

Single-cell immune aging clocks reveal inter-individual heterogeneity during infection and vaccination

Received: 8 August 2024

Accepted: 24 January 2025

Published online: 5 March 2025

 Check for updates

Wenchao Li^{1,2}, Zhenhua Zhang^{1,2}, Saumya Kumar^{1,2}, Javier Botey-Bataller^{1,2,3}, Martijn Zoodma^{1,2}, Ali Ehsani^{1,2}, Qiuyao Zhan^{1,2}, Ahmed Alaswad^{1,2}, Liang Zhou^{1,2}, Inge Grondman³, Valerie Koeken^{3,4}, Jian Yang⁵, Gang Wang⁵, Sonja Volland⁶, Tania O. Crişan^{3,7}, Leo A. B. Joosten^{3,7}, Thomas Illig^{6,8,9}, Cheng-Jian Xu^{1,2}, Mihai G. Netea^{3,10} & Yang Li^{1,2,3,8,9} ✉

Aging affects human immune system functionality, increasing susceptibility to immune-mediated diseases. While gene expression programs accurately reflect immune function, their relationship with biological immune aging and health status remains unclear. Here we developed robust, cell-type-specific aging clocks (sc-ImmuAging) for the myeloid and lymphoid immune cell populations in circulation within peripheral blood mononuclear cells, using single-cell RNA-sequencing data from 1,081 healthy individuals aged from 18 to 97 years. Application of sc-ImmuAging to transcriptome data of patients with COVID-19 revealed notable age acceleration in monocytes, which decreased during recovery. Furthermore, inter-individual variations in immune aging induced by vaccination were identified, with individuals exhibiting elevated baseline interferon response genes showing age rejuvenation in CD8⁺ T cells after BCG vaccination. sc-ImmuAging provides a powerful tool for decoding immune aging dynamics, offering insights into age-related immune alterations and potential interventions to promote healthy aging.

Aging substantially impacts the human immune system, with two main functional consequences: a decrease in immune cell responsiveness, combined with increased inappropriate systemic inflammation (also called “inflammaging”). This leads to both a reduced ability to defend against external or internal challenges¹ and increased susceptibility to inflammation-driven disorders such as atherosclerosis and neurodegenerative diseases^{2,3}. While chronological age is a major indicator of biological aging, the substantial heterogeneity in the kinetics of immune aging among individuals of the same chronological age suggests that chronological time alone does not fully account for the changes observed⁴. Hence, measuring “biological age” rather than merely lifespan can more accurately and comprehensively assess the process of aging⁵, offering a unique view of age-related immune biology.

Although studies during the past decade have identified a relatively large number of age-related biomarkers, the intricate nature of biological aging presents challenges to fully uncover its underlying mechanisms. The concept of an “aging clock” emerged as a promising approach by using machine-learning methods to comprehensively capture the dynamics of aging through the integration of aging-related markers at the molecular level.

As a hallmark of aging⁶, DNA methylation-based epigenetics data were the first to be used to establish epigenetic aging clocks^{7–10}, which have been widely used to predict both lifespan and healthspan. The impact of specific diseases on epigenetic aging clocks has been also reported, such as age acceleration in coronavirus disease 2019 (COVID-19)-infected individuals¹¹ and hepatitis C virus-infected

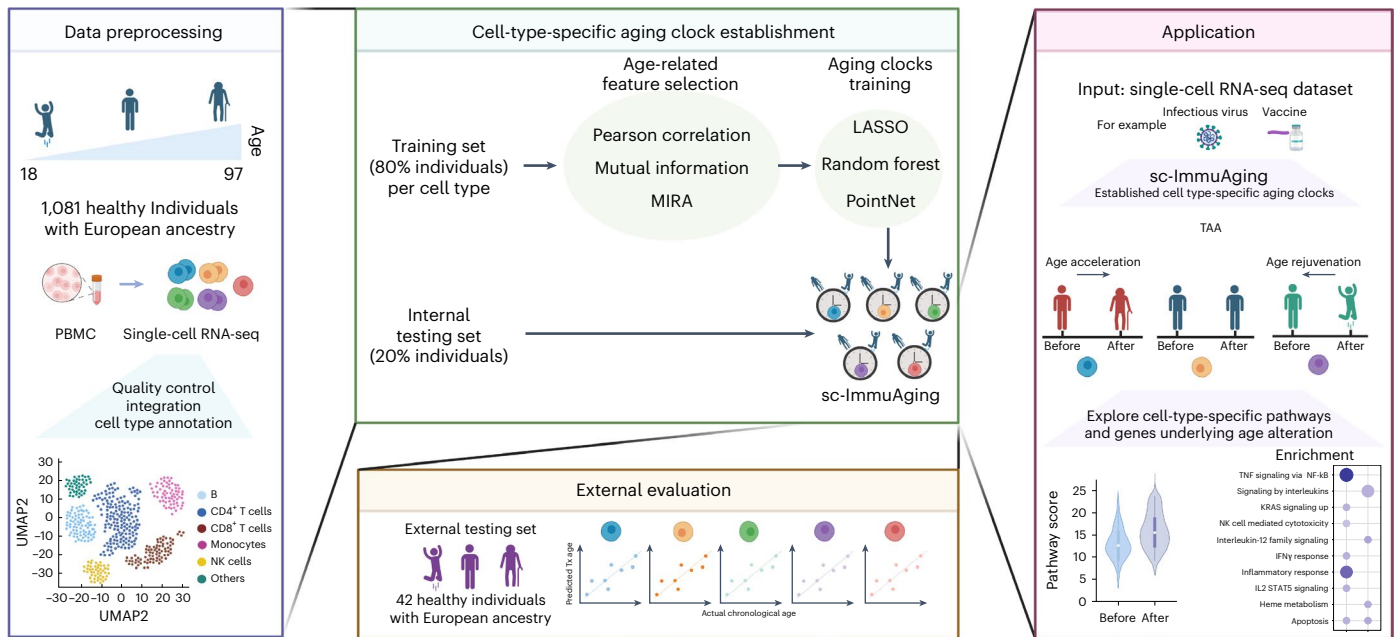


Fig. 1 | Workflow of establishing cell-type-specific aging clocks, sc-ImmuAging. The establishment and application of our aging clocks involve four main steps. (1) Data pre-processing: five publicly available scRNA-seq datasets from human PBMC are collected, including 1,081 healthy European individuals aged from 18 to 97 years. Next, quality control, integration, and cell-type annotation are conducted. In the following analysis, we focus on five major cell types, including B cells, NK cells, CD4⁺ T cells, CD8⁺ T cells and monocytes. (2) Model establishment: 80% of individuals are randomly selected for model training. Pearson correlation, mutual information and MIRA are used and compared in terms of feature selection. The selected features are trained by LASSO, random forest, and PointNet, respectively. Finally, cell-type-specific aging clocks are developed for each cell type. Subsequently, evaluation is

performed using an internal testing set, including 20% individuals. See Methods for details. (3) External validation: An independent external validation dataset of 42 healthy European individuals from five published datasets is used. Then the Pearson correlation coefficient, r.m.s.e., and m.a.e. between actual age and predicted transcriptome age (Tx age) are calculated to evaluate the accuracy of the model. (4) Application: We apply our established cell-type-specific aging clock model on vaccination and infectious diseases cohorts as case studies. We aim to examine which cell types exhibit the substantial age acceleration/rejuvenation in response to infectious diseases or vaccination. Furthermore, we explore the genes and pathways that contribute to age alterations. Fig. 1 was created with [BioRender.com](https://www.biorender.com).

individuals¹². In the field of immune biological aging, ref. 13 proposed an inflammatory aging clock (iAge) based on proteomics data and identified that the chemokine CXCL9 is an important player in age-related chronic inflammation. The immune aging score (IMM-AGE), developed by the group of Shen-Orr and colleagues, is an immune system aging clock that uses longitudinal data on inflammatory markers to estimate an individual's immune status, showing its ability to predict age-related health outcomes and mortality risks¹⁴. It is worth noting that age-related changes are also evident at the level of gene expression. One of the first transcriptomic aging clocks, proposed in 2015, used age-related marker genes identified from a meta-analysis of whole blood RNA sequencing data¹⁵. Subsequently, ref. 16 developed a machine learning-based aging clock for human fibroblasts derived from skin tissues but lacked independent datasets to evaluate the generalizability and robustness of the model. There is growing evidence that aging impacts various immune cells in a heterogeneous manner¹⁷, and single-cell RNA-sequencing (scRNA-seq) technologies provide the potential to develop cell-type-specific aging clocks. While the first single-cell transcriptome aging clock was proposed in 2023, focusing on mouse brain tissue¹⁸, research into aging in the human immune system at the single-cell resolution remained unexplored.

Given the important role of circulating immune cells in immune responses, examining their functional changes attributable to age largely improves our understanding of susceptibility to infections, cancer, and inflammatory diseases. As vaccination efficacy experiences a substantial decline with advanced age¹⁹, such knowledge would also be important to improve vaccination strategies. In this Article, we established a robust cell-type-specific aging clock, sc-ImmuAging, covering monocytes, CD4⁺ T cells, CD8⁺ T cells, natural killer (NK)

cells, and B cells, based on single-cell transcriptomic profiles of 1,081 samples of peripheral blood mononuclear cells (PBMCs) collected from European healthy adults. Both internal and external validations consistently showed the outperformance of our constructed models. We also demonstrated the value of these aging clocks to identify notable age alterations in monocytes after COVID-19 infection and to reveal the heterogeneity of age alterations in CD8⁺ T cells among individuals vaccinated with BCG (Bacillus Calmette–Guérin). Our research sheds light on understanding biological age alterations in response to vaccinations and diseases, indicating its potential application in improving personalized disease treatment and assessing age rejuvenation intervention.

Results

Single-cell transcriptome aging clocks predict biological age

We developed immune cell-type-specific transcriptome aging clocks, sc-ImmuAging, for human PBMCs (Fig. 1), using scRNA-seq datasets from five studies^{20–24}, encompassing 1,081 healthy individuals of European ancestry aged 18 to 97 years (Methods). After quality control, approximately 1.3 million high-quality cells were retained for further analysis (Methods) (Extended Data Fig. 1a,b). Focusing on the five most prevalent immune cell types in the circulation within PBMCs, CD4⁺ T cells, CD8⁺ T cells, monocytes, NK cells, and B cells, we built independent aging clocks for each using machine learning (least absolute shrinkage and selection operator (LASSO)²⁵ and random forest²⁶) and deep learning (PointNet²⁷) methods, assessing performance through various metrics.

We evaluated aging clocks performance by comparing the predicted and chronological ages (Pearson's *R*, root mean square error

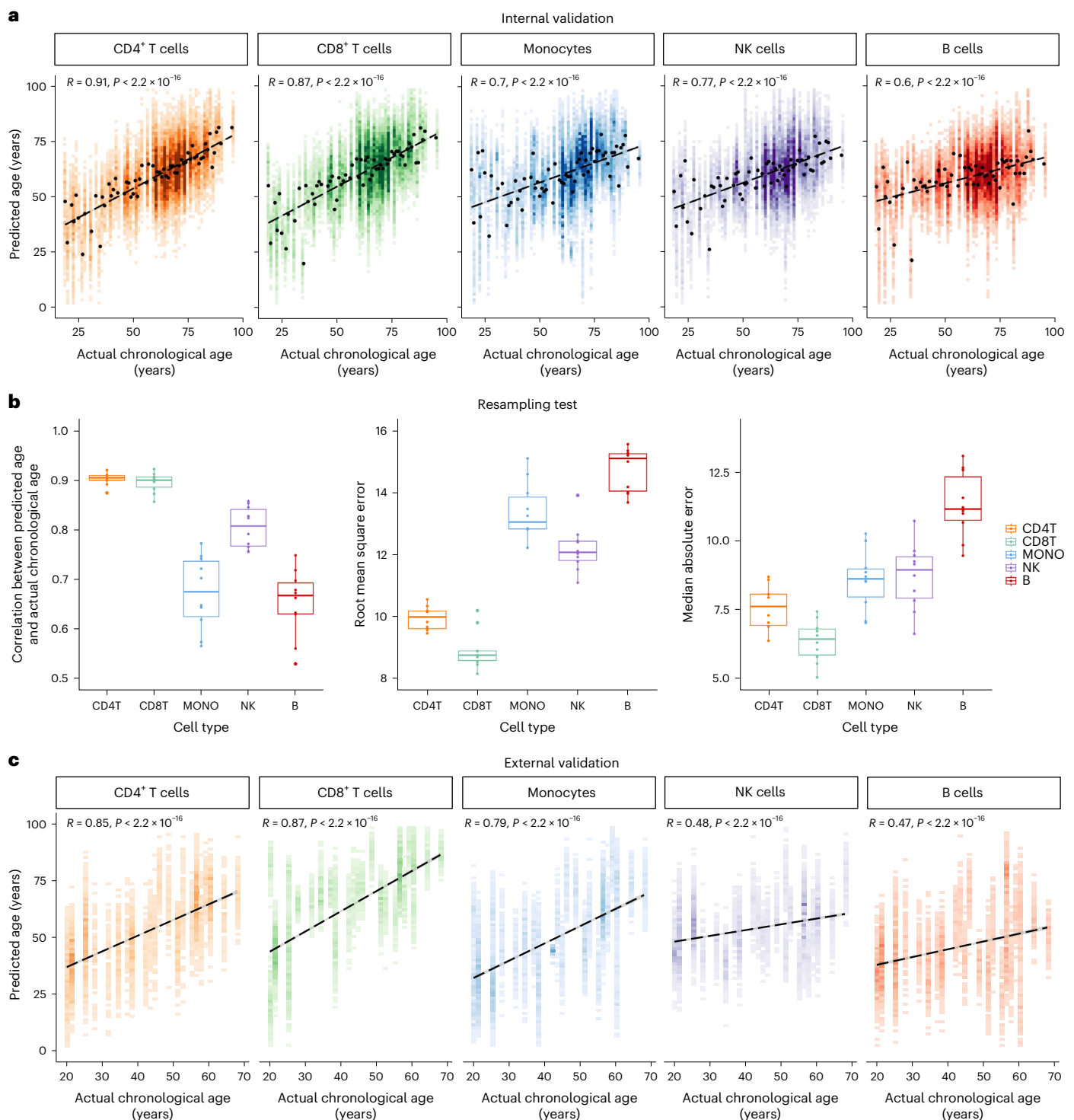


Fig. 2 | Evaluations on internal and external dataset show accuracy and robustness of aging clocks. **a**, Internal validation: Pearson correlation (two-sided test) is plotted between actual chronological age and predicted age for each cell type. Linear fit is plotted as a dashed line. **b**, To verify the robustness of the aging clock models, the training set and internal validation set are randomly resampled, and the model is trained for 10 times. Pearson correlation (two-sided

test), r.m.s.e., and m.a.e. are used as evaluation criteria. In the boxplots: center, median; box limits, upper and lower quartiles; points, resampling; whiskers, 1.5× interquartile range. CD4T, CD4⁺ T cells; CD8T, CD8⁺ T cells; MONO, monocytes; NK, NK cells; B, B cells. **c**, External validation: Pearson correlation (two-sided test) is plotted between actual chronological age and predicted age in each cell type. Linear fit is plotted as a dashed line.

(r.m.s.e.), and median absolute error (m.a.e.). A cross-validation step was implemented in which 80% (864 individuals) were randomly selected for model training, while the remaining 20% (217 individuals) constituted the internal validation set. A comprehensive evaluation of different methods used for building aging clocks is summarized

in Extended Data Table 1. LASSO regression showed superior performance across cell types. The strong correlation between chronological age and predicted transcriptomic age ($R = [0.6, 0.91]$; Fig. 2a and Extended Data Table 2) underscored the accuracy of our aging clocks. Through a 10-times repeated cross-validation analysis, we ensured the

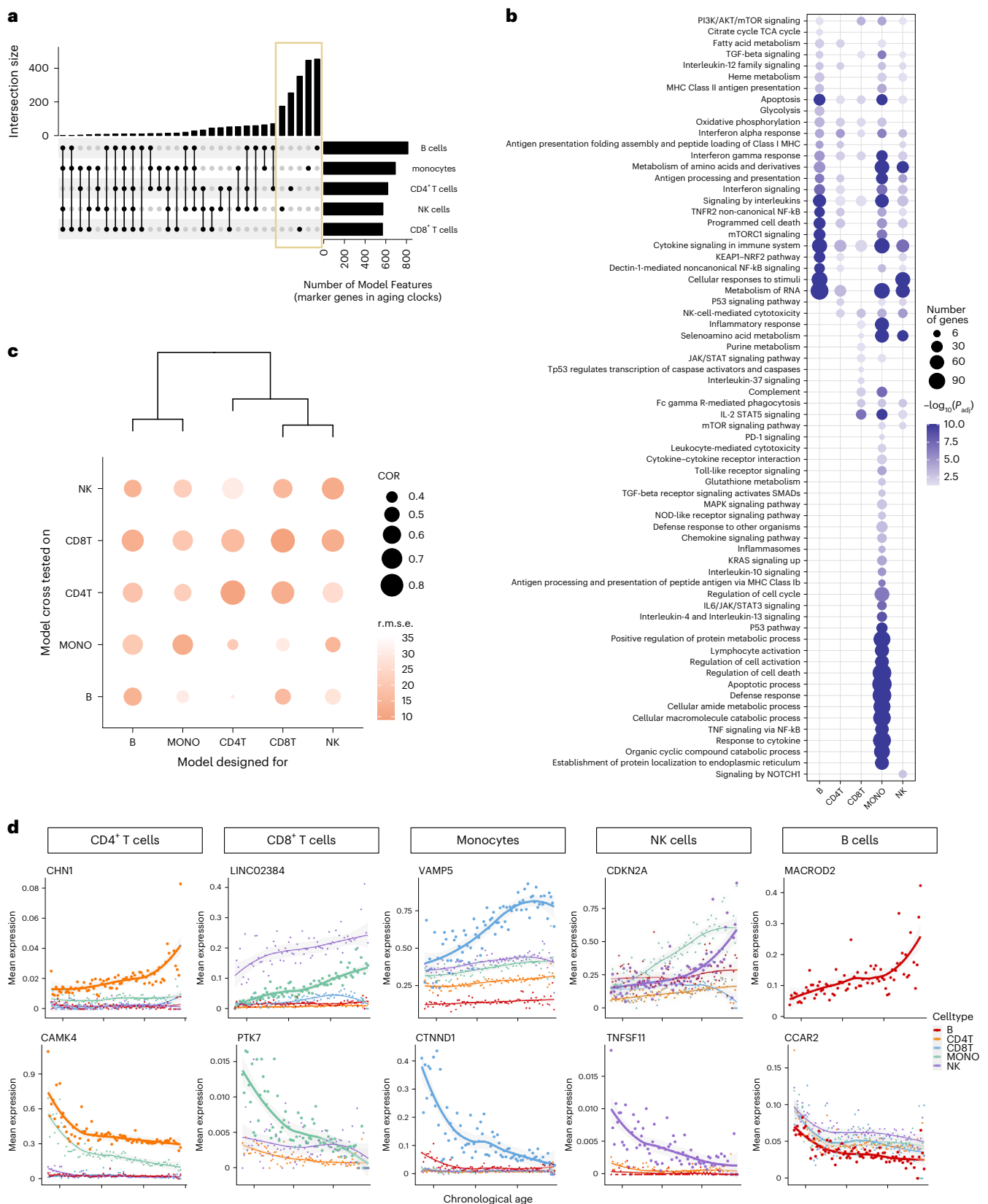


Fig. 3 | Characteristics of the established cell-type-specific aging clocks.
a, The upset plot of the intersections of the captured marker genes, with non-zero coefficients, between cell-type-specific aging clocks. **b**, Top enriched pathways of the captured marker genes in each cell-type-specific aging clock model. Pathways with Benjamini–Hochberg adjusted $P < 0.05$ are considered significant. **c**, Evaluation of the specificity of cell-type-specific aging clocks across cell types.

The x axis shows the cross tested of each model across cell types, and the y axis represents the trained model. Dot size denotes the Pearson correlation (COR) between actual chronological age and predicted age. The darkness of color represents the r.m.s.e. **d**, Trajectory of mean expression of top marker gene in each cell type. The genes on the first row have positive coefficients, and genes in the second row have negative coefficients.

robustness of our aging clock models, with subtle performance variations observed across iterations, in terms of Pearson's R , m.a.e., and r.m.s.e. (Fig. 2b). While PointNet, a deep learning approach, exhibited comparable performance in constructing cell-type-specific aging clocks (Extended Data Table 3), we proceed with the LASSO regression model due to its interpretability advantages.

To assess the generalizability of sc-ImmuAging, we tested it on independent datasets consisting of an additional 42 healthy individuals collected from five studies (healthy individuals from a thyroid cancer cohort, a sepsis cohort, an influenza cohort, a MMR (measles, mumps, and rubella vaccine) cohort, and a gout cohort)^{28–32}. Similar age distribution of both training and external validation sets is shown in Extended Data Fig. 1c. The results showed reliable predictions, confirming the efficacy of the transcriptional aging clocks (Fig. 2c and Extended Data Fig. 2a–e).

In summary, our developed single-cell transcriptome aging clocks, sc-ImmuAging, accurately predict immune aging across various ages and cell types, offering avenues for exploring both the biology of aging and the dynamics of immune aging alteration during infection and vaccination at a single-cell resolution.

Characterizing marker genes for cell-type aging clocks

For each of the cell-type-specific aging clocks constructed above, hundreds of marker genes whose expression levels correlate with age were selected as predictors of biological age (Supplementary Table 1). These aging marker genes provide insights into the molecular mechanisms underlying aging and age-related diseases.

Directly comparing selected marker genes across cell types revealed a considerable proportion unique to each cell type (Fig. 3a), while enriched pathways could still be shared (Fig. 3b). Pathway enrichment analysis using these marker genes revealed common pathways such as interferon-gamma (IFN γ) response, apoptosis, and cytokine signaling across all cell types, consistent with previous research linking inflammation to aging³³. In addition, pathways related to interleukin-12 and cytotoxicity were enriched in multiple cell types, aligning with existing literatures^{34–36}.

As the marker genes of different cell types' aging clocks share similar functions, we aimed to assess the performance of these aging clocks for predicting immune aging across various cell types. As illustrated in Fig. 3c, we observed the highest correlation and lowest r.m.s.e. between predicted and chronological age for the corresponding cell types. However, when applying an aging clock designed for one cell type to predict the age of another cell type, the performance was less accurate, indicating the specificity of the constructed aging clocks (Fig. 3c). Nevertheless, aging clocks showed similar predictability across similar immune cell types, such as CD4⁺ T cells and CD8⁺ T cells, which are consistent with previous study³⁷. We further applied a bulk aging clock¹⁵ to the pseudo-bulk data generated from our single-cell dataset and found that it performed less effectively compared to our single-cell aging clocks. These analyses underscore the importance of cell-type-specific aging clocks.

To validate our marker genes, we compared them with reported aging-related genes from human whole blood¹⁵. Among the shared marker genes, a substantial proportion exhibited the same direction across cell types, notably in CD4⁺ T cells (75%) and CD8⁺ T cells (62%) (Extended Data Table 4). Examining the gene expression trajectories of identified marker genes with age further supported our findings, revealing the expected age-dependent expression patterns for markers in their corresponding cell types (Fig. 3d). For instance, the marker gene *CAMK4* in CD4⁺ T cells, known for its role in immune response and inflammation³⁸, exhibited expression dynamics consistent with previous report¹⁵. Similarly, *MACROD2*, identified as the top marker gene for the B-cell aging clock and exhibiting an increased aging-dependent expression specifically in B cells, is recognized as a prominent feature in age-associated B cells³⁹. These results underscore the importance of

establishing cell-type-specific aging clocks and validate the accuracy of our selected marker genes.

Monocytes show aging shifts in response to COVID-19 infection

We next applied sc-ImmuAging to investigate the immune aging alterations during and after infection at a single-cell level. Using single-cell transcriptome data from four independent cohorts of patients with COVID-19, we first estimated the transcriptome age alteration (TAA) defined as the predicted age difference between patient and control at the same age (Methods). Our analysis revealed a notable transcriptome age acceleration (TAA > 0) caused by infection across immune cell types, especially in patients with severe COVID-19 (Fig. 4a). It is worth noting that monocytes showed the strongest and most consistent age acceleration across all the cohorts (Extended Data Fig. 3a–c). Given these findings, we focused on monocytes and compared TAA for age-matched individuals of control, mild, and severe conditions. Age acceleration was observed in monocytes from all patients with COVID-19, with a higher acceleration noted in patients with severe COVID-19 compared to those with mild COVID-19 across various age groups (Extended Data Fig. 3d).

To investigate the biological pathways underlying age acceleration due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection, we performed enrichment analysis of the genes that were both marker genes and differentially expressed in patients compared to control (Fig. 4b,c). We found up-regulation of TNF signaling via nuclear factor- κ B (NF- κ B) in patients with mild COVID-19, consistent with validated activation post infection^{40,41}. Patients with severe COVID-19 showed elevated circulating concentrations of pro-inflammatory cytokines⁴², which are paradoxically often accompanied by poor responsiveness to secondary stimulation, aligning with the observed up-regulation of interleukin signaling pathways. In addition, up-regulated genes in both patients with mild and patients with severe COVID-19 were all enriched in apoptosis pathways, suggesting a potential contribution of these up-regulations to aging acceleration. The down-regulated IFN signaling pathway in patients with severe COVID-19 corroborated previous reports of suppressed Type I IFN production in severe cases^{43,44}.

Furthermore, a Pearson's correlation analysis revealed a significant positive association between WHO (World Health Organization) severity score and TAA ($P = 6.66 \times 10^{-6}$; Fig. 4d). Applying sc-ImmuAging to two independent time series COVID-19 cohorts revealed a decrease in age acceleration in monocytes during the days after the symptoms started to subside (Fig. 4e). This pattern was further replicated in an independent single-nucleus dataset from a long-COVID cohort, where TAA was highest during the acute phase and decreased over time. The majority of samples exhibited age acceleration even after 90 days from the onset of infection, possibly due to the impact of long COVID on the immune system, as these patients all exhibited dysfunctional lung capacity (Fig. 4f).

In summary, our sc-ImmuAging applied to patients with COVID-19 revealed notable aging acceleration in monocytes, with further analysis identifying relevant pathways involved in COVID-19 pathogenesis. The observed decrease in age alterations over time suggested a gradual return of biological ages to healthy states.

Inter-individual variations in vaccination-induced immune age

Next, we investigated aging clock dynamics using sc-ImmuAging in immune cells from individuals before and 3 months after BCG vaccination. Our analysis revealed substantial inter-individual variations in BCG-induced TAAs across various immune cells (Fig. 5a), with individuals exhibiting either accelerated (AA) or rejuvenated (AR) TAA. It is worth noting that we found a significant negative correlation (Pearson correlation coefficient = -0.4 , $P = 0.025$; Fig. 5b) between BCG-induced TAAs in CD8⁺ T cells and vaccine efficacy. This efficacy was

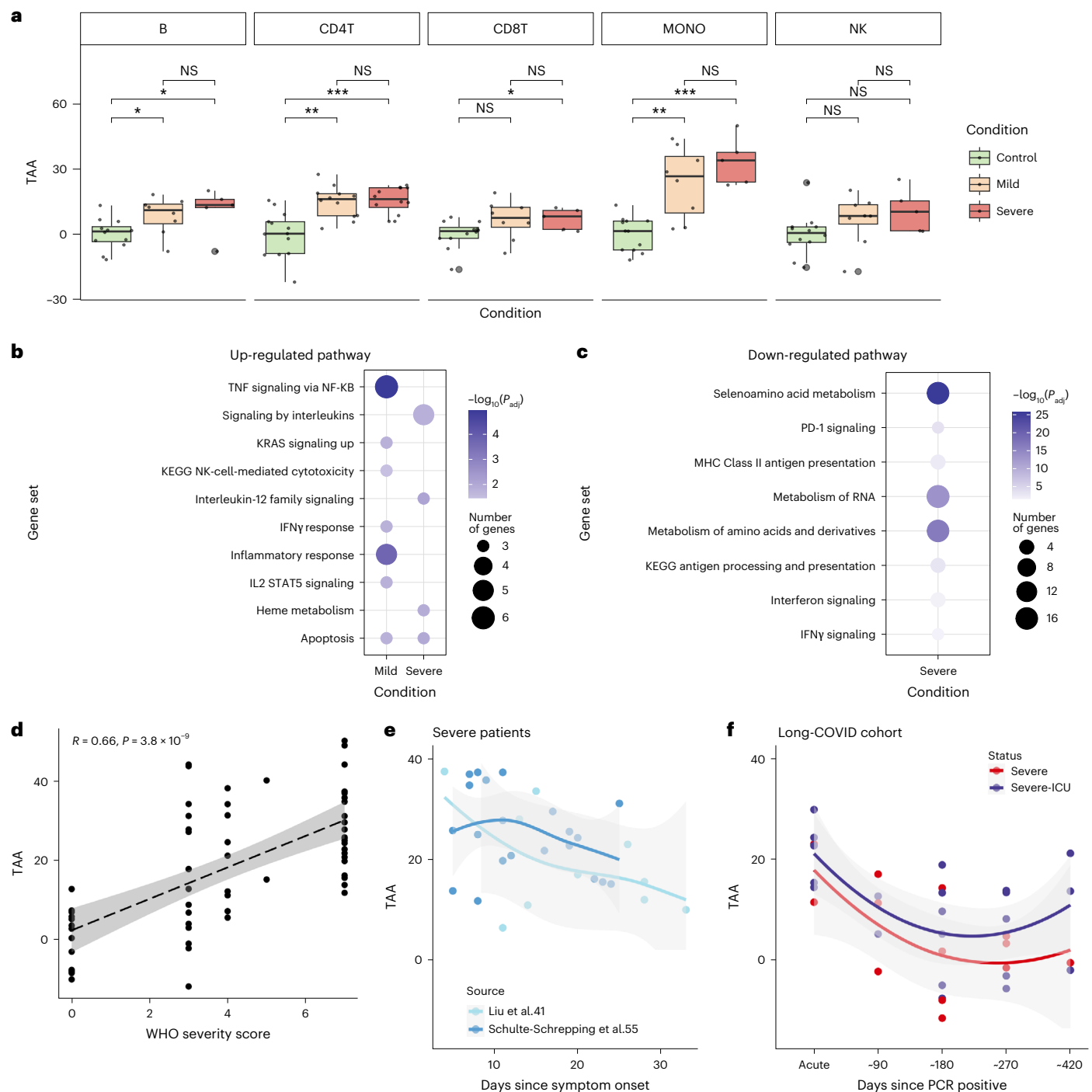


Fig. 4 | Monocytes undergo substantial age acceleration in response to COVID-19. **a**, Boxplot of the age shifts in healthy control and patients with mild and severe COVID-19 in each cell type, calculated as the predicted age subtract by the actual chronological age. See details in Methods. Boxplots: center, median; box limits, upper and lower quartiles; whiskers, 1.5 \times interquartile range. Two-sided Wilcoxon test: $P_{B:control-mild} = 0.0314, P_{B:control-severe} = 0.0485, P_{CD4T:control-mild} = 0.0011, P_{CD4T:control-severe} = 0.000449, P_{CD8T:control-severe} = 0.0348, P_{MONO:control-mild} = 0.00298, P_{MONO:control-severe} = 0.000323$; NS, not significant. **b**, Enriched pathways are calculated by the intersected genes between marker genes and up-regulated genes in patients with mild/severe COVID-19 compared to healthy control in monocytes. Pathways with Benjamini–Hochberg adjusted $P < 0.05$ are considered significant. **c**, Enriched pathways are calculated by the

intersected genes between marker genes and down-regulated genes in patients with mild/severe COVID-19 compared to healthy control in monocytes. Pathways with Benjamini–Hochberg adjusted $P < 0.05$ are considered significant. **d**, Correlation between age shifts in monocytes and WHO severity score. Dashed line represents the linearly fitted curve between WHO severity score and TAA, with the 95% confidence interval (CI). **e**, Time series age shifts of patients with severe COVID-19 from two independent published scRNA-seq datasets. The x axis is days since symptom onset, and the y axis is the age shifts. Curve is fitted using “loess” function, with the 95% CI. **f**, Time series age shifts of severe and severe ICU patients from a single-nucleus RNA-seq dataset. ICU, intensive care unit. The x axis is days since positive PCR test result, and the y axis is the age shifts. Curve is fitted using loess function, with the 95% CI.

measured by the fold change in IFN γ production upon *Mycobacterium tuberculosis* stimulation between before and after BCG vaccination, a well-established marker of BCG efficacy^{45,46}. These findings are consistent with previous research indicating the active involvement of CD8⁺ T cells in BCG-induced biological processes⁴⁷.

To investigate the biological pathways underlying BCG-induced changes in transcriptome age, we performed an enrichment analysis of genes that both served as marker genes of the transcriptomic aging clock and were differentially expressed in CD8⁺ T cells before and after vaccination (Fig. 5c). Our analysis revealed that energy-related pathways, such as oxidative phosphorylation, were down-regulated following vaccination in the group decreased TAA. This finding is consistent with previous research indicating a decrease in oxidative phosphorylation after BCG administration, suggesting a reset in cellular metabolism^{48,49}. Conversely, cytotoxicity-related pathways were up-regulated in the AA group showing increased TAA, which may suggest a contribution of these pathways to the relatively poorer vaccine responses observed in this group (Fig. 5c).

To gain a deeper understanding of the differences between the AR and AA groups after BCG vaccination, we systematically analyzed the expression dynamics within CD8⁺ T cells from each group. Our examination revealed that individuals in the AR group exhibited higher expression levels of anti-viral and IFN response genes before vaccination compared to those in the AA group (Fig. 5d). Following BCG vaccination, these genes were substantially down-regulated in the AR group, likely contributing to the observed rejuvenation in transcriptome age after vaccination. Conversely, individuals in the AA group exhibited a notable up-regulation of cytotoxicity-related genes following BCG vaccination, a pattern not observed in the AR group in which these genes remained lowly expressed. This observed pattern was further confirmed through enrichment analysis based on the differential expressed genes (Fig. 5e,f).

To further validate the higher level of cytotoxicity in the AA group, we tested whether the fold change in the proportion of CD8⁺ terminally differentiated effector memory T cells re-expressing CD45RA (TEMRA) correlates with TAA. We found a significant positive correlation ($P = 0.0074$), indicating that individuals showing age acceleration have an increased proportion of CD8⁺ TEMRA cells after BCG vaccination. This finding is consistent with the elevated expression level of cytotoxicity pathways in the CD8⁺ T cells of the AA group.

Our subsequent goal was to identify the transcription factors that potentially regulate vaccination-induced changes in TAA. Through a transcription factor enrichment analysis⁵⁰, we uncovered distinctive associations between transcription factors and IFN response genes in the AR group, highlighting the pivotal role of *IRF7* (Fig. 5g). Many of the IFN pathway genes undergoing decreased expression after BCG vaccination were under control of *IRF7*, potentially contributing to the observed age rejuvenation. By contrast, within the AA group, we observed an up-regulation of *IRF7* and its predicted targets after BCG vaccination. Further examination of the gene regulatory network in the

AA group revealed *IRF1* as the primary transcription factor governing cytotoxicity genes (Fig. 5h). It is worth noting that previous research has underscored the significance of *IRF1* in inducing T-cell cytotoxicity⁵¹. Also, we observed a down-regulation of *IRF1* and its target genes in the AR group after vaccination.

As summarized in Fig. 5i, following BCG vaccination, individuals can be stratified into AA and AR groups based on the calculated age alteration (that is, TAA) in CD8⁺ T cells using our aging clocks. In the AR group, BCG administration led to a decrease in the expression level of *IRF7*-dependent IFN response genes, while at the same time *IRF1*-dependent cytotoxicity genes were down-regulated. Conversely, the pattern observed in the AA group showed opposite directions of change, with high *IRF1*-dependent cytotoxicity genes. These findings shed light on the intricate regulatory mechanisms underlying age alterations in response to BCG administration, with *IRF7* and *IRF1* playing distinct roles in modulating IFN response and cytotoxicity, respectively, and potentially contributing to the differential outcomes observed in the TAA after vaccination.

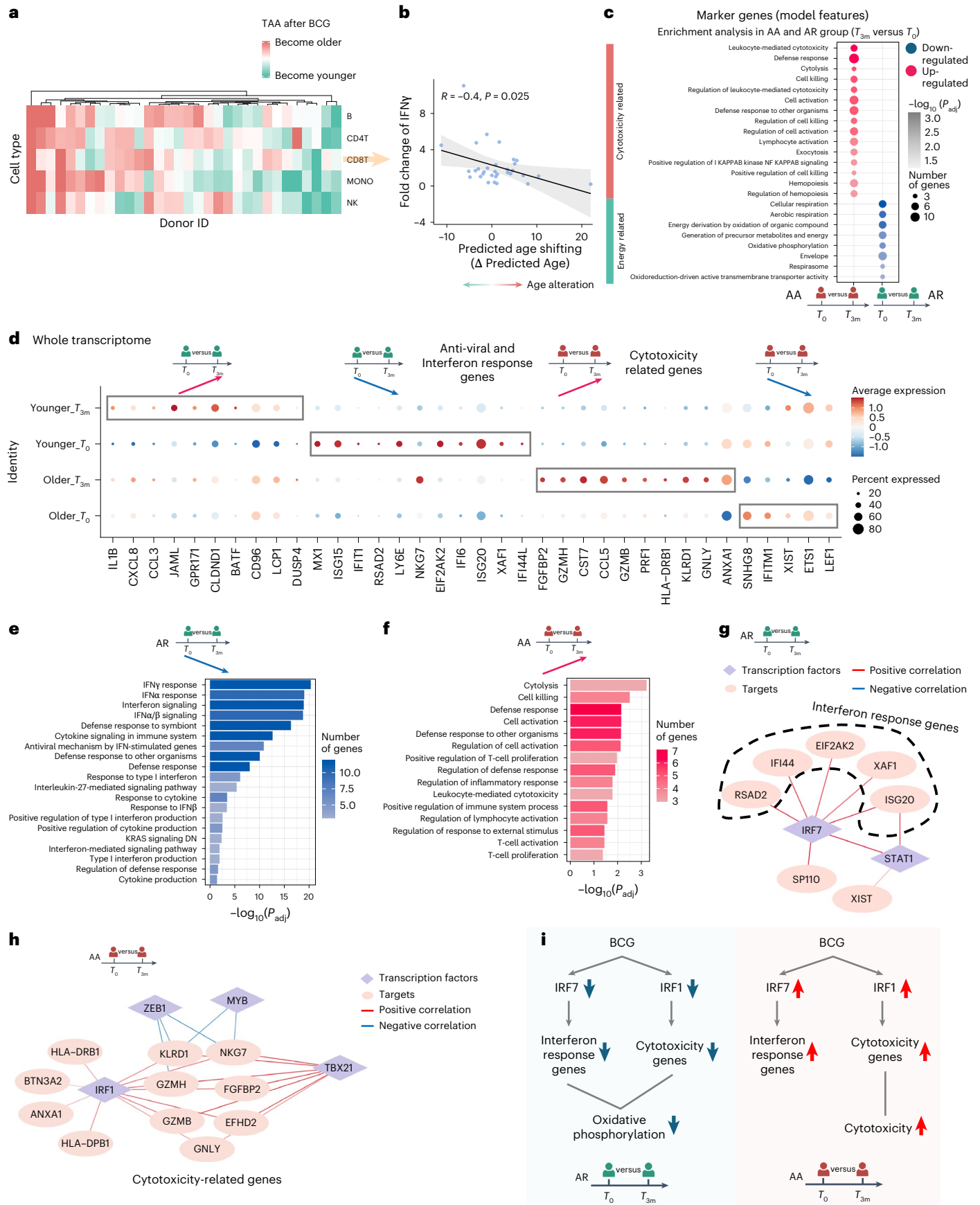
To validate these findings, we conducted a similar analysis using scRNA-seq data from nine individuals both before and 4 weeks after receiving either the MMR vaccine or a placebo (Methods). In the placebo-vaccinated group, we did not observe any notable age alterations, as expected and confirming the validity of sc-ImmuAging (Extended Data Fig. 4a). When comparing predicted ages of CD8⁺ T cells before and after MMR vaccination, we found that 3 out of 4 individuals exhibited lower TAA, equivalent with immune age rejuvenation (Extended Data Fig. 4b). It is worth noting that we observed a substantial decrease in the expression levels of IFN response gene after MMR vaccination (Extended Data Fig. 4c and Methods), which mirrors the findings observed in the BCG-induced AR group.

Discussion

In this study we collected scRNA-seq data from PBMCs of 1,081 healthy European individuals aged from 18 to 97 years and developed a cell-type-specific transcriptomic aging clock, sc-ImmuAging, to investigate the immune aging alterations in response to infection and vaccination. Through comprehensive internal and independent external validation, our proposed transcriptomics aging clocks exhibited an accurate and robust performance, enabling its potential for further application to gain a fundamental understanding of the impact of various diseases and vaccinations effect on the biological aging of the immune system. The marker genes captured by our models are cell type specific. For instance, age-related changes in *TNFSF11* were found to be specific to NK cells. It has been hypothesized that this gene may play a role in the regulatory process of apoptosis⁵². *PTK7* exhibited a unique age-dependent trajectory in CD8⁺ T cells: this gene has been proven to be associated with apoptosis in human bone marrow immune progenitors⁵³. As these marker genes presented diverse trajectories with age across different immune cell types, this suggests a model of cell-type-specific transcriptional aging in immune cells. In addition, sex

Fig. 5 | CD8⁺ T cells exhibit diverse age alterations among individuals in response to BCG vaccination. **a**, Age shifts in each cell type per individual. Red color means age acceleration after BCG vaccination, and green means the opposite. **b**, Significant correlation (Pearson correlation, $P = 0.025$, $R = 0.4$) between BCG efficacy, defined as the fold change of IFN γ between before and after vaccination with *M. tuberculosis* stimulation, and age shifts in CD8⁺ T cells. We define individuals with age rejuvenation as AR group and individuals with age acceleration as AA group. **c**, Enrichment analysis by using the genes which are marker genes and DEGs (compared between before (T_0) and after (T_m) BCG vaccination, $P < 0.05$) in AR and AA groups. **d**, Dot plot of DEGs (adjusted $P < 0.05$), comparing before and after BCG vaccination at the whole transcriptome level in AR and AA groups, respectively. **e**, Pathway enrichment analysis of down-regulated genes after BCG vaccination compared with before BCG vaccination in the AR group. Pathways with Benjamini–Hochberg adjusted $P < 0.05$ are

considered significant. **f**, Pathway enrichment analysis of up-regulated genes after BCG vaccination compared to before BCG vaccination in the AA group. Pathways with Benjamini–Hochberg adjusted $P < 0.05$ are considered significant. **g**, Unique gene regulatory network in the AR group. Targets are down-regulated after BCG vaccination. Diamonds refer to transcription factors, and circles refer to targets. The color of edges represents correlation between transcription factors and targets. Red means positive correlation, and blue means negative correlation. **h**, Unique gene regulatory network in the AA group. Targets are up-regulated after BCG vaccination. Diamonds refer to transcription factors, and circles refer to targets. The color of edges represents correlation between transcription factors and targets. Red means positive correlation, and blue means negative correlation. **i**, Schematic plot summarizing the age shifts and the involved biological processes in CD8⁺ T cells in AR group and AA group, respectively. Panels c–i were created using BioRender.com.



may markedly influence the accuracy of aging clocks, as reported by ref. 54. However, our results (not shown) indicate that overall prediction accuracy does not notably benefit from separate sex-specific analysis, likely due to reduced power from the smaller sample sizes for each sex-based subset. Future studies with larger collections of single-cell RNA-seq data for aging clock construction will allow us to refine and enhance the accuracy of sex-specific aging clocks.

To validate the functional implication of sc-ImmuAging, we extended our analysis to predict the biological age of individuals subjected to vaccination or suffering from an infection. In individuals with COVID-19, monocytes showed the most striking changes compared to other cell types, aligning with previous studies positing monocytes as key immunological players in the pathophysiology of SARS-CoV-2 infection^{55,56}. By assessing two independent cohorts, the age acceleration in monocytes, which was strongly induced in the acute phase of the infection, consistently decreased within days since symptoms appeared. This observation was independently validated when we applied our aging clock of monocytes to single-nucleus data including patients with long COVID, with the accelerated biological ages gradually returning toward normal levels. While these data show the increase of immune aging during the acute phase, with a gradual return to normal levels during recovery, these findings raise an important question: does the post-infectious TAA decrease reach the normal immune age as before infection, or does a “residual TAA” persist after recovery? Even if such a residual TAA may be relatively minor (for example, 5% of the increased TAA during the acute phase), this would have big consequences in time if an individual is prone to recurrent infections. Such a scenario is supported by a study in Bangladeshi children in a highly infectious environment, showing accelerated immune aging compared with American counterparts⁵⁷. Our sc-ImmuAging score provides an important tool to assess this hypothesis in future studies, as well as the possibility to assess the efficacy of potential interventions to slow immune aging.

Vaccinations are the most effective approach to decrease susceptibility to infections. By mimicking an infection, they induce long-term immune effects that are on the one hand protective but on the other hand could also impact immune aging. We therefore aimed to investigate such potential effects and as a model of vaccination. We explored the age alterations in response to BCG administration, one of the most used vaccines worldwide. It is also important to observe that BCG-induced transcriptional age alteration shows a notable inter-individual variation. It is worth noting that the changes in transcriptome age alteration (TAA) induced by BCG vaccination in CD8⁺ T cells negatively correlated IFN γ production upon ex vivo *M. tuberculosis* stimulation ($P = 0.025$). Using mycobacterial growth inhibition assays to evaluate *M. tuberculosis* growth control, we found no substantial correlation between phenotype from these assays and TAA before and after BCG vaccination. We used the Horvath aging clock⁵⁸ to calculate epigenetic age alteration before and after BCG vaccination. However, we did not find a notable correlation between epigenetic age alteration and BCG efficacy. This result highlights the unique insights provided by single-cell transcriptome aging clocks. We further categorized individuals who responded to BCG vaccination with either a decreased (AR) or increased (AA) TAA, and we explore the differences between these groups at the molecular level. One first important observation was a decrease in transcription of genes important for oxidative phosphorylation in the BCG-vaccinated group showing decreased TAA: this observation argues for a role of oxidative metabolism in immune aging, and it is in line with the anti-aging properties of metformin—an inhibitor of mitochondrial complex I⁵⁹. Considering the distinct transcriptome landscape between young and old individuals, in our analysis, only young individuals were considered to reduce potential biases. Future studies could include the individuals with a wider age range to evaluate the effects of vaccination on immune aging across different age groups.

To gain more insights in the molecular mechanisms responsible for changes in immune aging after vaccination, we performed a comparison of the whole transcriptome between AA and AR groups. We subsequently revealed that IFN response and cytotoxicity pathways were enriched among down-regulated genes in the AR group after BCG. Combining public databases with our scRNA-seq data, *IRF7* and *IRF1* were identified to regulate IFN response genes and cytotoxicity genes, respectively. By contrast, in individuals that experienced an increased TAA after BCG vaccination, the IFN genes under *IRF7* regulation were increased. This observation was further independently validated in a group of individuals vaccinated with MMR, suggesting that vaccines likely have consistent effects on immune aging and *IRF7/IRF1*-regulated pathways are likely to regulate these effects. An important area of future investigation will be to understand the factors that determine the heterogeneity of *IRF7/IRF1* pathways and TAA effects in different individuals after vaccination. Collectively, it was demonstrated that sc-ImmuAging can be applied in diverse PBMC datasets, with consistent results verifying the reliability of this model and the potential implementation in understanding human immune system in response to vaccine and infectious diseases.

While our prediction model exhibited an accurate prediction ($R = 0.91$) in CD4⁺ T cells, it is crucial to acknowledge certain limitations and the need for further improvement. First, the performance on external validation datasets in NK cells and B cells need to be improved by future studies. The reasons behind the less robust performance in B cells and NK cells remain unclear, and one potential approach to address it is including additional datasets for validation and exploration. To ensure the robustness and accuracy of sc-ImmuAging, we concentrated our efforts on the five main cell types with sufficient cell numbers from PBMCs. However, immune cell subpopulations exhibit distinct transcriptional profiles that play crucial roles in aging^{60,61}. Therefore, further exploration of rare immune cell populations, such as dendritic cells and $\gamma\delta$ T cells, holds great promise. Complementary technologies such as SMART-seq³⁶² provide high-resolution single-cell data, offering a valuable opportunity for future validation of single-cell transcriptome aging models. In addition, neutrophils are recognized for their involvement in inflammation⁶³ and their capacity to acquire long-term memory⁶⁴. Investigating the immune age alterations of neutrophils would be highly valuable in future studies. Another limitation is that our training, testing, and application focused exclusively on the European population from whom enough datasets were available. It is our intention to pursue the assessment across populations of African and Asian ancestry in future work. Another area of great interest in the future is to intersect transcriptomic aging clocks and the known epigenetic aging clocks. The integration between these two layers of analysis or even other omics data could yield even deeper insights into aging alteration⁶⁵. Beyond chronological age, current studies are moving forward to include clinical information as features or train the model based on age-related phenotypes. For instance, the second-generation methylation aging clock is more closely associated with mortality risk due to the inclusion of additional clinical parameters as features⁶⁶ or because it is trained on aging-related phenotypes, such as telomere length⁶⁷.

Looking ahead, our sc-ImmuAging holds promising potential for applications in personalized medicine. It can be used to evaluate the efficacy of drugs in slowing down or even reversing aging in individuals. It is worth noting that sc-ImmuAging has already shown its ability to capture inter-individual heterogeneity of age shift in response to vaccination, revealing that some individuals exhibit signs of rejuvenation while others may age more rapidly. This capability can be further explored, enabling the tailoring of interventions designed to promote rejuvenation. In addition, sc-ImmuAging can help determine the most appropriate vaccinations or drugs to administer based on the individual's unique molecular profile. For instance, recent studies⁶⁸ have indicated that metformin can decelerate the aging clock in male monkeys, showcasing a potential avenue for therapeutic application.

Furthermore, as illustrated in our COVID-19 case study, sc-ImmuAging serves as a powerful tool for identifying molecular markers within specific cell types that drive aging in the context of disease. These insights could ultimately pave the way for targeted therapeutic strategies aimed at improving age-related conditions.

Overall, our study proposed cell-type-specific transcriptomic aging clocks, sc-ImmuAging, using scRNA-seq dataset with considerable sample size from human PBMCs, with the capacity to characterize the key molecules involved in immune responses and aging, thus contributing substantially to aging studies in human immunity. In addition to the biological insights of this study, sc-ImmuAging provides a framework to investigate immune aging in various diseases and monitor future approaches aimed to slow biological aging.

Methods

Study cohorts

OneK1K cohort. The Data1 dataset²¹ was accessed from CELLxGENE (<https://cellxgene.cziscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1>), which is a cohort including 981 individuals, aged above 17 years at recruitment. Healthy individuals without any stimulations were included in our study for model training and internal validation. PBMCs were barcoded with the Single Cell 3' Library and Gel Bead Kit (10x Genomics) and sequenced by Illumina NovaSeq 2000. Cell Ranger (v2.2.0) was used to generate a gene expression matrix. The processed RDS file was downloaded, and raw counts were extracted from the Seurat object.

COVID-19-UK cohort. The Data2 dataset²² was accessed from European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) (<https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-10026>), which is a cohort including 143 individuals from London, Newcastle, and Cambridge. Only PBMCs from 11 healthy individuals aged above 17 years without any stimulations were included into model training. Included in the case study were 26 patients with mild and 15 patients with severe COVID-19. In detail, the Newcastle cohort used Chromium NextGEM Single Cell V(D)J Reagent kits v1.1. Cambridge used Single Cell V(D)J 5' version 1.1 (1000020), while London used 10 × 5' single-cell capture (Chromium Next GEM Single Cell V(D)J Reagent Kit v1.1). All three cohorts were sequenced and processed according to the 10x Genomics manufacturer's protocols. PBMCs were sequenced by Illumina NovaSeq 6000. Cell Ranger (v4.0) was used to generate a gene expression matrix. The processed h5ad file was downloaded, and raw counts were extracted from the Seurat object.

Influenza A virus (IAV) cohort. The Data3 dataset²³ was accessed from the Gene Expression Omnibus (GEO) database using accession number [GSE162632](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE162632). This cohort includes 90 male individuals aged between 21 and 69 years. Only 42 PBMCs from European healthy individuals (EUR > 0.9) without any stimulations were included in our study. Selected individuals in this cohort were used in model training and internal testing. Library preparation was done using Single Cell 3' Reagent Kits v2 User Guide (10x Genomics). PBMCs were sequenced by Illumina NovaSeq 6000. Cell Ranger (v3.0.2) was used to generate a gene expression matrix. The processed RDS file was downloaded, and raw counts were extracted from the Seurat object.

COVID-19-the Netherland cohort. The Data4 dataset²⁴ was accessed from European Genome-Phenome Archive (EGA) database using accession number [EGAS00001005529](https://www.ebi.ac.uk/ega/datasets/EGAS00001005529). This cohort includes European individuals aged above 17 years. Nine healthy individuals were selected in model training, and 15 convalescence patients were used in the case study. Library preparation was done using Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index), and PBMCs were sequenced by Illumina NovaSeq 6000. Cell Ranger (v4.0.0) was used to generate a gene expression matrix.

BCG cohort. The Data5 dataset²⁰ was accessed from the EGA database using accession number [EGAS00001006990](https://www.ebi.ac.uk/ega/datasets/EGAS00001006990). This cohort includes 38 European individuals aged above 17 years. To make results comparable, we selected 32 individuals aged between 18 and 30 years. PBMCs without any stimulations were selected in model training. BCG vaccination condition was used in the case study. Library preparation used Chromium Next GEM Single Cell 3' Library & Gel Bead Kit v3.1. PBMCs were sequenced by Illumina NovaSeq 6000. Cell Ranger (v3.1.0) was used to generate a gene expression matrix.

COVID-19-Germany cohort. The Data6 dataset⁵⁵ was accessed from FASTGenomics (https://beta.fastgenomics.org/p/schulte-schrepping_covid19), which is a cohort including 119 European individuals from Berlin and Bonn; 25 healthy controls did not have detailed age information from the Berlin cohort. In total, 19 healthy individuals and 22 patients with COVID-19 from the Bonn cohort were included in the case study. Individuals' ages were given as a range within 5 years, and therefore we calculated the mean age as the real chronological age. Library preparation was done using 10x Single Cell 30 Reagent Kit v3.1. PBMCs were sequenced in paired-end mode using Illumina NovaSeq 6000. Cell Ranger (v3.1.0) was used to generate a gene expression matrix.

Thyroid carcinoma cohort. The Data7 dataset²⁸ was accessed from the EGA database using accession number [EGAS00001005594](https://www.ebi.ac.uk/ega/datasets/EGAS00001005594). This cohort includes 36 individuals aged above 17 years. Eight healthy individuals were used in external validation. Library preparation used 10x Genomics Chromium Single-Cell 3' v2 RNA-sequencing. PBMCs were sequenced by NextSeq500-v2 150 cycle kit. Cell Ranger (v3.1.0) was used to generate a gene expression matrix.

COVID-19-MHH cohort. The Data8 dataset⁵⁶ was accessed from the EGA database using accession number [EGAD00001009331](https://www.ebi.ac.uk/ega/datasets/EGAD00001009331). This cohort includes 78 European individuals aged above 17 years old. Included in the case study were 32 convalescent patients and 46 hospitalized patients with mild or severe COVID-19. Library preparation used Next GEM Single Cell 30 Reagent Kits v3.1 (Dual Index). PBMCs were sequenced by Illumina NovaSeq 6000. Cell Ranger (v4.0.0) was used to generate a gene expression matrix.

Sepsis cohort. The Data9 dataset³¹ is a cohort including 22 European individuals. Six healthy controls were included in external validation. Library preparation used 10x Genomics Chromium Single Cell 3 v2 RNA sequencing specification. PBMCs were sequenced by NextSeq500-v2 150 cycle kit. Cell Ranger was used to generate a gene expression matrix.

Influenza cohort. The Data10 dataset²⁹ was accessed from the EGA database using accession number [EGAS00001005446](https://www.ebi.ac.uk/ega/datasets/EGAS00001005446). This cohort includes 10 healthy European individuals aged above 17 years. We only included PBMCs without any stimulations into our study as an external validation dataset. Cell Ranger (v4.0.0) was used to generate a gene expression matrix. Library preparation used Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index). PBMCs were sequenced by Illumina NovaSeq 6000.

MMR cohort. The Data11 dataset³⁰ was accessed from the EGA database using accession number [EGAS00001006787](https://www.ebi.ac.uk/ega/datasets/EGAS00001006787). This cohort includes European individuals aged above 17 years. Five placebo-vaccinated individuals were used as external validation. Four MMR-vaccinated individuals were used for BCG cohort replication. Cell Ranger (v4.0.0) was used to generate a gene expression matrix. Library preparation used Chromium Next GEM Single Cell 3' Library & Gel Bead Kit v3.1 and Chromium Next GEM Chip G Single Cell Kit (10x Genomics). PBMCs were sequenced by Illumina NovaSeq 6000.

Gout cohort. The Data12 dataset³² is a cohort including European individuals aged above 17 years. Nine healthy individuals were included for external validation. Cell Ranger (v4.0.0) was used to generate a gene expression matrix. Library preparation used Chromium Next GEM Single Cell 3' Reagent Kits v3.1-Dual Index (CG000315, 10x genomics). PBMCs were sequenced by Illumina NovaSeq 6000.

COVID-19-Yale cohort. The Data13 dataset⁴¹ was accessed from the GEO database using accession number GSE161918. This cohort includes European individuals aged above 17 years. Included in the case study were 33 SARS-CoV-2-infected patients. Cell Ranger (v3.1.0) was used to generate a gene expression matrix. Library preparation used 10 × 50 Chromium Single Cell Immune Profiling Next GEM v1.1 chemistry. PBMCs were sequenced by Illumina NovaSeq platform.

Long-COVID cohort. The Data14 dataset was accessed from the EGA database using accession number EGAS5000000142. This is a single-nuclear dataset, with European individuals aged above 17 years. PBMCs were collected from five time points: acute phase and 3, 6, 9 and 15 months after positive PCR test result. Data were used in the case study. Cell Ranger ARC (v2.0.2) was used to generate a gene expression matrix. Library preparation used 10 × Single Cell Multiome ATAC + Gene Expression chemistry. PBMCs were sequenced using Illumina NovaSeq 6000 platform.

Cell-type-specific transcriptome aging clock

In our study, the model was trained on a large dataset of 1,081 samples, using the chronological age of the samples as the ground truth. The main focus here was to establish cell-type-specific aging clocks for predicting the chronological age of various human immune cells based on scRNA-seq data. This was achieved by selecting marker genes that exhibited age-related expression changes specific to each cell type. The underlying assumption here is that the expression changes of these specific marker genes are strongly correlated with the passage of time and reflect the biological aging process. By identifying specific marker genes, the transcriptome aging clock provides insights into the molecular mechanisms underlying aging. This can help us in understanding the biological pathways driving aging and age-related diseases.

Preprocessing and analysis of published single-cell datasets

The first five cohorts were used in model training and internal testing. Raw counts from the Seurat object were extracted. Seurat (v4.0)⁶⁹ was used to create the new object. Cells with less than 200 or greater than 5,000 features and with mitochondrial percentage greater than 25% were excluded. Genes expressed in less than 3 cells were also excluded. Thereafter, the data matrix was normalized and scaled by using `NormalizeData()` and `ScaleData()` functions. `RunPCA()` was performed for dimension reduction based on the identified top 2,000 most variable genes before harmony integration⁷⁰. Subsequently, `RunNeighbors()` and `RunCluster()` were used to cluster cells based on the shared nearest-neighbor method. In summary, five datasets were integrated and clustered. Annotation was based on manually selected marker genes (B cells: *MS4A1*; CD4⁺ T cells: *IL7R*, *CD3D*; CD8⁺ T cells: *CD8A*, *CD8B*, *CD3D*; monocytes: *CD14*, *FCGR3A*, *LYZ*; NK cells: *GNLY*, *NGK7*; platelets: *PPBP*; dendritic cells: *CD74*) expression levels in each cell type⁷¹. For data visualization, we used Uniform Manifold Approximation and Projection based on the harmony data integration.

Feature selection

We first calculated the average expression of each gene in each individual by using the `AverageExpression()` function. Then we correlated each gene with real age and corrected the *P* value using Benjamini–Hochberg method and filtered out genes with adjusted *P* > 0.05. MIRA⁷² (1.0.4) is a variational auto-encoder-based model to jointly integrate scRNA-seq

data and single-cell approaches for the assay for transposase-accessible chromatin using sequencing (scATAC-seq) data. One of the advantages of using this method is it enables us to find the interpretable latent space. We used MIRA's topic model, which is a generative probabilistic model, as an unsupervised method for feature selection. MIRA found the strongly associated features with the latent topics by calculating the normalized activation of a specific gene given a topic. In addition to Pearson correlation, mutual information⁷³ was also used in feature selection. Different from the linear correlation method, mutual information could capture the non-linear relationship. Considering the non-linearity of the scRNA-seq dataset, we adopted the union set of results from Pearson correlation and mutual information. To select the informative gene lists from two methods, we adopted a widely used non-parametric method in biology, rank product⁷⁴. Simply, we permuted the ranked features 100 times and then calculated the rank product of each gene in 100 permuted matrices. Next, we counted the times *c*, in which the rank of the gene is smaller than the observed rank product. We selected those genes with adjusted $E(rp)/rank < 0.05$, where $E(rp) = c/100$; rank is the observed rank of the gene.

Specifically, these strategies included (1) examining the correlation between gene expression and chronological age, selecting significant genes based on Benjamini–Hochberg adjusted *P* < 0.05 within each cell type. To determine the optimal number of features to incorporate into the model, we input the ranked features with the step size to 100. Therefore, this approach enabled us to pinpoint the optimal hyperparameters for each specific cell type; (2) applying grid search on the gene set identified in (1) to identify the appropriate number of top features; (3) using the MIRA method for unsupervised feature selection; and (4) using the rank product to select the informative features identified by mutual information and Pearson correlation. All selected methods described below used the same feature selection strategies to ensure comparability between models.

Converting scRNA-seq into pseudo-cells

Given the sparsity of scRNA sequencing data, we applied a pseudo-cell conversion. We investigated the BootstrapCell method and Ensemble-Cell method proposed in a previous study¹⁸ to convert single cells into pseudo-cells, with both methods using LASSO regression for model training. In internal validation, the bootstrapping approach consistently outperformed the ensemble method (Extended Data Table 5). Therefore, we used the BootstrapCell method in our study. In detail, cells from one cell type one individual were down-sampling by 100 times. In each iteration, 15 cells were randomly selected, and the average value of each gene was calculated.

Model establishment and evaluation criteria

In this study, we selected LASSO regression, random forest, and PointNet to evaluate a range of methodological approaches. LASSO was chosen for its ability to identify linear associations and its stringent feature selection, which is crucial given the high correlation among genes. Random forest was included for its effectiveness in capturing nonlinear relationships and efficient parallel computation. In addition, we explored PointNet as a deep learning method that processes data as sets of points rather than relying on spatial arrangements, making it suitable for scRNA-seq data. Specifically, LASSO regression applies a penalty to the absolute value of the coefficients, effectively driving some coefficients to 0 to reduce overfitting. Furthermore, LASSO shows greater accuracy and robustness compared to ElasticNet (data not shown). For these reasons, we have selected LASSO for our downstream analysis. Fivefold cross-validation was used in the LASSO regression model, using the `glmnet`⁷⁵ package in R (v4.2.1). To balance the age distribution, we counted the number of individuals at each age n_i , $i = 18, \dots, 97$ and calculated the added $1/n_i$ as weight for each individual. We set 500 trees in the random forest model, using the `sklearn` ensemble package in python (v3.9.8). We named these methods as

LASSO, MIRA + LASSO, MIRP + LASSO, RF, MIRA + RF, and MIRP + RF. We calculated the median predicted age for each individual measured using Pearson correlation, r.m.s.e., and m.a.e. between median predicted age and the actual chronological age to estimate the accuracy of the model. Actual chronological age was defined as years since birth.

PointNet structure

We investigated a deep learning framework, PointNet²⁷, as a prediction model in our study. Specifically designed for tasks involving point cloud data within the realm of deep learning, PointNet is suited for the subset of points with unordered characteristics. Therefore, it could be potentially applied and explored in the scRNA-seq dataset. In particular, the cells can be formulated as $\{P_i | i = 1, 2, \dots, M\}$, where P_i is a vector of the gene expression profile of cell_{*i*}. To enable the feature extraction to be more robust and invariant to geometric transformations, we introduced the feature transform T-Net in our study. The structure of T-Net included three 1D convolutional layers (the numbers of filters are 32, 64, 256) with kernel size of 1 along the cell, followed by a maxpooling layer and three fully connected layers. In our PointNet model, we initially used four 1D convolutional layers, with the number of filters being 128, 64, 32, and 32. Subsequently, T-Net was used and connected by a 1D convolutional layer with 32 filters. To jointly reason the low-level features and high-level features, we further incorporated the architecture of U-Net in layers 5 through 7, which were skip-connected to the 3rd, 2nd, and 1st layers, respectively. Finally, three fully connected layers were designed to progressively aggregate the features and regress the label of cells. Except the last layer using linear activation, in each layer, BatchNormalization and rectified linear unit (ReLU) activation were used.

Description of established cell-type-specific aging clocks

Genes with nonzero coefficients were selected for downstream analysis. Genes with positive coefficients represented the higher expression level in the elderly, and vice versa. To check the specificity of the aging clock tailed for each cell type, we used the UpSet() function⁷⁶ to show the intersection and uniqueness. To clarify the contributing marker genes were specific in each cell type, we ranked the marker genes based on the absolute value of their coefficients and calculated the trajectory of marker genes over age in each cell type. We calculated the average expression level per gene per age by using the AverageExpression() function in Seurat and visualized it by ggplot2().

TAA analysis in vaccination or disease cohorts

Considering the individualized differences, for the BCG cohort, we only selected the individuals who were participants in both before and after vaccination sampling. Aging alterations induced by vaccination were compared within each individual between two time points. For infectious diseases, to keep the predicted age comparable, we selected patients and healthy controls having comparable actual chronological ages and performed the same analysis as mentioned above. For COVID-19 cohorts, four groups of individuals with real chronological ages of 38, 31, and 39 years (for healthy, mild, and severe), as well as 57, 52, and 52 years (for corresponding conditions), in addition to 63 years (representing all conditions) and 69 years (all groups) were selected. To show the aging alterations after infection, we additionally adopted the strategy proposed by ref. 77 to adjust the real actual age based on our model. In brief, we first fit a linear regression between the predicted bias and actual chronological age. Then the actual chronological age was adjusted by the fitted model.

Differentially expressed genes identification

In the 300BCG cohort, scRNA-seq data were collected at before (T_0) and three months after (T_{3m}) BCG vaccination. Differentially expressed genes (DEGs) were calculated in the following comparisons: (1) AR versus AA at both T_0 and T_{3m} , respectively, and (2) the comparison

between T_{3m} and T_0 within AR and AA, respectively. FindMarkers() function was used, and significant genes (adjusted $P < 0.05$, two-sided) were considered for downstream analysis.

The assessment of effects of marker genes in vaccination or infectious diseases cohorts

For the vaccination cohort, we calculated the fold change of the marker genes by using FindMarkers() function and selected those with P value below 0.05 (two-sided test) for downstream analysis. For the infectious disease cohort, genes were selected from the intersections between marker genes and DEGs.

Pathway enrichment analysis

We used the online database of FUMA⁷⁸ (<https://fuma.ctglab.nl>) for enrichment analysis. The databases used for enrichment analysis are the following: WikiPathways (v20191010), MsigDB (v2023.1.Hs), Kyoto Encyclopedia of Genes and Genomes (KEGG) (MsigDB c2), Reactome (MsigDB c2), Gene Ontology (GO) biological processes (MsigDB c5), and GO molecular functions (MsigDB c5). Only pathways with Benjamini–Hochberg adjusted $P < 0.05$ (two-sided test) were considered. ggplot2 package in R (v3.4.1) was used to visualize the enrichment results.

IFN response genes score in the MMR cohort

We applied our aging clocks to a MMR vaccine cohort³⁰, before vaccination and 4 weeks after vaccination. Four MMR-vaccinated individuals and five placebo-vaccinated individuals aged below 30 years were included in this study. We calculated the differential expression level of IFN response genes collected from the hallmark gene set⁷⁹. We used those genes with $P < 0.05$ (two-sided test) and calculated the gene set score for MMR-vaccinated and placebo-vaccinated groups, respectively.

Transcription factors identification in the BCG cohort

DEGs identified from comparisons between before and after vaccination in AA group or AR group were set as input to the R package: RcisTarget (v1.16.0)⁵⁰. We used hg38_refseq-r80_10kb_up_and_down_tss.mc9nr.feather as the reference database. We used those transcription factors with “directAnnotation” for downstream analysis. The R package GENIE3 (v1.18.0)⁸⁰ was adopted to reconstruct the gene regulatory network for the AA group and AR group, respectively. Based on the importance score, only the top 100 transcription factor–target links were considered, and we compared the links before and after vaccination within AA and AR groups, respectively. Only unique links that lead to age alterations were considered. Cytoscape v3.9.1⁸¹ was used for transcription factor–target network visualization.

Statistics and reproducibility

No statistical methods were used to predetermine sample sizes, but our sample sizes are similar to or larger than those reported in previous publications. We excluded individuals without age information, those aged below 18 years, and individuals not of European ancestry. The training set and internal validation set were randomly selected, and this analysis was repeated 10 times. Data distribution was assumed to be normal, but this was not formally tested. Data collection and analysis were not performed blind to the conditions of the experiments.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

OneK1K cohort data were obtained from CELLxGENE (<https://cellxgene.cziscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1>). COVID-19-UK cohort data were obtained from European Molecular Biology Laboratory’s European Bioinformatics

Institute (EMBL-EBI) (<https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-10026>). Influenza A virus (IAV) cohort data were obtained from Gene Expression Omnibus (GEO) (GSE162632). COVID-19-the Netherland cohort data were obtained from the European Genome-Phenome Archive (EGA) (EGAS00001005529). BCG cohort data were obtained from EGA (EGAS00001006990). COVID-19-Germany cohort data were obtained from FASTGenomics (https://beta.fast-genomics.org/p/schulte-schrepping_covid19). Thyroid carcinoma cohort data were obtained from EGA (EGAS00001005594). COVID-19-MHH cohort data were obtained from EGA (EGAD00001009331). Sepsis cohort data were obtained from ref. 31. Influenza cohort data were obtained from EGA (EGAS00001005446). MMR cohort data were obtained from EGA (EGAS00001006787). Gout cohort data were obtained from ref. 32. COVID-19-Yale cohort data were obtained from GEO (GSE161918). Long-COVID cohort data were obtained from EGA (EGAS0000000142). The databases used for enrichment analysis are the following: WikiPathways (v20191010), MsigDB (v2023.1.Hs), Kyoto Encyclopedia of Genes and Genomes (KEGG) (MsigDB c2), Reactome (MsigDB c2), Gene Ontology (GO) biological processes (MsigDB c5), and GO molecular functions (MsigDB c5).

Code availability

The codebase and R package for the established aging clocks, sc-ImmuAging, is available via GitHub at <https://github.com/CiiM-Bioinformatics-group/scImmuAging>.

References

- Albright, J. F. & Albright, J. W. *Aging, Immunity, and Infection* (Humana, 2003).
- Head, T., Daunert, S. & Goldschmidt-Clermont, P. J. The aging risk and atherosclerosis: a fresh look at arterial homeostasis. *Front. Genet.* **8**, 216 (2017).
- Hou, Y. et al. Ageing as a risk factor for neurodegenerative disease. *Nat. Rev. Neurol.* **15**, 565–581 (2019).
- Jazwinski, S. M. & Kim, S. Examination of the dimensions of biological age. *Front. Genet.* **10**, 263 (2019).
- Han, J.-D. J. The ticking of aging clocks. *Trends Endocrinol. Metab.* **35**, 11–22 (2024).
- López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
- Xu, C.-J. et al. The emerging landscape of dynamic DNA methylation in early childhood. *BMC Genomics* **18**, 25 (2017).
- Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
- Zhang, Y. et al. DNA methylation signatures in peripheral blood strongly predict all-cause mortality. *Nat. Commun.* **8**, 14617 (2017).
- Belsky, D. W. et al. Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *Elife* **9**, e54870 (2020).
- Cao, X. et al. Accelerated biological aging in COVID-19 patients. *Nat. Commun.* **13**, 2135 (2022).
- Oltmanns, C. et al. Reverse inflammaging: long-term effects of HCV cure on biological age. *J. Hepatol.* **78**, 90–98 (2023).
- 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).
- Alpert, A. et al. A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nat. Med.* **25**, 487–495 (2019).
- Peters, M. J. et al. The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* **6**, 8570 (2015).
- Fleischer, J. G. et al. Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biol.* **19**, 221 (2018).
- Nikolich-Zugich, J. The twilight of immunity: emerging concepts in aging of the immune system. *Nat. Immunol.* **19**, 10–19 (2018).
- 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).
- Allen, J. C., Toapanta, F. R., Chen, W. & Tennant, S. M. Understanding immunosenescence and its impact on vaccination of older adults. *Vaccine* **38**, 8264–8272 (2020).
- Li, W. et al. A single-cell view on host immune transcriptional response to in vivo BCG-induced trained immunity. *Cell Rep.* **42**, 112487 (2023).
- Yazar, S. et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**, eabf3041 (2022).
- Stephenson, E. et al. Single-cell multi-omics analysis of the immune response in COVID-19. *Nat. Med.* **27**, 904–916 (2021).
- Randolph, H. E. et al. Genetic ancestry effects on the response to viral infection are pervasive but cell type specific. *Science* **374**, 1127–1133 (2021).
- Liu, Z. et al. Multi-omics integration reveals only minor long-term molecular and functional sequelae in immune cells of individuals recovered from COVID-19. *Front. Immunol.* **13**, 838132 (2022).
- Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. Ser. B* **58**, 267–288 (1996).
- Ho, T. K. Random decision forests. In *Proc. 3rd International Conference on Document Analysis and Recognition* vol. 1, 278–282 (IEEE, 1995).
- Charles, R. Q., Su, H., Kaichun, M. & Guibas, L. J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 77–85 (IEEE, 2017).
- Rabold, K. et al. Reprogramming of myeloid cells and their progenitors in patients with non-medullary thyroid carcinoma. *Nat. Commun.* **13**, 6149 (2022).
- Debisarun, P. A. et al. Induction of trained immunity by influenza vaccination - impact on COVID-19. *PLoS Pathog.* **17**, e1009928 (2021).
- Röring, R. J. et al. MMR vaccination induces trained immunity via functional and metabolic reprogramming of $\gamma\delta$ T cells. *J. Clin. Invest.* **134**, e170848 (2024).
- Grondman, I. et al. Single-cell transcriptomics identifies different immune signatures between macrophage activation-like syndrome and immune paralysis in sepsis. Preprint at *medRxiv* <https://doi.org/10.1101/2023.03.17.23287390> (2023).
- Alaswad, A. et al. Integrative analysis reveals the multilateral inflammatory mechanisms of CD14 monocytes in gout. *Ann. Rheum. Dis.* (in the press).
- de Almeida, A. J. P. O., Ribeiro, T. P. & de Medeiros, I. A. Aging: molecular pathways and implications on the cardiovascular system. *Oxid. Med. Cell. Longev.* **2017**, 7941563 (2017).
- Itazawa, T. et al. Developmental changes in interleukin-12-producing ability by monocytes and their relevance to allergic diseases. *Clin. Exp. Allergy* **33**, 525–530 (2003).
- Castle, S. C. et al. Age-related impaired proliferation of peripheral blood mononuclear cells is associated with an increase in both IL-10 and IL-12. *Exp. Gerontol.* **34**, 243–252 (1999).
- Ponnappan, S. & Ponnappan, U. Aging and immune function: molecular mechanisms to interventions. *Antioxid. Redox Signal.* **14**, 1551–1585 (2011).
- Yu, Y., Xu, T., Yu, Y., Hao, P. & Li, X. Association of tissue lineage and gene expression: conservatively and differentially expressed genes define common and special functions of tissues. *BMC Bioinformatics* **11**, S1 (2010).
- Illario, M. et al. Calmodulin-dependent kinase IV links Toll-like receptor 4 signaling with survival pathway of activated dendritic cells. *Blood* **111**, 723–731 (2008).

39. Domínguez Conde, C. et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* **376**, eabl5197 (2022).
40. Kircheis, R. et al. NF- κ B pathway as a potential target for treatment of critical stage COVID-19 patients. *Front. Immunol.* **11**, 598444 (2020).
41. Liu, C. et al. Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. *Cell* **184**, 1836–1857.e22 (2021).
42. Chen, G. et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *J. Clin. Invest.* **130**, 2620–2629 (2020).
43. Akamatsu, M. A., de Castro, J. T., Takano, C. Y. & Ho, P. L. Off balance: interferons in COVID-19 lung infections. *EBioMedicine* **73**, 103642 (2021).
44. Hadjadj, J. et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).
45. Flynn, J. L. et al. An essential role for interferon gamma in resistance to *Mycobacterium tuberculosis* infection. *J. Exp. Med.* **178**, 2249–2254 (1993).
46. Roche, P. W., Triccas, J. A. & Winter, N. BCG vaccination against tuberculosis: past disappointments and future hopes. *Trends Microbiol.* **3**, 397–401 (1995).
47. Boer, M. C., van Meijgaarden, K. E., Joosten, S. A. & Ottenhoff, T. H. M. CD8⁺ regulatory T cells, and not CD4⁺ T cells, dominate suppressive phenotype and function after in vitro live *Mycobacterium bovis*-BCG activation of human cells. *PLoS ONE* **9**, e94192 (2014).
48. Kührtreiber, W. M. et al. Long-term reduction in hyperglycemia in advanced type 1 diabetes: the value of induced aerobic glycolysis with BCG vaccinations. *npj Vaccines* **3**, 23 (2018).
49. Faustman, D. L. Benefits of BCG-induced metabolic switch from oxidative phosphorylation to aerobic glycolysis in autoimmune and nervous system diseases. *J. Intern. Med.* **288**, 641–650 (2020).
50. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
51. Duncan, G. S., Mittrücker, H. W., Kägi, D., Matsuyama, T. & Mak, T. W. The transcription factor interferon regulatory factor-1 is essential for natural killer cell function in vivo. *J. Exp. Med.* **184**, 2043–2048 (1996).
52. Parodi, M. et al. Hypoxia modifies the transcriptome of human NK cells, modulates their immunoregulatory profile, and influences NK cell subset migration. *Front. Immunol.* **9**, 2358 (2018).
53. Prebet, T. et al. The cell polarity PTK7 receptor acts as a modulator of the chemotherapeutic response in acute myeloid leukemia and impairs clinical outcome. *Blood* **116**, 2315–2323 (2010).
54. Horvath, S. et al. An epigenetic clock analysis of race/ethnicity, sex, and coronary heart disease. *Genome Biol.* **17**, 171 (2016).
55. Schulte-Schrepping, J. et al. Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* **182**, 1419–1440.e23 (2020).
56. Zhang, B. et al. Altered and allele-specific open chromatin landscape reveals epigenetic and genetic regulators of innate immunity in COVID-19. *Cell Genom.* **3**, 100232 (2023).
57. Wagar, L. E. et al. Increased T cell differentiation and cytolytic function in Bangladeshi compared to American children. *Front. Immunol.* **10**, 2239 (2019).
58. Horvath, S. Erratum to: DNA methylation age of human tissues and cell types. *Genome Biol.* **16**, 96 (2015).
59. Liu, J.-K. Antiaging agents: safe interventions to slow aging and healthy life span extension. *Nat. Prod. Bioprospect.* **12**, 18 (2022).
60. Mittelbrunn, M. & Kroemer, G. Hallmarks of T cell aging. *Nat. Immunol.* **22**, 687–698 (2021).
61. Hao, Y., O'Neill, P., Naradikian, M. S., Scholz, J. L. & Cancro, M. P. A B-cell subset uniquely responsive to innate stimuli accumulates in aged mice. *Blood* **118**, 1294–1304 (2011).
62. Hagemann-Jensen, M. et al. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol.* **38**, 708–714 (2020).
63. Greenlee-Wacker, M. C. Clearance of apoptotic neutrophils and resolution of inflammation. *Immunol. Rev.* **273**, 357–370 (2016).
64. Netea, M. G. et al. Trained immunity: a program of innate immune memory in health and disease. *Science* **352**, aaf1098 (2016).
65. Rutledge, J., Oh, H. & Wyss-Coray, T. Measuring biological age using omics data. *Nat. Rev. Genet.* **23**, 715–727 (2022).
66. Lu, A. T. et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging* **11**, 303–327 (2019).
67. Lu, A. T. et al. DNA methylation-based estimator of telomere length. *Aging* **11**, 5895–5923 (2019).
68. Yang, Y. et al. Metformin decelerates aging clock in male monkeys. *Cell* 10.1016/j.cell.2024.08.021 (2024).
69. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
70. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
71. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
72. Lynch, A. W. et al. MIRA: joint regulatory modeling of multimodal expression and chromatin accessibility in single cells. *Nat. Methods* **19**, 1097–1108 (2022).
73. Meyer, P. E. *Information-Theoretic Variable Selection And Network Inference From Microarray Data*. (Universite Libre de Bruxelles, 2008).
74. Breitling, R., Armengaud, P., Amtmann, A. & Herzyk, P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett.* **573**, 83–92 (2004).
75. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
76. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
77. 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).
78. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
79. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
80. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5**, e12776 (2010).
81. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

Acknowledgements

This work was supported with funds provided by European Research Council starting grant 948207, Netherlands Organisation for Scientific Research (NWO) ASPASIA and Radboud University Medical Center Hypatia grant. The COVID-19 Research Network of the State of Lower Saxony (COFONI) was funded by the Ministry of Science and Culture of Lower Saxony, Germany (14-76403-184-3), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2155 (390874280), and the Lower Saxony Center for AI and Causal Methods in Medicine (CAIMed) grant to Y.L.; European Research Council advanced grant 833247 and Spinoza grant of the Netherlands Organization for

Scientific research to M.G.N.; and Lower Saxony Ministry of Science and Culture (MWK) Sprung grant to C.-J.X. (19777006). MWK COVID-19 biobank of Hannover Medical School (MHH) was funded by the Lower Saxony Ministry of Science and Culture (MWK; to T.I. and S.V.) China Scholarship Council, and a PhD scholarship to W.L.

Author contributions

Y.L. developed the concept of the work. W.L. contributed to the design and implementation of the method and the analysis of computational experiments. W.L., S.K., J.B.-B., M.Z., M.G.N., and Y.L. contributed to the interpretation of the results. L.Z. and A.A. contributed to the performance of the wet lab experiments of the long-COVID data. S.K. contributed to the preprocessing of the long-COVID data. T.O.C. and A.A. contributed to the performance of the wet lab experiments of the gout data. A.A. contributed to the preprocessing of the gout data. Z.Z., A.E., and Q.Z. contributed to the validation of the model. Y.L. and W.L. drafted the initial version of the manuscript. All the authors contributed to the revision of the work.

Funding

Open access funding provided by Helmholtz-Zentrum für Infektionsforschung GmbH (HZI).

Competing interests

M.G.N. is a scientific founder of Lemba, TTXD, and Biotrip. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s43587-025-00819-z>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43587-025-00819-z>.

Correspondence and requests for materials should be addressed to Yang Li.

Peer review information *Nature Aging* thanks David Furman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

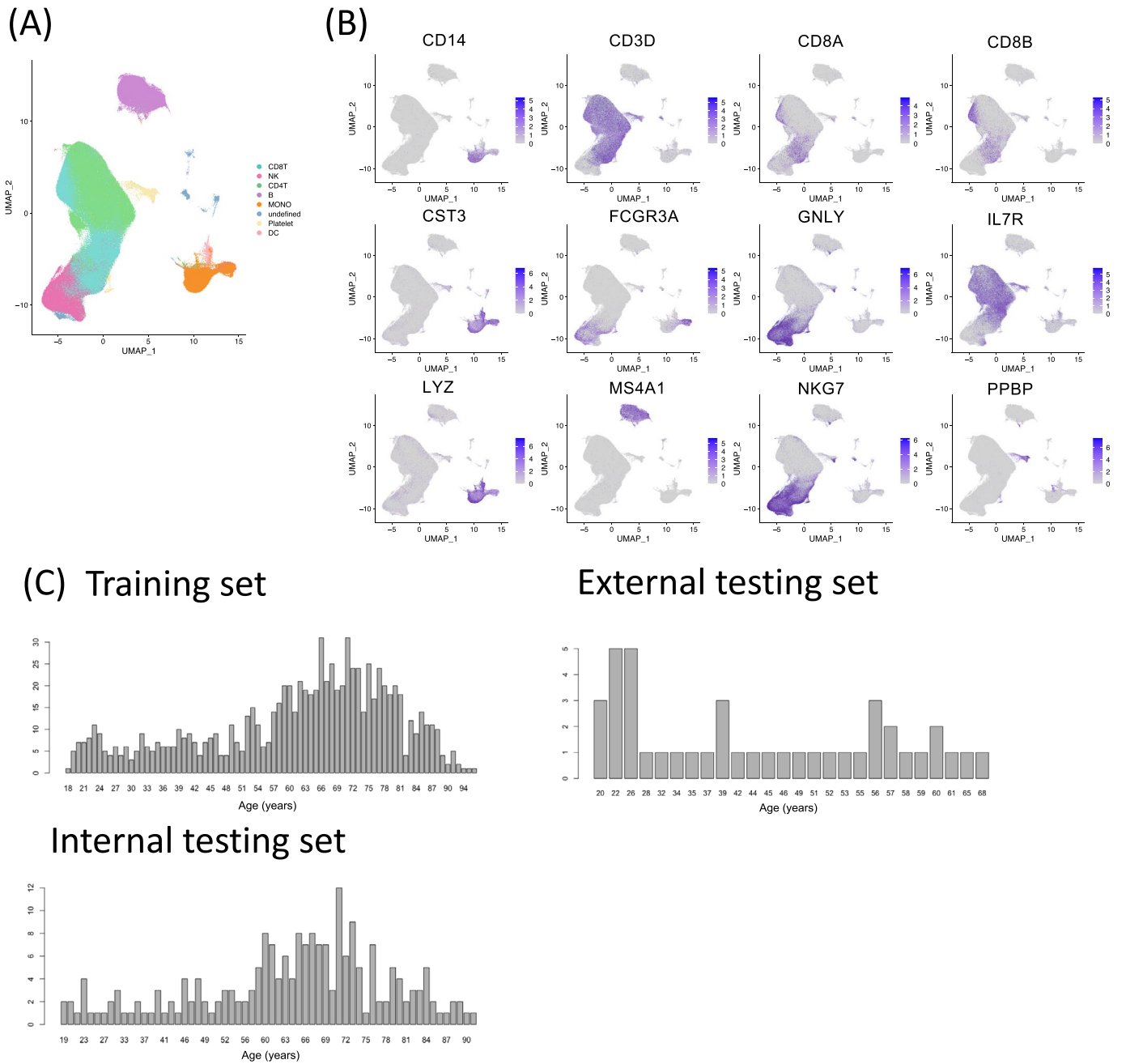
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

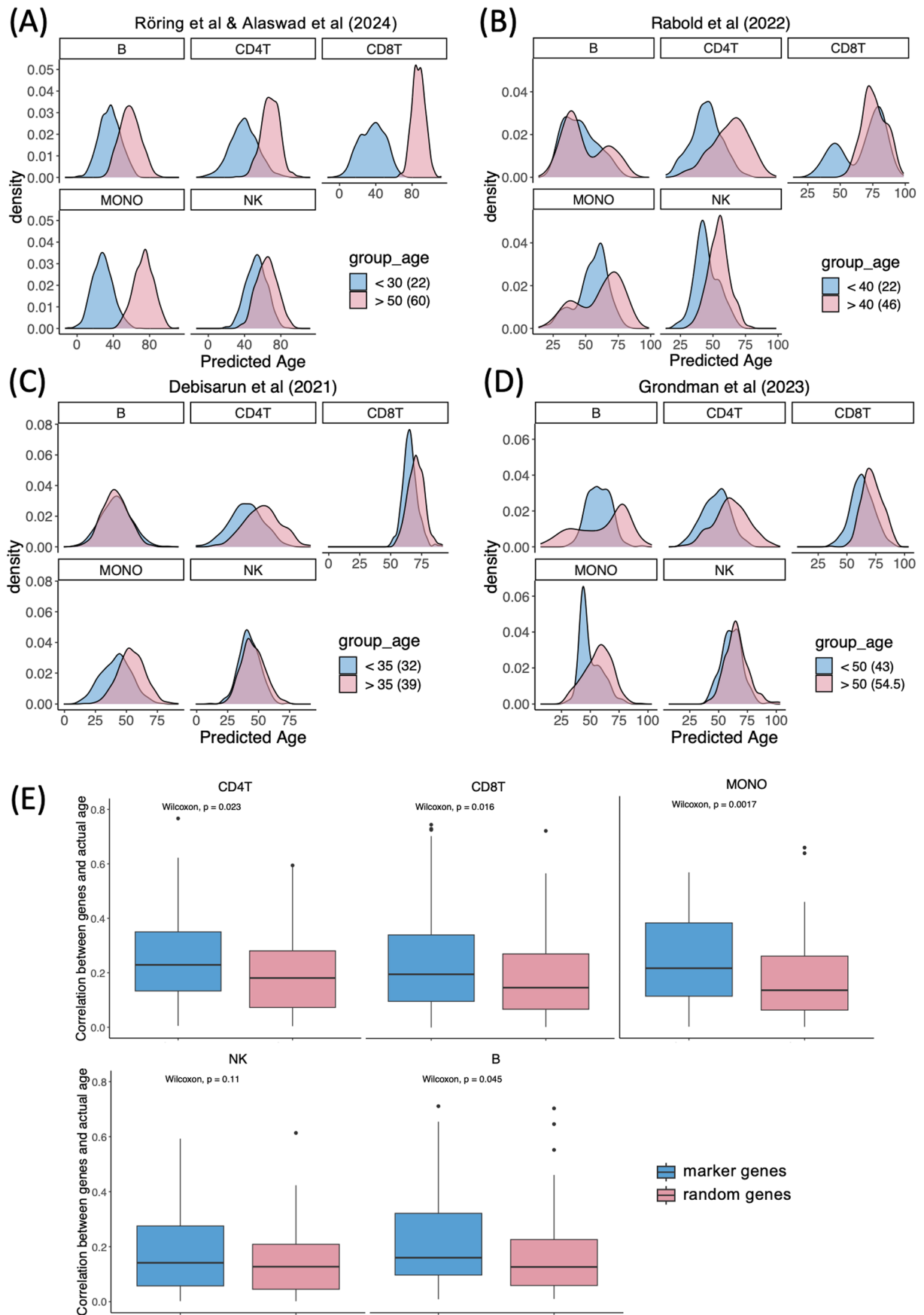
¹Department of Computational Biology of Individualised Medicine, Centre for Individualised Infection Medicine (CiiM), a Joint Venture Between the Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany. ²TWINCORE, Centre for Experimental and Clinical Infection Research, a Joint Venture Between the Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany.

³Department of Internal Medicine and Radboud Center for Infectious Diseases, Radboud University Medical Center, Nijmegen, The Netherlands.

⁴Research Centre Innovations in Care, Rotterdam University of Applied Sciences, Rotterdam, The Netherlands. ⁵The National Clinical Research Center for Mental Disorders and Beijing Key Laboratory of Mental Disorders, Beijing Anding Hospital, Capital Medical University, Beijing, China. ⁶Hannover Unified Biobank, Hannover Medical School, Hannover, Germany. ⁷Department of Medical Genetics, Iuliu Hațieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania. ⁸Cluster of Excellence RESIST (EXC 2155), Hannover Medical School, Hannover, Germany. ⁹Lower Saxony center for artificial intelligence and causal methods in medicine (CAIMed), Hannover, Germany. ¹⁰Department for Immunology and Metabolism, Life and Medical Sciences Institute (LIMES), University of Bonn, Bonn, Germany. ✉e-mail: Yang.Li@helmholtz-hzi.de



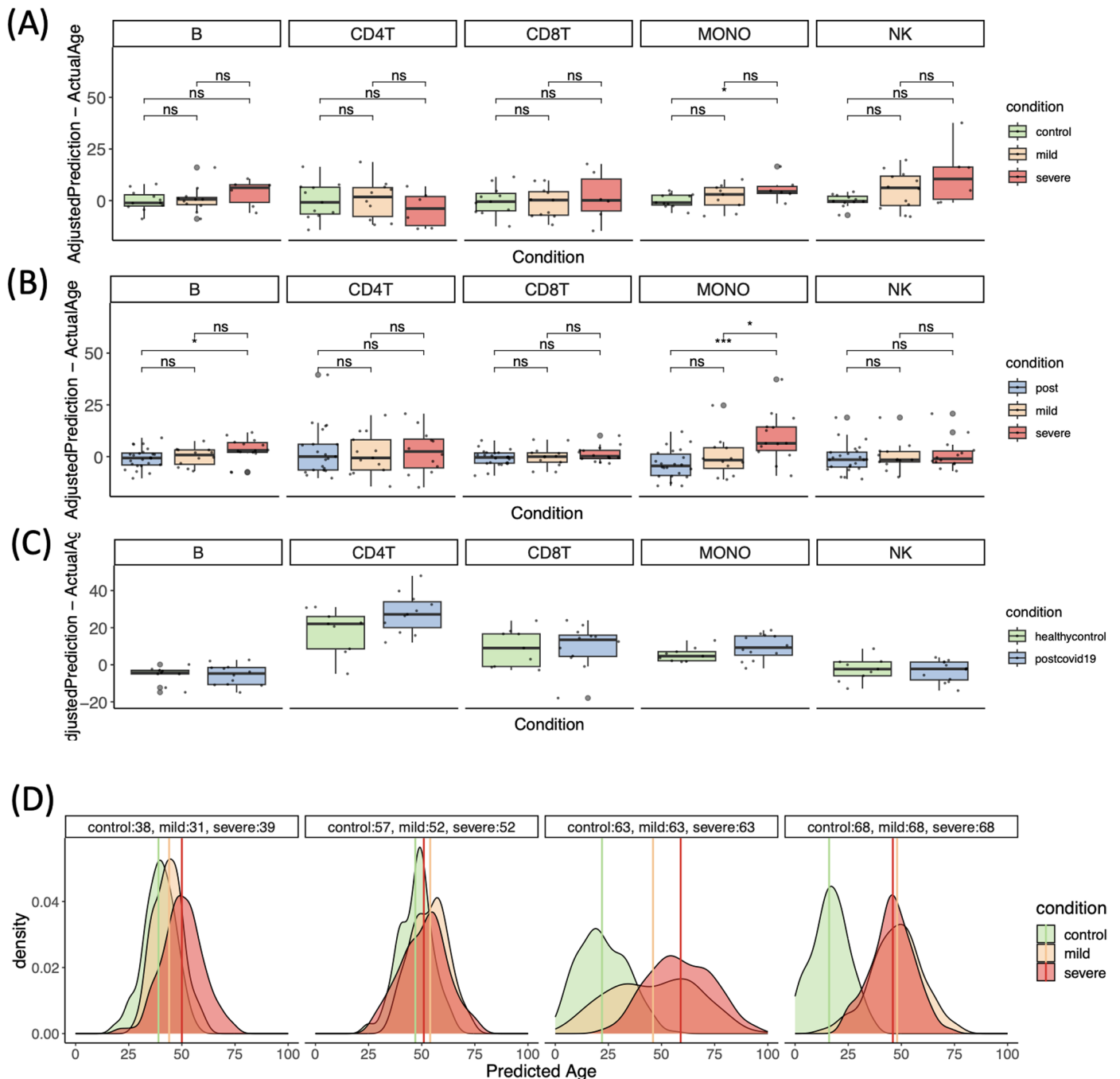
Extended Data Fig. 1 | Overview of collected scRNA-seq datasets. (a) UMAP of the scRNA-seq data. **(b)** Expression of marker genes for cell type annotations. **(c)** Age distributions for training set, internal testing set, and external testing set, respectively.



Extended Data Fig. 2 | See next page for caption.

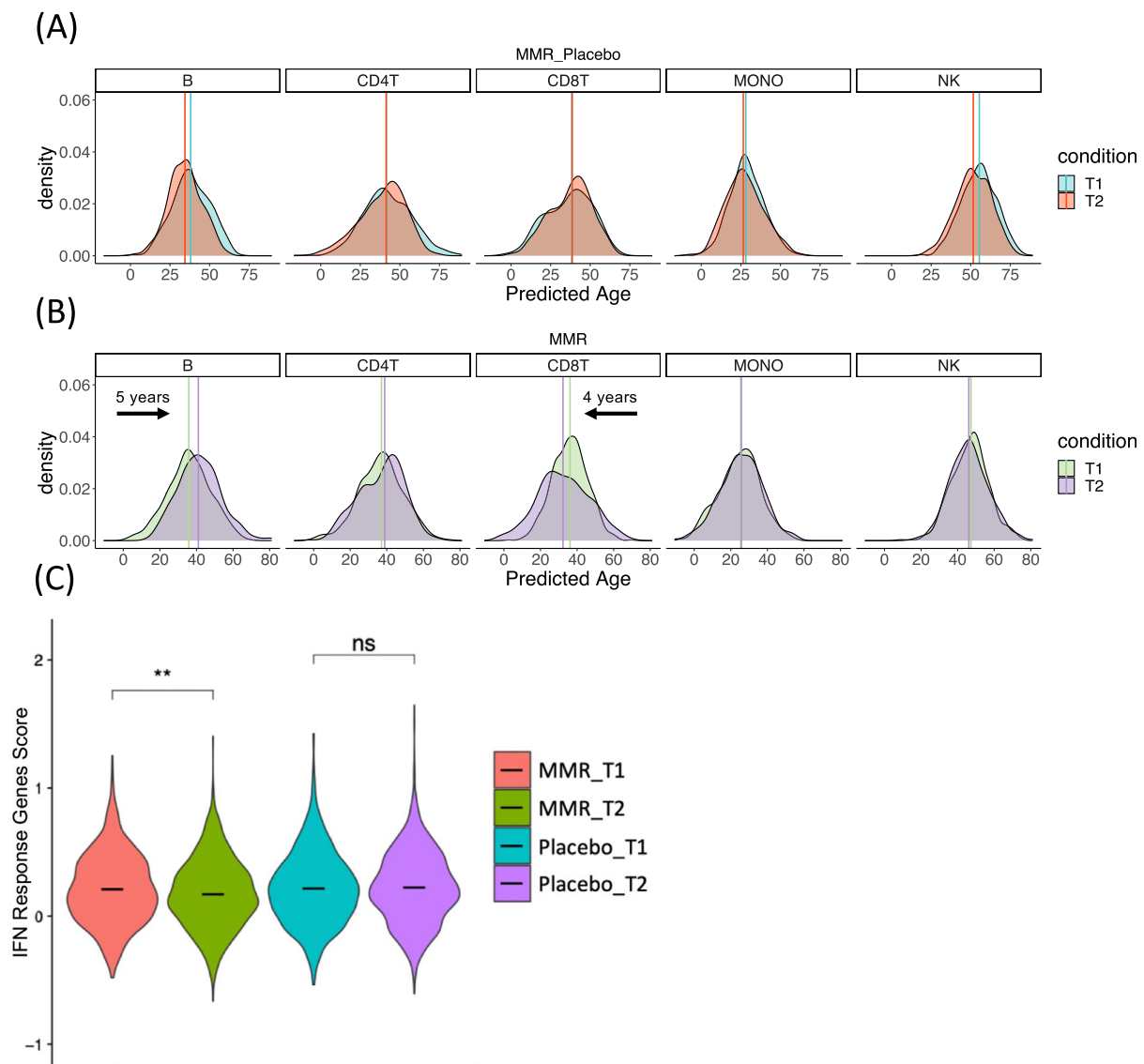
Extended Data Fig. 2 | External validation performance. (a) External validation on Röring et al.³⁰ dataset, with average chronological age of 22 years, and Alaswad et al.³² dataset, with average chronological age of 60 years. Density plots of predicted age for these two datasets are compared. (b) Individuals from Rabold et al.²⁸ dataset were categorized into those under 40 years old (average actual chronological age = 22, blue density plot) and over 40 years old (average actual chronological age = 46, red density plot), based on age distribution of the samples. Density plots represent the predicted age for each group. (c) Individuals from Debisarun et al.²⁹ dataset are split into those under 35 years old (average actual chronological age = 32, blue density plot) and above 35 years old (average actual chronological age = 39, red density plot). Density plots represent the

predicted age for each group. (d) Individuals from Grondman et al.³¹ dataset are grouped into those under 50 years old (average actual chronological age = 43, blue density plot) and over 50 years (average actual chronological age = 54.5, red density plot). Density plots represent the predicted age for each group. (e) Boxplot of comparison between top 100 marker genes (ranked by absolute coefficient) of each aging clock and randomly selected genes using training set ($n = 864$). The y axis is the $-\log_{10}(P \text{ value})$ (two-sided wilcoxon test). Blue color means genes captured by the model and red color means random genes. boxplots: centre, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range.



Extended Data Fig. 3 | Application of sc-ImmuAging on the independent COVID-19 cohorts. (a) Boxplot of TAA for the Stephenson et al.²² cohort (n = 26). boxplots: centre, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range. *two-sided wilcoxon test*: $P_{MONO:control-severe} = 0.0202$; ns, not significant. (b) Boxplot of TAA for Zhang et al.⁵⁶ cohort (n = 48). boxplots: centre, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range.

two-sided wilcoxon test: $P_{MONO:post-severe} = 0.000336$, $P_{MONO:mild-severe} = 0.0107$; ns, not significant. (c) Boxplot of TAA for Liu et al.²⁴ cohort (n = 24). boxplots: centre, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range. (d) Density plot of the predicted age for healthy control, mild and severe patients. The median value of each distribution is highlighted as a vertical line. Four groups of individuals with different age stages are plotted.



Extended Data Fig. 4 | Application of sc-ImmuAging on the MMR cohort³⁰.

(a) Density Plot of the predicted age of the Placebo-vaccinated group at before and after administration. Vertical line represents the median predicted age of each group. No differences are detected between two timepoints, suggesting the prediction accuracy of our model. **(b)** Density Plot of the predicted age of the MMR-vaccinated group at before and after administration. Vertical line represents the median predicted age of each group. We observe five years age

acceleration in B cells, whereas four years age rejuvenation in CD8 + T cells. **(c)** Violin plot of the Interferon response gene score (Methods) for the MMR- and Placebo-vaccinated group at two time points. The y axis is the interferon gene score. We observe a notable decrease in the MMR-vaccinated group after administration, which is consistent with the result from the BCG-vaccinated group. No significant changes are observed in the Placebo-vaccinated group. *two-sided wilcoxon test*: $P_{MMR:T1-T2} = 0.0019$; ns, not significant.

Extended Data Table 1 | Methods comparison

		Gridsearch + LASSO		LASSO		MIRA + LASSO		MI_COR+LASSO		RF		MIRA + RF		MI_COR + RF	
		Correlation	RMSE	Correlation	RMSE	Correlation	RMSE	Correlation	RMSE	Correlation	RMSE	Correlation	RMSE	Correlation	RMSE
B	Internal	0.6	14.08	0.69	13.02	0.69	13.43	0.63	14.05	0.59	13.94	0.56	14.32	0.5	14.74
	External	0.47	15.07	0.034	20.15	0.012	19.51	0.2	19.07	0.35	20.59	0.36	19.09	0.63	16.04
NK	Internal	0.77	11.98	0.73	12.17	0.75	12.16	0.71	12.95	0.65	13.46	0.6	14.03	0.55	14.47
	External	0.48	18.02	-0.016	18.01	-0.38	20.57	0.43	16.41	0.11	20.88	-0.059	20.08	0.25	15.97
CD8T	Internal	0.87	9.47	0.88	9.42	0.86	9.56	0.83	9.98	0.82	10.79	0.81	10.91	0.73	12.28
	External	0.87	21.28	0.82	23.38	0.79	24.47	0.84	11.8	0.38	24.59	0.64	19.26	0.52	13.19
CD4T	Internal	0.91	9.22	0.91	8.68	0.88	10.06	0.84	11.13	0.75	12.54	0.7	13.14	0.65	13.83
	External	0.85	13.18	0.87	13.84	0.68	11.59	0.9	12.19	0.55	17.64	0.46	15.54	0.6	12.69
MONO	Internal	0.73	11.79	0.72	12.03	0.75	11.86	0.7	12.29	0.52	14.6	0.51	14.77	0.57	14.22
	External	0.55	13.4	-0.2	27.48	0.22	15.72	0.81	10.45	-0.37	21.5	-0.16	20.18	-0.16	19.59

Extended Data Table 2 | Internal and external validation performance

	Dataset	Correlation	MedianAbsError	MeanAbsError	RMSE
B	Internal	0.6	9.71	11.7	14.08
	External	0.47	13.54	13.13	15.07
NK	Internal	0.77	9	9.79	11.98
	External	0.48	12.42	14.79	18.02
MONO	Internal	0.7	5.65	8.78	12.29
	External	0.79	9.4	9.99	11.65
CD8T	Internal	0.87	7.24	7.97	9.47
	External	0.87	20.21	19.44	21.28
CD4T	Internal	0.91	7.61	7.71	9.22
	External	0.85	9.99	11.18	13.18

Extended Data Table 3 | PointNet evaluation

	Method	Correlation	MedianAbsError	MeanAbsError	RMSE
B	ours	0.6	9.71	11.7	14.08
	Pointnet	0.65	6.3	10.06	13.55
NK	ours	0.77	9	9.79	11.98
	Pointnet	0.73	7.83	9.01	12.14
MONO	ours	0.7	5.65	8.78	12.29
	Pointnet	0.74	5.7	8.6	12.44
CD8T	ours	0.87	7.24	7.97	9.47
	Pointnet	0.84	7.1	7.81	9.75
CD4T	ours	0.91	7.61	7.71	9.22
	Pointnet	0.91	4.8	6.25	8.02

Extended Data Table 4 | Replication with public data

	proportion of overlapped features	with same direction	correlation from the paper	proportion with same direction
B	0.12	47	NA	0.5
NK	0.1	35	NA	0.5
MONO	0.16	130	0.78	0.62
CD8T	0.37	89	0.67	0.75
CD4T	0.19	60	0.42	0.54

Extended Data Table 5 | PseudoCell Method evaluation

	Bootstrap Correlation	Ensemble Correlation
B	0.69	0.57
NK	0.73	0.63
CD8T	0.88	0.75
CD4T	0.91	0.7
MONO	0.72	0.55

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	R 4.2.1, python 3.9.8
Data analysis	R packages: Seurat 4.0, FUMA 1.5.2, ggplot2 3.4.1, readr 2.1.3, tidyverse 1.3.2, glmnet 4.1.4, ggpubr 0.4.0, biomaRt 2.52.0, infotheo 1.2.0.1, purrr 0.3.4, ggridges 0.5.4, ComplexHeatmap 2.12.1, RcisTarget 1.16.0, GENIE3 1.18.0; python packages: pandas 1.4.4, numpy 1.19.5, tensorflow 2.5.3, keras 2.5.0, scipy 1.9.3, sklearn 0.23.2, scanpy 1.9.13, mira 1.0.4, Biorender code: https://github.com/CiiM-Bioinformatics-group/sclmmuAging

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

OneK1K cohort data from CELLxGENE (<https://cellxgene.cziscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1>), COVID-19-UK cohort data from

EMBL-EBI (<https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-10026>), IAV cohort data from Gene Expression Omnibus (GEO, GSE162632), COVID-19-the Netherland cohort data from European Genome-Phenome Archive (EGA, EGAS00001005529), BCG cohort data from EGA (EGAS00001006990), COVID-19-Germany cohort data from FASTGenomics (https://beta.fastgenomics.org/p/schulte-schrepping_covid19), Thyroid Carcinoma cohort data from EGA (EGAS00001005594), COVID-19-MHH cohort data from EGA (EGAD00001009331), Sepsis cohort data from Grondman et al.31, Influenza cohort data from EGA (EGAS00001005446), MMR cohort data from EGA (EGAS00001006787), Gout cohort data from Alaswad et al.32, COVID-19-Yale cohort data from GEO (GSE161918), Long COVID cohort data from EGA (EGAS0000000142). Database for enrichment analysis: WikiPathways (v20191010), MsigDB (v2023.1.Hs), KEGG (MsigDB c2), Reactome (MsigDB c2), GO biological processes (MsigDB c5), GO molecular functions (MsigDB c5).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The cell type-specific aging clocks for human PBMC are established based on 1.3 million cells from scRNA-seq datasets. We only included individuals aged above 18 years old with European ancestry. And the data should provide detailed age information. In total, our study comprises 1,081 individuals. We randomly split them into training set (80%) and internal testing set (20%). Additionally, we also established an independent external dataset including 42 individuals to further validate the accuracy of our model. To reduce bias, we only focus on main cell types (monocytes, NK cells, B cells, CD8+ T cells and CD4+ T cells) with sufficient cell counts in PBMC, ensuring robust results for each cell type. To our knowledge, the sample size of our study is comparable to or exceeds that of other aging clock studies. Therefore, this dataset is sufficient to provide reliable and accurate results.
Data exclusions	We included all the healthy individuals aged above 18 years old for model training and internal/external testing. COVID-19 patients or vaccinated individuals are used as case study.
Replication	No new data was generated for this study. The code is available on GitHub: https://github.com/CiiM-Bioinformatics-group/sclmmuAging .
Randomization	The training and internal validation samples are randomly assigned. The external dataset is independent with training/internal testing dataset.
Blinding	All datasets used in this study were publicly available. The training data were collected from five independent studies, and the external data were obtained from another five independent studies. The training set and internal validation were randomly assigned. Additionally, we repeated this analysis 10 times. Data collection and analysis were not performed blind to the conditions of the experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

- | n/a | Involvement |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Plants |

- | n/a | Involvement |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.