuous increase cluster in females and early increase cluster in males. Nodes represent pathways colored by cell types showing significant enrichment. Edges connect pathways with Jaccard index >0.25. **b**, Scaled pseudobulk expression trajectories of representative features from the 531 features across age, colored by sex. **c**, Top GOBP terms from pathway enrichment analysis of the female late increase cluster. **d**, Scaled pseudobulk expression trajectories of representative late increase features across age, colored by sex. **e**, *FTH1* and *FTL* scaled pseudobulk expression trajectories across age in different cell types in male donors. **f**, Serum ferritin levels across 10-year age groups by sex from NHANES 2017-2018 dataset (n = 3,378 females, 3,211 males). **g**, Network visualization of GOBP pathway enrichment results ($q < 0.01$) for the inverted U-shape cluster. Nodes represent pathways colored by cell type showing significant enrichment. Edges connect pathways with Jaccard index >0.25. **h**, Scaled pseudobulk expression trajectories of representative inverted U-shape features across age, colored by sex.

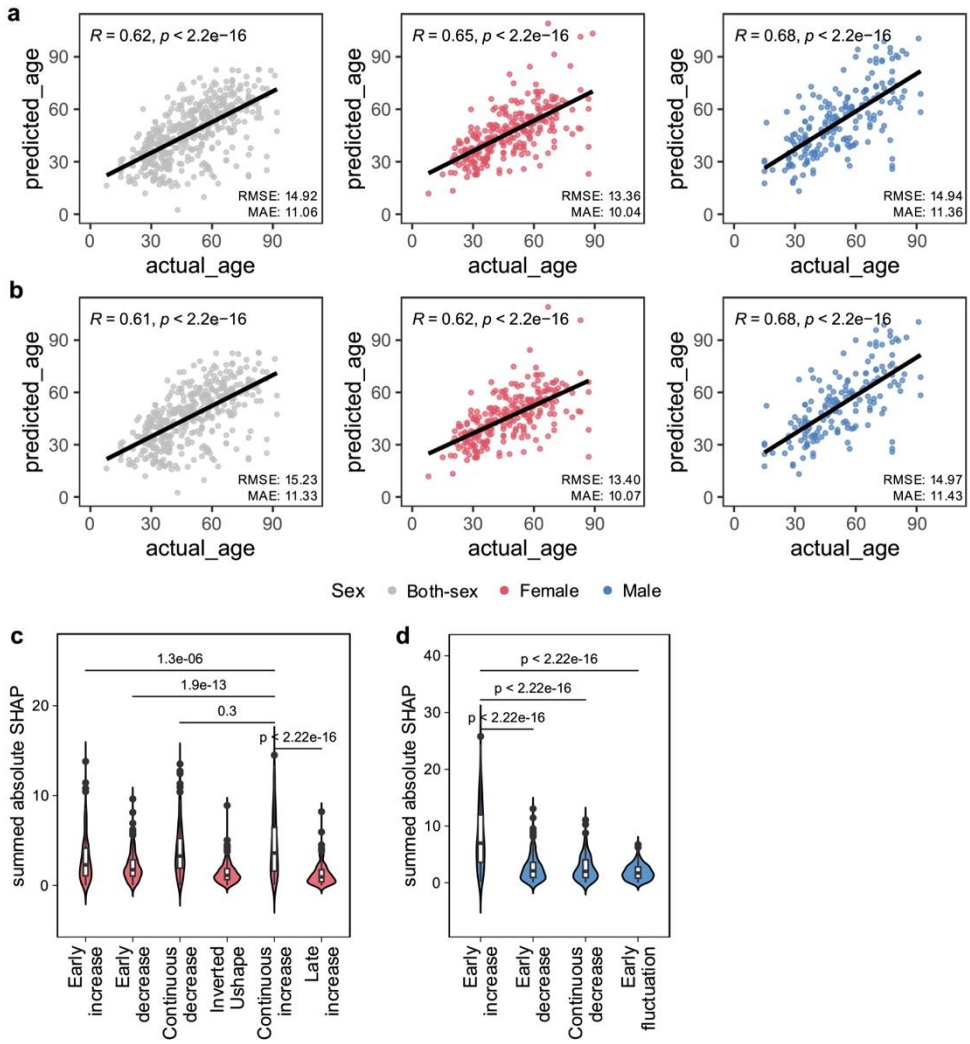


Extended Data Fig. 6: Epigenetic mechanisms underlying male-specific trajectory. a, Network visualization of GOBP pathway enrichment results (q < 0.01) for the male-specific early fluctuation cluster. Nodes represent pathways colored by cell type showing significant enrichment. Edges connect pathways with Jaccard index >0.25. **b**, Top GOBP terms from pathway enrichment analysis of the early fluctuation cluster. **c**, Age and sex distribution in the DNA methylation array dataset (GSE87571). **d**, Chromosomal positions of two chr6p21 DMRs from young vs. middle-aged and young vs. old comparisons. **e**, **g**, Mean beta values

grouped by age group and stratified by sex for each CpG in the corresponding DMRs. **f**, **h**, Mean beta values grouped by age group and stratified by sex for all CpGs in the corresponding DMRs.



Extended Data Fig. 7: Model performance comparison and sex-stratified clock validation. **a**, **b**, RMSE (**a**) and Pearson *r* (**b**) comparison among Elastic Net, XGBoost, and MLP across 5 cross-validation folds in internal test sets. **c**, **d**, RMSE (**c**) and Pearson *r* (**d**) comparison among Elastic Net, XGBoost, and MLP across 5 cross-validation folds in external validation sets (healthy and disease cohorts combined). **e**, **f**, Performance metrics comparison of sex- combined versus sex-specific models in female (**e**) and male donors (**f**) in internal test sets. **g**, **h**, Performance metrics comparison of sex-combined versus sex-specific models in female (**e**) and male donors (**f**) in external validation sets. **i**, RMSE, MAE, and Pearson *r* comparison of MLP models on healthy versus disease cohorts.



Extended Data Fig. 8: Age prediction performance and SHAP analysis of final sex-stratified biological aging clocks. a, **b**, Chronological age versus predicted age by sex-combined and sex-stratified models on full external validation set (healthy and disease cohorts combined) (**a**) and disease cohort only (**b**). **c**, **d**, Summed absolute SHAP values by trajectory cluster in female (**c**) and male donors (**d**).

Methods

Human PBMC scRNA-seq datasets

OneK1K dataset

The scRNA-seq data from OneK1K cohort[15] was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1) in h5ad format with raw counts. PBMCs were barcoded with the Single Cell 3' Library and Gel Bead Kit (10x Genomics) and sequenced on Illumina NovaSeq 2000. The original dataset included 981 healthy donors of Northern European ancestry (Austrailian) aged 19-97 years (57.6% female). After quality control, 949 donors aged 19–97 years (57.2% female) were included as an internal dataset for DE-SWAN analysis, aging gene expression trajectory clustering analysis, and biological age clock construction.

AIDA v2 dataset

The scRNA-seq data from the Asian Immune Diversity Atlas (AIDA) study[16] was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/ced320a1-29f3-47c1-a735-513c7084d508) in h5ad format with raw counts. PBMCs from 16 donors each were pooled and demultiplexed computationally using genotyping data. Cells were captured with the Chromium Single Cell 5' v2 Reagent Kit (10x Genomics) and sequenced on Illumina NovaSeq 6000. The original dataset included 625 healthy donors across 8 self-reported ethnicity groups (Korean, Japanese, Singaporean Chinese, Singaporean Indian, Singaporean Malay, Thai, Indian, and European) aged 19-77 years (56% female). After quality control, 622 donors aged 19-77 years (55.9% female) were included as an internal dataset for downstream analyses mentioned above.

M et al. dataset

The scRNA-seq data from ref[5] was downloaded from Synapse.org (https://www.synapse.org/Synapse:syn49637038/files/) in h5ad format with raw counts. PBMCs were captured with Chromium Single Cell 5' v2 Reagent Kit (10x Genomics) and sequenced on a NovaSeq S4 (300 cycles) flow cell. The dataset included 317 samples from 166 healthy donors of Caucasian lineage aged 25-81 years (21.7% female), including multiple visits from some donors. After quality control and retaining only the first visit for donors with multiple visits, 161 individuals aged 25-80 years (22.4% female) were included as an internal dataset for downstream analyses mentioned above.

Perez et al. dataset

The scRNA-seq data from ref[17] was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/436154da-bcf1-4130-9c8b-120ff9a888f2) in h5ad format with raw counts. PBMCs were profiled using Chromium Single Cell 3' v2 chemistry (10x genomics) and sequenced on HiSeq4000 or NovaSeq6000. The original dataset included 99 healthy donors and 162 SLE patients of both European and Asian descent aged 20-83 years (93.5% female). After quality control, 88 healthy donors aged 21-75 years (97.9% female) were included as an internal dataset for downstream analyses mentioned above. Data from 148 SLE patients aged 20-83 years (89.9% female) were included in the external dataset for biological age clock validation.

Ren et al. dataset (COVID-19)

The scRNA-seq data from ref[88] was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/0a839c4b-10d0-4d64-9272-684c49a2c8ba) in h5ad format with raw counts. Libraries were prepared using the Chromium Single Cell 3' v2, 3' v3, or 5' v2 Reagent Kits (10x Genomics) and sequenced using either DIPSEQ, BGISEQ or Illumina platforms. The original dataset included blood, saliva, and lung samples from 25 healthy donors and 171 COVID-19 patients with varying severity, aged 6-92 (35.7% female). After quality control, restricting to blood samples with known sex, 199 samples from 25 healthy donors and 141 COVID-19 patients aged 8-87 (37.3% female) were included as an external dataset for biological age clock validation.

Wellcome dataset (COVID-19)

The scRNA-seq data from ref[89] was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/ddfad306-714d-4cc0-9985-d9072820c530) in h5ad format with raw counts. Libraries were prepared using Chromium NextGEM Single Cell V(D)J Reagent kits v1.1 (10x Genomics) and sequenced on NovaSeq 6000. The original dataset included 29 healthy donors and 91 COVID-19 patients with varying severity, aged 20-92 (49.2% female). After quality control, excluding resampled timepoints and including only healthy and COVID-19 samples, 14 healthy donors and 71 COVID-19 patients aged 20-92 (48.2% female) were included as an external dataset for biological age clock validation.

GSE268936 (Glaucoma)

The scRNA-seq data was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/de2cde16-c8d3-4a6d-80be-1be9e879aaca) in h5ad format with raw counts. Libraries were prepared using the Chromium Single Cell 5' v2 Reagent Kit (10x Genomics). The original dataset included 3 healthy donors and 5 glaucoma patients aged 58-82 (62.5% female). After quality control, all 8 individuals were included as an external dataset for biological age clock validation.

Heimlich et al. dataset (Clonal hematopoiesis)

The scRNA-seq data from ref[90] was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/0aab20b3-c30c-4606-bd2e-d20dae739c45) in h5ad format with raw counts. Libraries were prepared using the Chromium Single Cell 3' v3 Reagent Kit (10x Genomics) and sequenced on NovaSeq 6000. The original dataset included 7 healthy donors and 17 Clonal hematopoiesis (CH) patients aged 39-91 (37.5% female). After quality control, all individuals were included as an external dataset for biological age clock validation.

Preprocessing and pseudobulk expression generation from single-cell datasets

Downloaded datasets were filtered for quality control using scanpy (v1.11.4)[91]. We excluded cells with <1,000 or >20,000 UMIs, <300 or >6,000 features, and >10% mitochondrial reads. Raw counts and metadata from h5ad files were extracted to create Seurat objects using anndata (v0.8.0)[92]. Cell types were annotated using RunAzimuth function in Azimuth (v0.5.0) with the human PBMC reference[93]. Donors with fewer than three cells in any major cell types (CD4 T cells, CD8 T cells, B cells, NK cells, Monocytes) were excluded prior to pseudobulk generation. To construct pseudobulk profiles, raw counts were normalized using NormalizeData and average expression per cell was calculated using AverageExpression rather than summed counts using AggregateExpression in Seurat (v5.0.1)[94]. This approach

minimizes biases introduced by differences in cell number, sequencing depth, and dataset-specific sampling, which is particularly important when modeling continuous age-associated trajectories across multiple datasets (Supplementary Figure 3). Pseudobulk expression matrices from all datasets were merged, retaining only 7,385 features (cell type-gene pairs) that were expressed across all donors in all four datasets for downstream analyses.

Sex-stratified differential expression sliding window analysis (DE-SWAN)

DE-SWAN was applied on the merged pseudobulk expression matrix as previously described[2]. Briefly, for each central age a_0, individuals were partitioned into two groups based on an age window of width w (default w = 20 years): one comprising individuals with ages in the range $[a_0 - w/2, a_0)$ and the other comprising individuals with ages in the range $(a_0, a_0 + w/2]$. Gene expression between these two groups was compared using a linear model with dataset identity as a covariate. This comparison was iteratively applied using central ages from 20 to 90 years, incremented by one year, to capture linear and nonlinear expression trajectories. P-values were calculated for all genes at each sliding window and adjusted for multiple testing using the Benjamini–Hochberg correction.

Analyses were conducted using all samples combined (both-sex) or sex-stratified (female, male) to identify sex-specific transcriptional dynamics. Features showing significant differential expression ($q < 0.05$, default w) at any central age were designated sex-aware aging features (SAFs). Robustness was evaluated using multiple window sizes (w = 5, 10, 15, 20 years) and q-value thresholds ($q < 0.0001$, 0.001, 0.01, 0.05). To assess statistical significance, analyses were repeated with randomly permuted ages in both combined and sex-stratified datasets.

Sex-aware aging feature (SAF) trajectory clustering

The merged pseudobulk expression matrix was corrected for dataset-specific batch effects using removeBatchEffect function from limma (v3.62.1)[95]. Three matrix subsets were generated for separate trajectory estimation and clustering: both SAFs with all donors, female SAFs with female donors only, and male SAFs with male donors only. For each subset, expression values were z-score normalized, and locally estimated scatterplot smoothing (LOESS) models were fitted to estimate transcriptional trajectories. The span parameter for each LOESS model was optimized through cross-validation to minimize overfitting.

Trajectories were clustered using fuzzy c-means clustering implemented in Mfuzz (v2.66.0)[96]. Fuzzification parameter (*m1*) was estimated using mestimate function, and the optimal number of clusters was determined by identifying the elbow point in minimum centroid distances across 2-40 clusters (step size = 2). Following initial clustering, highly correlated clusters were merged using a graph-based approach: clusters were represented as nodes, edges were drawn between clusters with highly correlated centroid trajectories (Pearson $r > 0.8$), and all clusters within the same connected component were merged. To assess robustness, the full clustering and merging procedure was repeated 10 times with different random initializations. The number and composition of merged clusters were highly consistent across runs for all SAFs (Supplementary Figure 2). Accordingly, a single representative clustering per SAF was used for downstream analyses. Only features with membership scores > 0.5 were included in centroid correlation calculations, cluster merging, and subsequent analyses.

GSEA network analysis based on spearman correlation

Spearman correlations between each SAF and age were calculated from the three corrected pseudobulk expression matrices (both-sex, female, and male) using the using cor.test function in stats (rbase v4.4.3). Features in each SAF set were ranked by Spearman correlation within each cell type to generate gene lists for GSEA using gseWP and gsePathway functions in clusterProfiler (v3.18.0)[97]. Pathways with q < 0.01 were visualized as networks using ggraph (v2.2.1). Nodes represent terms, colored by cell type showing significant enrichment. Edges connect terms with pairwise Jaccard index >0.25, representing the extent of gene overlap between terms.

Functional annotation of trajectory cluster features

We performed functional annotation on features from each trajectory cluster separately for each SAF set. Pathway enrichment analysis using Gene Ontology Biological Process (GOBP) terms was conducted with the enrichGO function in clusterProfiler (v3.18.0, minGSSize = 3)[97]. The top 3 GOBP terms per cell type per SAF set were manually annotated into functional modules based on Gene Ontology hierarchies and literature. We also performed positional gene set enrichment using the C1 collection from MSigDB (v7.5.1)[98] via the enricher function in clusterProfiler (v3.18.0, minGSSize = 3)[97] to assess chromosomal location enrichment. Significant results (qvalue < 0.01) were visualized as networks using ggraph (v2.2.1) as described above.

Cell type proportion analysis

Spearman correlations between donor age and cell type proportions (level 1 and level 2 Azimuth annotations, v0.5.0) were calculated for all donors, female-only, and male-only cohorts for sex-stratified analysis using cor.test function from stats (rbase v4.4.3). FDR was calculated to correct for multiple testing using p.adjust function from stats.

Serum ferritin analysis from NHANES data

NHANES 2017-2018 demographics and serum ferritin data were downloaded from the Centers for Disease Control and Prevention website (https://wwwn.cdc.gov/nchs/nhanes/). Participants with available age and sex information were grouped into 10-year age intervals and stratified by sex to analyze age-associated changes in serum ferritin levels.

Differentially methylated region (DMR) analysis

Raw IDAT files from Infinium HumanMethylation450K BeadChip were downloaded from GSE87571[41], retaining 729 samples with unique age and sex information. Methylation data were processed using minfi (v1.52.0)[99]. CpG probes with poor detection (P > 0.01), SNP-associated probes, and cross-hybridizing probes were excluded. Beta values were extracted using getBeta function in minfi (v1.44.0)[100] and normalized via beta-mixture quantile normalization using champ.norm function in ChAMP (v.2.28.0)[101]. Normalized beta values were converted to M values using B2M function in Enmix (v1.34.0)[102].

M values were analyzed for DMRs using dmrcate function in DMRcate (v3.2.1)[103] with the contrast (age group * sex) to identify DMRs with sex-age interactions. DMRs were visualized using DMR.plot function in DMRcate, and mean beta values per CpG and per sample were calculated for plotting.

Constructing sex-stratified scRNA-seq biological aging clocks

We constructed biological age clocks using the integrated pseudobulk expression matrix from DE-SWAN and trajectory clustering analyses as the internal training and testing set. SAFs identified from DE-SWAN, capturing both linear and nonlinear age-associated transcriptional dynamics, served as input features. For both-sex clocks, we used both-sex SAFs with all donors. For sex-stratified clocks, we used sex-specific SAFs (male or female) with corresponding donor subsets. To avoid sample size bias, we randomly subsampled half of the donors (sex-balanced) using StratifiedShuffleSplit function from scikit-learn (v1.3.0)[104] for both-sex clock training.

We implemented three algorithms to model biological age: Elastic Net Regression (ENET), eXtreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP) neural network. ENET was chosen for its simplicity and robustness in handling multicollinearity while capturing primarily linear associations. XGBoost, a gradient boosting framework based on decision trees, is well-suited for modeling complex nonlinear interactions. MLP, a feedforward artificial neural network, was employed as a baseline deep learning approach to capture nonlinear dependencies and potential higher-order interactions between features.

Models were trained using five-fold cross-validation, performed independently for the both-sex and sex-stratified settings. Model performance was evaluated using root mean squared error (RMSE), mean absolute error (MAE), and Pearson correlation coefficient (r) between predicted and chronological ages.

ENET and XGBoost models were built using the scikit-learn (v1.7.2)[104] and py-xgboost-gpu (v2.1.4)[105], respectively. ENET hyperparameters (α and L1 ratio) were optimized using grid search. XGBoost hyperparameters (learning rate η, maximum tree depth, number of boosting rounds) were tuned via randomized grid search. MLP was implemented in PyTorch (v2.5.1)[106] with a fully connected architecture (1–3 hidden layers, 64–256 units per layer), ReLU activation, dropout regularization, and early stopping based on validation loss. Models were trained with the Adam optimizer (learning rate 1e-4 to 1e-2, weight decay 1e-6 to 1e-2) for up to 50 epochs, using ReduceLROnPlateau scheduler (mode: min, patience: 3, factor: 0.5).

External validation

For external validation, we curated publicly available human PBMC scRNA-seq datasets from independent studies with healthy controls and explicit donor age information. The combined dataset served as an external validation cohort, with samples categorized as healthy or diseased based on clinical metadata. Age prediction on external data was performed using the same pipelines as internal validation: ENET (scikit-learn v1.7.2)[104], XGBoost (py-xgboost-gpu v2.1.4)[105], and MLP (PyTorch v2.5.1)[106].

SHAP analysis and correction

For SHAP analysis, we used a fixed background reference of 100 donors (50 male, 50 female) randomly selected from the training data of each model to ensure representativeness while maintaining sex balance and avoiding test data leakage. SHAP values were computed using DeepExplainer function from shap (v0.48.0)[107].

To mitigate batch effects, we applied batch correction to raw SHAP values. For each feature, the mean raw SHAP contribution was calculated per dataset and subtracted from all samples in that dataset, yielding corrected SHAP values. This centers feature contributions relative to dataset baselines, reducing systematic offsets from dataset differences.

Statistics and reproducibility

We excluded individuals without age and sex information and included only the first visit for donors with multiple visits. No statistical method was used to pre-determine sample size. Data distribution was assumed to be normal, but this was not formally tested. The training and validation sets were randomly selected.

Data availability

Data from OneK1K cohort was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1). Data from Asian Immune Diversity Atlas (AIDA) study was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/ced320a1-29f3-47c1-a735-513c7084d508). Data from M et al. was downloaded from Synapse.org (https://www.synapse.org/Synapse:syn49637038/files/). Data of healthy and SLE patients from Perez et al. was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/436154da-bcf1-4130-9c8b-120ff9a888f2). Data of healthy and COVID-19 patients from Ren et al. was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/0a839c4b-10d0-4d64-9272-684c49a2c8ba). Data of healthy and COVID-19 patients from Wellcome cohort was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/ddfad306-714d-4cc0-9985-d9072820c530). Data from COvid-19 Multi-omics Blood Atlas (COMBAT) Consortium was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/8f126edf-5405-4731-8374-b5ce11f53e82). Glaucoma cohort data was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/de2cde16-c8d3-4a6d-80be-1be9e879aaca). Rheumatoid arthritis cohort data was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/de2cde16-c8d3-4a6d-80be-1be9e879aaca). Clonal hematopoiesis cohort data was downloaded from CELLxGENE (https://cellxgene.cziscience.com/collections/0aab20b3-c30c-4606-bd2e-d20dae739c45). DNA methylation data was downloaded from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE87571). NHANES 2017-2018 data was downloaded from (https://wwwn.cdc.gov/nchs/nhanes/).

Code availability

The codes for non-linear aging, clustering, DNA methylation, and biological age clock analyses are available via GitHub at https://github.com/HaileyHryPark/pbmc-aging-snakemake.

Acknowledgements

This work was supported by funding from the National Medical Research Council (OFIRG24jan-0083/ MOH-001622 to J.B. and OFLCG22may-0011/ MOH-001327-05 to J.B.), Ministry of Education Singapore (MOE-T2EP30125-0010 to J.B.), Agency for Science and Technology (Industry Alignment Fund Prepositioning grant, H24J4a0019 to J.B.) and Duke NUS (Duke-NUS-GCR/2024/0034 to J.B.)

Author contributions

H.P. and J.B. conceptualized the study. H.P. designed and performed all bioinformatics analysis. N.L.B, A.B., N.T., and J.G. edited the manuscript and provided critical input on the

results. H.P. and J.B. wrote the manuscript. All authors reviewed and approved the final version.

Competing interests

The authors declare no competing interests.

Reference

1. Shen, X. *et al*. Nonlinear dynamics of multi-omics profiles during human aging. *Nat. Aging* 1–16 (2024) doi:10.1038/s43587-024-00692-2.
2. Lehallier, B. *et al*. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
3. Márquez, E. J. *et al*. Sexual-dimorphism in human immune system aging. *Nat. Commun.* **11**, 751 (2020).
4. Mogilenko, D. A. *et al*. Comprehensive Profiling of an Aging Immune System Reveals Clonal GZMK+ CD8+ T Cells as Conserved Hallmark of Inflammaging. *Immunity* **54**, 99-115.e12 (2021).
5. M, T. *et al*. 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**, (2023).
6. Bohacova, P. *et al*. Multidimensional profiling of human T cells reveals high CD38 expression, marking recent thymic emigrants and age-related naive T cell remodeling. *Immunity* **57**, 2362-2379.e10 (2024).
7. Austad, S. N. & Fischer, K. E. Sex Differences in Lifespan. *Cell Metab.* **23**, 1022–1033 (2016).
8. Hägg, S. & Jylhävä, J. Sex differences in biological aging with a focus on human studies. *eLife* **10**, e63425 (2021).
9. Gilmer, G. *et al*. Female aging: when translational models don't translate. *Nat. Aging* **3**, 1500–1508 (2023).
10. Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nat. Rev. Immunol.* **16**, 626–638 (2016).

11. Fish, E. N. The X-files in immunity: sex-based differences predispose immune responses. *Nat. Rev. Immunol.* **8**, 737–744 (2008).

12. Ortona, E. *et al.* Sex-based differences in autoimmune diseases. *Ann. Ist. Super. Sanita* **52**, 205–212 (2016).

13. Gilmer, G. *et al.* Menopause-induced 17β-estradiol and progesterone loss increases senescence markers, matrix disassembly and degeneration in mouse cartilage. *Nat. Aging* **5**, 65–86 (2025).

14. Han, A., Kim, J. Y., Kwak-Kim, J. & Lee, S. K. Menopause is an inflection point of age-related immune changes in women. *J. Reprod. Immunol.* **146**, 103346 (2021).

15. Yazar, S. *et al.* Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease. *Science* https://doi.org/10.1126/science.abf3041 (2022) doi:10.1126/science.abf3041.

16. Kock, K. H. *et al.* Asian diversity in human immune cells. *Cell* **188**, 2288-2306.e24 (2025).

17. Perez, R. K. *et al.* Single-cell RNA-seq reveals cell type–specific molecular and genetic associations to lupus. *Science* **376**, eabf1970 (2022).

18. Zhu, H. *et al.* 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 (2023).

19. Li, M. *et al.* Age related human T cell subset evolution and senescence. *Immun. Ageing A* **16**, 24 (2019).

20. Gheitasi, R. *et al.* Age- and sex-associated differences in immune cell populations. *iScience* **28**, 113092 (2025).

21. Ding, Y. *et al.* Comprehensive human proteome profiles across a 50-year lifespan reveal aging trajectories and signatures. *Cell* **0**, (2025).

22. Colin-York, H. *et al.* Cytoskeletal Control of Antigen-Dependent T Cell Activation. *Cell Rep.* **26**, 3369-3379.e5 (2019).

23. Wang, Z. *et al.* Interleukin-37 promotes colitis-associated carcinogenesis via SIGIRR-mediated cytotoxic T cells dysfunction. *Signal Transduct. Target. Ther.* **7**, 19 (2022).

24. Doherty, E. *et al.* Pathogenic role of MIF receptor (CD74) expressing T cells in inflammatory arthritis. *Proc. Natl. Acad. Sci.* **123**, e2509156123 (2026).

25. Gérard, A. *et al.* Detection of Rare Antigen-Presenting Cells through T Cell-Intrinsic Meandering Motility, Mediated by Myo1g. *Cell* **158**, 492–505 (2014).

26. Lei, M. *et al.* ITGB2 and ICAM3 predict increased survival of sepsis with decreased intercellular communication in cytotoxic CD8+ T cells. *Sci. Rep.* **15**, 12635 (2025).

27. Zhao, H. *et al.* Genome-wide fitness gene identification reveals Roquin as a potent suppressor of CD8 T cell expansion and anti-tumor immunity. *Cell Rep.* **37**, 110083 (2021).

28. Wang, Y. *et al.* LAPTM5 mediates immature B cell apoptosis and B cell tolerance by regulating the WWP2-PTEN-AKT pathway. *Proc. Natl. Acad. Sci.* **119**, e2205629119 (2022).

29. Ouchida, R. *et al.* A lysosomal protein negatively regulates surface T cell antigen receptor expression by promoting CD3zeta-chain degradation. *Immunity* **29**, 33–43 (2008).

30. Maus, M. *et al.* Iron accumulation drives fibrosis, senescence and the senescence-associated secretory phenotype. *Nat. Metab.* 1–20 (2023) doi:10.1038/s42255-023-00928-2.

31. Kim, M. *et al.* Accelerated increase in ferritin levels during menopausal transition as a marker of metabolic health. *Sci. Rep.* **15**, 29311 (2025).

32. Murphy, K. E., Meng, F. W., Makowski, C. E. & Murphy, P. J. Genome-wide chromatin accessibility is restricted by ANP32E. *Nat. Commun.* **11**, 5063 (2020).

33. Patwari, P. *et al.* The Arrestin Domain-Containing 3 Protein Regulates Body Mass and Energy Expenditure. *Cell Metab.* **14**, 671–683 (2011).

34. Khan, S. *et al.* B cells drive CD4 T cell immunosenescence and age-associated health decline. *Sci. Immunol.* **11**, eadv7615 (2026).

35. Huber, F. M. *et al.* Histone-binding of DPF2 mediates its repressive role in myeloid differentiation. *Proc. Natl. Acad. Sci.* **114**, 6016–6021 (2017).

36. Battistello, E. *et al.* Stepwise activities of mSWI/SNF family chromatin remodeling complexes direct T cell activation and exhaustion. *Mol. Cell* **83**, 1216-1236.e12 (2023).

37. Ji, M. *et al.* The SWI/SNF chromatin-remodeling subunit DPF2 regulates macrophage inflammation in intestinal injury via the CACNA1D-mediated MAPK pathway. *Proc. Natl. Acad. Sci.* **122**, e2518762122 (2025).

38. Jahan, A. S. *et al.* OTUB1 Is a Key Regulator of RIG-I-Dependent Immune Signaling and Is Targeted for Proteasomal Degradation by Influenza A NS1. *Cell Rep.* **30**, 1570-1584.e6 (2020).

983 39. Mulas, F. *et al*. The deubiquitinase OTUB1 augments NF-κB-dependent immune
984 responses in dendritic cells in infection and inflammation by stabilizing UBC13. *Cell*.
985 *Mol*. *Immunol*. **18**, 1512–1527 (2021).
986 40. Lee, S. G. *et al*. Non-canonical deubiquitination of OTUB1 induces IFNγ-
987 mediated cell cycle arrest via regulation of p27 stability. *Oncogene* **43**, 1852–1860
988 (2024).
989 41. Johansson, Å., Enroth, S. & Gyllensten, U. Continuous Aging of the Human DNA
990 Methylome Throughout the Human Lifespan. *PLOS ONE* **8**, e67378 (2013).
991 42. Dozmorov, M. G., Coit, P., Maksimowicz-McKinnon, K. & Sawalha, A. H. Age-
992 associated DNA methylation changes in naive CD4+ T cells suggest an evolving
993 autoimmune epigenotype in aging T cells. *Epigenomics* **9**, 429–445 (2017).
994 43. Gross, A. M. *et al*. Methylome-wide analysis of chronic HIV infection reveals five-
995 year increase in biological age and epigenetic targeting of HLA. *Mol*. *Cell* **62**, 157–168
996 (2016).
997 44. Hughes, T. *et al*. Analysis of autosomal genes reveals gene-sex interactions and
998 higher total genetic risk in men with systemic lupus erythematosus. *Ann*. *Rheum*. *Dis*.
999 **71**, 694–699 (2012).
1000 45. Harb, H. *et al*. A regulatory T cell Notch4–GDF15 axis licenses tissue
1001 inflammation in asthma. *Nat*. *Immunol*. **21**, 1359–1370 (2020).
1002 46. Harb, H. *et al*. Notch4 signaling limits regulatory T-cell-mediated tissue repair
1003 and promotes severe lung inflammation in viral infections. *Immunity* **54**, 1186-
1004 1199.e7 (2021).
1005 47. Gubbels Bupp, M. R., Potluri, T., Fink, A. L. & Klein, S. L. The Confluence of Sex
1006 Hormones and Aging on Immunity. *Front*. *Immunol*. **9**, 1269 (2018).

48. Voskuhl, R. & Itoh, Y. The X factor in neurodegeneration. *J. Exp. Med.* **219**, e20211488 (2022).

49. Goldberg, E. L. & Dixit, V. D. Drivers of age-related inflammation and strategies for healthspan extension. *Immunol. Rev.* **265**, 63–74 (2015).

50. Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune–metabolic viewpoint for age-related diseases. *Nat. Rev. Endocrinol.* **14**, 576–590 (2018).

51. Jang, I. H., Niedernhofer, L. J., Robbins, P. D. & Camell, C. D. The ageing immune system as a driver of systemic ageing. *Nat. Rev. Immunol.* 1–18 (2026) doi:10.1038/s41577-026-01269-3.

52. Franceschi, C. *et al.* Inflamm-aging: An Evolutionary Perspective on Immunosenescence. *Ann. N. Y. Acad. Sci.* **908**, 244–254 (2000).

53. Goronzy, J. J. & Weyand, C. M. Mechanisms underlying T cell ageing. *Nat. Rev. Immunol.* **19**, 573–583 (2019).

54. Nikolich-Žugich, J. The twilight of immunity: emerging concepts in aging of the immune system. *Nat. Immunol.* **19**, 10–19 (2018).

55. Franceschi, C., Olivieri, F., Moskalev, A., Ivanchenko, M. & Santoro, A. Toward precision interventions and metrics of inflammaging. *Nat. Aging* **5**, 1441–1454 (2025).

56. Almeida-Oliveira, A. *et al.* Age-related changes in natural killer cell receptors from childhood through old age. *Hum. Immunol.* **72**, 319–329 (2011).

57. Le Garff-Tavernier, M. *et al.* Human NK cells display major phenotypic and functional changes over the life span. *Aging Cell* **9**, 527–535 (2010).

58. Hetz, C. & Saxena, S. ER stress and the unfolded protein response in neurodegeneration. *Nat. Rev. Neurol.* **13**, 477–491 (2017).

59. Brehme, M. *et al.* A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease. *Cell Rep.* **9**, 1135–1150 (2014).

60. Kaushik, S. & Cuervo, A. M. Proteostasis and aging. *Nat. Med.* **21**, 1406–1415 (2015).

61. Taylor, R. C. & Dillin, A. Aging as an event of proteostasis collapse. *Cold Spring Harb. Perspect. Biol.* **3**, a004440 (2011).

62. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243–278 (2023).

63. Di Fraia, D. *et al.* Altered translation elongation contributes to key hallmarks of aging in the killifish brain. *Science* **389**, eadk3079 (2025).

64. Guldner, I. H. *et al.* Ageing promotes microglial accumulation of slow-degrading synaptic proteins. *Nature* 1–12 (2026) doi:10.1038/s41586-025-09987-9.

65. Hesterberg, R. S. *et al.* Lymphoma accelerates T cell and tissue aging. *Cancer Cell* **43**, 1917-1936.e8 (2025).

66. Wang, Y. *et al.* Proteotoxic stress response drives T cell exhaustion and immune evasion. *Nature* **647**, 1025–1035 (2025).

67. Morimoto, R. I. Cell-Nonautonomous Regulation of Proteostasis in Aging and Disease. *Cold Spring Harb. Perspect. Biol.* **12**, a034074 (2020).

68. Klaips, C. L., Jayaraj, G. G. & Hartl, F. U. Pathways of cellular proteostasis in aging and disease. *J. Cell Biol.* **217**, 51–63 (2018).

69. Wherry, E. J. & Kurachi, M. Molecular and cellular insights into T cell exhaustion. *Nat. Rev. Immunol.* **15**, 486–499 (2015).

70. Nasrallah, R. *et al.* A distal enhancer at risk locus 11q13.5 promotes suppression of colitis by Treg cells. *Nature* **583**, 447–452 (2020).

71. Gómez-Morón, Á. *et al.* Human T-cell receptor triggering requires inactivation of Lim kinase-1 by Slingshot-1 phosphatase. *Commun. Biol.* **7**, 918 (2024).
72. Botey-Bataller, J. *et al.* Genetic and molecular landscape of comorbidities in people living with HIV. *Nat. Med.* **31**, 3350–3359 (2025).
73. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, 1–20 (2013).
74. Lu, A. T. *et al.* DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
75. Lu, A. T. *et al.* DNA methylation GrimAge version 2. *Aging* **14**, 9484–9549 (2022).
76. Fuentealba, M. *et al.* A blood-based epigenetic clock for intrinsic capacity predicts mortality and is associated with clinical, immunological and lifestyle factors. *Nat. Aging* **5**, 1207–1216 (2025).
77. Sayed, N. *et al.* An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging. *Nat. Aging* **1**, 598–615 (2021).
78. Fong, S. *et al.* Principal component-based clinical aging clocks identify signatures of healthy aging and targets for clinical intervention. *Nat. Aging* **4**, 1137–1152 (2024).
79. Oh, H. S.-H. *et al.* Organ aging signatures in the plasma proteome track health and disease. *Nature* **624**, 164–172 (2023).
80. Buckley, M. T. *et al.* Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. *Nat. Aging* **3**, 121–137 (2023).
81. Li, W. *et al.* Single-cell immune aging clocks reveal inter-individual heterogeneity during infection and vaccination. *Nat. Aging* **5**, 607–621 (2025).

82. Pilley, S. E. *et al.* A metabolic atlas of mouse aging. *Cell Metab.* **38**, 399-418.e7 (2026).

83. Schaum, N. *et al.* Ageing hallmarks exhibit organ-specific temporal signatures. *Nature* **583**, 596–602 (2020).

84. Roberts, A. L. *et al.* Age acquired skewed X chromosome inactivation is associated with adverse health outcomes in humans. *eLife* **11**, e78263 (2022).

85. Abdulai-Saiku, S. *et al.* The maternal X chromosome affects cognition and brain ageing in female mice. *Nature* **638**, 152–159 (2025).

86. Furman, D. *et al.* 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 (2014).

87. Yazar, S. *et al.* The impact of sex on the immune system explored at the single-cell level. 2025.07.27.667062 Preprint at https://doi.org/10.1101/2025.07.27.667062 (2025).

88. Ren, X. *et al.* COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* **184**, 1895-1913.e19 (2021).

89. Stephenson, E. *et al.* Single-cell multi-omics analysis of the immune response in COVID-19. *Nat. Med.* **27**, 904–916 (2021).

90. Heimlich, J. B. *et al.* Multiomic profiling of human clonal hematopoiesis reveals genotype and cell-specific inflammatory pathway activation. *Blood Adv.* **8**, 3665–3678 (2024).

91. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

92. Virshup, I., Rybakov, S., Theis, F. J., Angerer, P. & Wolf, F. A. anndata: Access and store annotated data matrices. *J. Open Source Softw.* **9**, 4371 (2024).

93. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

94. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).

95. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

96. Shen, X. *et al.* Multi-omics microsampling for the profiling of lifestyle-associated changes in health. *Nat. Biomed. Eng.* **8**, 11–29 (2024).

97. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* **16**, 284–287 (2012).

98. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).

99. Fortin, J.-P., Triche, T. J. & Hansen, K. D. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinforma. Oxf. Engl.* **33**, 558–560 (2017).

100. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

101. Tian, Y. *et al.* ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* **33**, 3982–3984 (2017).

102. Xu, Z., Niu, L., Li, L. & Taylor, J. A. ENmix: a novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res.* **44**, e20 (2016).
103. Peters, T. J. *et al.* Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate. *Nucleic Acids Res.* **49**, e109 (2021).
104. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**, 2825–2830 (2011).
105. Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, New York, NY, USA, 2016). doi:10.1145/2939672.2939785.
106. Paszke, A. *et al.* PyTorch: an imperative style, high-performance deep learning library. in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* 8026–8037 (Curran Associates Inc., Red Hook, NY, USA, 2019).
107. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. in *Proceedings of the 31st International Conference on Neural Information Processing Systems* 4768–4777 (Curran Associates Inc., Red Hook, NY, USA, 2017).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.xlsx
- pbmcagingsuppfiguresv5.pdf