RESEARCH

Biologically informed machine learning modeling of immune cells to reveal physiological and pathological aging process

Cangang Zhang[1,2,3†], Tao Ren[4,5†], Xiaofan Zhao[6†], Yanhong Su[1,2,3†], Qianhao Wang[1,2,3], Tianzhe Zhang[1,2,3], Boxiao He[7], Yabing Chen[8], Ling-Yun Wu[4,5], Lina Sun[1,2,3*], Baojun Zhang[1,2,3*] and Zheng Xia[6,9,10*]

Abstract

The immune system undergoes progressive functional remodeling from neonatal stages to old age. Therefore, understanding how aging shapes immune cell function is vital for precise treatment of patients at different life stages. Here, we constructed the first transcriptomic atlas of immune cells encompassing human lifespan, ranging from newborns to supercentenarians, and comprehensively examined gene expression signatures involving cell signaling, metabolism, differentiation, and functions in all cell types to investigate immune aging changes. By comparing immune cell composition among different age groups, HLA highly expressing NK cells and CD83 positive B cells were identified with high percentages exclusively in the teenager (Tg) group, whereas unknown_T cells were exclusively enriched in the supercentenarian (Sc) group. Notably, we found that the biological age (BA) of pediatric COVID-19 patients with multisystem inflammatory syndrome accelerated aging according to their chronological age (CA). Besides, we proved that inflammatory shift- myeloid abundance and signature correlate with the progression of complications in Kawasaki disease (KD). The shift- myeloid signature was also found to be associated with KD treatment resistance, and effective therapies improve treatment outcomes by reducing this signaling. Finally, based on those age-related immune cell compositions, we developed a novel BA prediction model PHARE (https://xiazlab.org/phare/), which can apply to both scRNA-seq and bulk RNA-seq data. Using this model, we found patients with coronary artery disease (CAD) also exhibit accelerated aging compared to healthy individuals. Overall, our study revealed changes in immune cell proportions and function associated with aging, both in health and disease, and provided a novel tool for successfully capturing features that accelerate or delay aging.

Keywords Immune aging, Biological age, Chronological age, Aging clock, Machine learning

†Cangang Zhang, Tao Ren, Xiaofan Zhao and Yanhong Su contributed equally to this work.

*Correspondence:
Lina Sun
lina.sun@xjtu.edu.cn
Baojun Zhang
bj.zhang@mail.xjtu.edu.cn
Zheng Xia
xiaz@ohsu.edu
Full list of author information is available at the end of the article

Introduction

Advanced age is often linked to increased morbidity and mortality from infectious diseases, and decreased vaccination efficacy [1, 2]. Aging is also a leading risk factor for the increased incidence of most cancer types [3, 4]. Although most children develop mild and self-limiting symptoms of infectious diseases, such as COVID-19, a severe and delayed post-SARS-CoV-2 inflammatory response in children has been recognized worldwide, correlating with disease severity [5, 6]. Moreover, compared with older individuals, children

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 2 of 21

appear to have a more robust immune response during acute infections, whereas young cancer patients treated with immune checkpoint inhibitors often have poor outcomes [7, 8]. The immune system exhibits highly age-specific features; however, the impact of age-related changes on different components of the immune system is not fully understood [2]. Therefore, the characterization of immune populations and function among different age groups could provide valuable insights into mechanisms underlying disease development and inform more precise disease treatments in the future.

Not everyone ages at the same rate. Usually, aging affects health and varies from person to person; therefore, it is not surprising that people with the same CA manifest diverse aging-related phenotypes [9]. In this regard, a person's BA can often differ from his/her CA. Arrojo et al. revealed that most organs, even in one person, are also a mix of cells and proteins of vastly different ages, which depend on their rates of regeneration [10]. The ability to accurately measure human aging from molecular profiles has practical implications in many fields, particularly in disease prevention and treatment [11]. As expected, previous studies have developed many BA measurements that attempt to capture physiological changes during the aging process, such as DNA methylation, telomere length, and frailty [9, 11, 12]. Meanwhile, there are tools have been developed for BA prediction based on blood transcriptome, given the ease of obtaining samples [13, 14]. Considering that most immune diseases present strong age characteristics, the functional states of immune cells do not always match their CA in these diseases [15]. Therefore, effective measurements to comprehensively depict the age distribution of immune cells to accurately assess the impact of various diseases on the aging process are urgently needed.

Single-cell RNA sequencing (scRNA-seq) is a powerful technology used for studying individual cells and delineating complex cell populations. Recently, scRNA-seq has been performed in several studies to profile the immune landscape of human peripheral blood mononuclear cells (PBMCs) at different ages [16–18]. However, most of these studies have primarily focused on certain age periods, rather than the entire lifespan from 0 to over 110 years old, and many studies only analyzed a single type of immune cell, such as T cells [19, 20], B cells [21], or myeloid cells [22, 23]. Owing to the complexity of age spans and cell populations, the previous studies mentioned above limit our understanding of how immune profiles contribute to disease development from a systematic perspective. Thus, for the first time, we systematically analyzed all immune cells and their transcriptomic signatures in human PBMCs across different age groups encompassing the entire lifespan to establish an elaborate and aging-focused immune landscape.

In the current study, we first integrated three public scRNA-seq datasets from 24 healthy individuals across different age groups. Through extensive analysis, we revealed dynamic changes in cell composition, signaling pathways, metabolism, differentiation, and functions of human PBMCs over the lifespan. Furthermore, we re-analyzed other independent PBMC data from COVID-19 and KD mapping to healthy reference data, and detected a functional shift: immune cells' BA differed from their CA, which was associated with the progress of vaccine efficacy and complications, respectively. Hereafter, we enrolled the largest PBMCs scRNA-seq datasets from 343 healthy individuals (over 2 million cells), and constructed a Physiological Age Prediction (PHARE) model. Collectively, our study provides essential insights for the precise treatment of patients at different life stages, and for building a novel machine-learning model to predict the patients' BA at both scRNA-seq and bulk RNA-seq levels.

Results

Depicting global features of human immune cell atlas over the lifespan

To determine the effects of age on the immune system, we initially integrated three public scRNA-seq datasets of PBMCs from 24 healthy individuals across children (Cd), teenagers (Tg), adults (Ad), elders (Ed), and supercentenarians (Sc) (Fig. 1A). After quality control and filtering, we obtained high-quality single-cell transcriptomes from 159,671 cells (Figures S1A and Supplementary Table 2). Then, the scRNA-seq data were normalized and used harmony to remove batch effects (Figures S1B). Based on an unbiased integrative analysis across all immune cells, 12 cell clusters (Figures S1C) across six main immune cell populations were annotated according to the most salient cell markers: myeloid cells ($CD14^+LYZ^+VCAN^+$), B cells ($MS4A1^+CD79A^+CD79B^+$), T cells ($C3D^+CD3E^+CD3G^+$), NK cells ($KLRD1^+KLRB1^+KLRF1^+$), platelets ($PPBP^+GP9^+$), and hematopoietic progenitor cells (Hpc, KIT^+CD34^+) (Fig. 1B, C, and Figures S1D). Immune cell compositions were compared among different age groups, and the results showed that all six immune cell types were present in each group, albeit in different proportions. To confirm this validity, the proportion of cells in each individual was analyzed, and no immune cells showed high degrees of interindividual heterogeneity (Fig. 1D). As shown in Fig. 1E, the composition of T cells significantly decreased, whereas that of NK and myeloid cells increased in old groups. Notably, the proportion
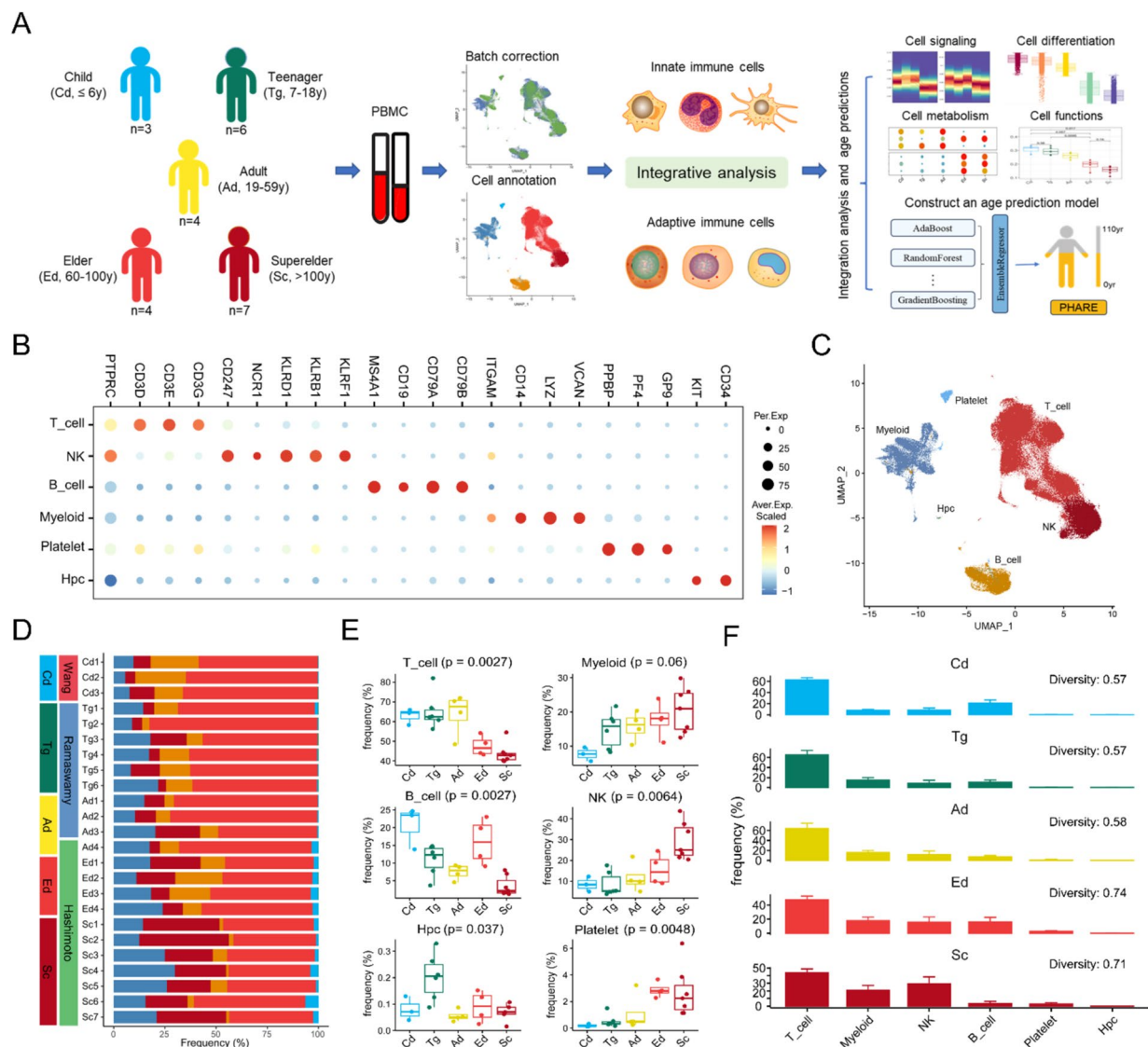
Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 3 of 21

Fig. 1 Depicting global features of human immune cell atlas over the lifespan. **A** Schematic diagram of the scRNA-Seq data collection, processing, and analysis design. **B** Dot plots showing representative signature gene expression in the main immune cell types. **C** UMAP projection of immune cell profiles in PBMC of different age groups. **D** Overview of data collection datasets, clinical age groups; quantification of main cell types per patient and color-coded by main cell type. **E** Box plot showing distribution of main cell types across different age groups. The *P* values are calculated with kruskal.test. **F** Bar plots showing the main immune cell types from different age groups (mean ± SD). Average diversity measured with the Shannon equitability index for each tissue is shown. Point size of dot plot shows the fraction of cells with non-zero expression

of platelets gradually increased in old groups (Figures S1E), which is likely associated with an increased incidence of cardiovascular disease in these populations [24]. The compositions of immune cells from the elderly displayed prominent differences because the diversities measured with Shannon equitability index were significantly higher than those in young groups (Fig. 1F). The results implied different aging processes

in different individuals, and this disparity was amplified with aging.

Pluripotency functions in innate immune subsets decline with aging

Given that innate immune cells provide the first line of defense to control pathogenic infections and instruct the subsequent adaptive immune response, we analyzed the functional alterations in innate immune cells associated

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 4 of 21

with aging. To classify each cell subpopulation in an unbiased manner, we re-clustered each cell lineage. Six cell clusters comprising five myeloid cell types with unique transcriptional features were revealed in the PBMCs of all age groups (Figures S2A, and 2B), including two types of CD14$^+$ monocytes (*CD14*_Mono1 and *CD14*_Mono2), one group of CD16$^+$ monocytes (*PCGR3A$^+$*), and two types of dendritic cells: *CD1C$^+$* cDC (conventional dendritic cell) and *JCHAIN$^+$* pDC (plasmacytoid dendritic cell) (Fig. 2A and B). Among the different age groups, we found that CD14_Mono1 was mainly enriched in the young groups (Cd, Tg and Ad), whereas CD14_Mono2 was markedly enriched in the elderly (Fig. 2C and Figures S2C). Besides, we included healthy samples from an independent PBMC bulk RNA-seq dataset (GSE180081) for analysis. We found the CD14_Mono2 score of the elderly group (≥ 60) was significantly higher than that of the young group (< 60), while the CD14_Mono1 score was significantly lower (Fig. 2D). GSEA analysis of CD14_Mono1 and CD14_Mono2 DEGs found the former was enriched for migration- and response- related pathways, while the latter was enriched for epigenetic alterations and oxidative phosphorylation (Fig. 2E and Supplementary Table 4). Consistent with previous findings [25], both cDC and pDC decreased with increasing age (Fig. 2C and Figures S2D). We further want to evaluated the functional changes of DC in the different age groups. Briefly, we extracted the highly expressed genes in DC subsets and clustered them based on their expression patterns across different age groups. We identified a total of 12 distinct gene modules and selected those that showed a gradual increase or decrease with age for further analysis. We found the genes involved in antigen presentation function of cDC gradually decreased with age, while the expression of splicing molecules increased (Fig. 2F and Supplementary Table 5). With aging, the metabolic pattern of pDC was remodeled, while the maintenance of protein homeostasis was progressively lost (Fig. 2F and Supplementary Table 5). To thoroughly

explore the function of all myeloid subsets, we evaluated the module scores associated with well-defined signatures of inflammation regulation and immune activation (Methods). The results indicated that TNF and IL6 signaling were highly activated in *CD14*_Mono1 cells, while *CD16*_Mono cells showed high IFN-induced signaling (Figures S2E). For DC populations, pDC displayed strong protein secretion ability, whereas cDC had high MHC II expression and antigen processing potency (Figures S2F), which was in agreement with their respective roles. In addition, unlike CD14$^+$ monocytes, the proportion of CD16$^+$ monocytes did not change substantially among different age groups (Fig. 2C). However, CD16$^+$ monocytes in the Tg group exhibited the highest levels of inflammation, IL6 and innate receptor signaling (Figures S2G). Taken together, the functions of monocytes and DCs gradually declined with aging.

NK cells are innate immune cells that play critical roles in coordinating tumor immunosurveillance and viral infection [26]. Five clusters of NK cells were identified in the PBMCs of all the age groups (Figures S3A), and identified four NK cell types with unique transcriptional features (Figures S3B), including three types of CD56^{dim} NK (classical_CD56^{dim}, inflamed_CD56^{dim}, and HLA_CD56^{dim}) and CD56^{bright} NK (Fig. 2G and H). We found that Classical_CD56^{dim} NK cells increased, whereas Inflamed_CD56^{dim} NK and CD56^{bright} NK cells decreased with aging (Fig. 2I). Notably, HLA_CD56^{dim} NK cells were specifically enriched in the Tg group (Figures S3C), although there was variation between healthy individuals (Figures S3D). Next, we used module scoring to evaluate the functional pathways related to NK based on the IOBR package, which provides a comprehensive investigation of the estimation of reported or user-built signatures [27]. Classical_CD56^{dim} NK cells had the lowest inflammatory signaling and CD56^{bright} NK cells had strong cytokine and chemokine secretion abilities (Figures S3E). Furthermore, we took advantage of PROGENy, which addressed both limitations by utilizing an extensive collection of publicly

(See figure on next page.)

Fig. 2 Pluripotency of functions in innate subsets declines with age. **A** UMAP projection of myeloid profiles in PBMC of different age groups. **B** Dot plots showing expression profiles of marker genes in different myeloid subsets. **C** Box plot showing distribution of myeloid subsets across different age groups. The *P* values are calculated with kruskal.test. **D** Box plot showing signature scores of monocyte subsets in different age groups based on ssGSEA algorithm, t-tests (two-sided) were performed. **E** Clustering network of significantly enriched GO pathways in the GSEA analysis. The nodes representing the significant GO pathways are colored by normalized enrichment score (NES). **F** The curves represent the changing trends of DCs' gene modules across different age groups (left), and heatmap showing gene expression of different modules (middle), and the represent functional pathways of enrichment analysis were shown within the boxes (right). **G** UMAP projection of NK profiles in PBMC of different age groups. **H** Dot plots showing expression profiles of marker genes in different NK subsets. **I** Box plot showing distribution of NK subsets across different age groups. The *P* values are calculated with kruskal.test. **J** Heatmap showing mean 14 PROGENy pathway scores of different NK subsets. **K** GO-based enrichment analysis illustrating indicated pathways upregulated in HLA_CD56^{dim} subset. **L** The curves represent the changing trends of CD56.^{bright} NK cells' gene modules across different age groups (left), and heatmap showing gene expression of different modules (middle), and the represent functional pathways of enrichment analysis were shown within the boxes (right)

Zhang et al. Immunity & Ageing (2024) 21:74

Page 5 of 21
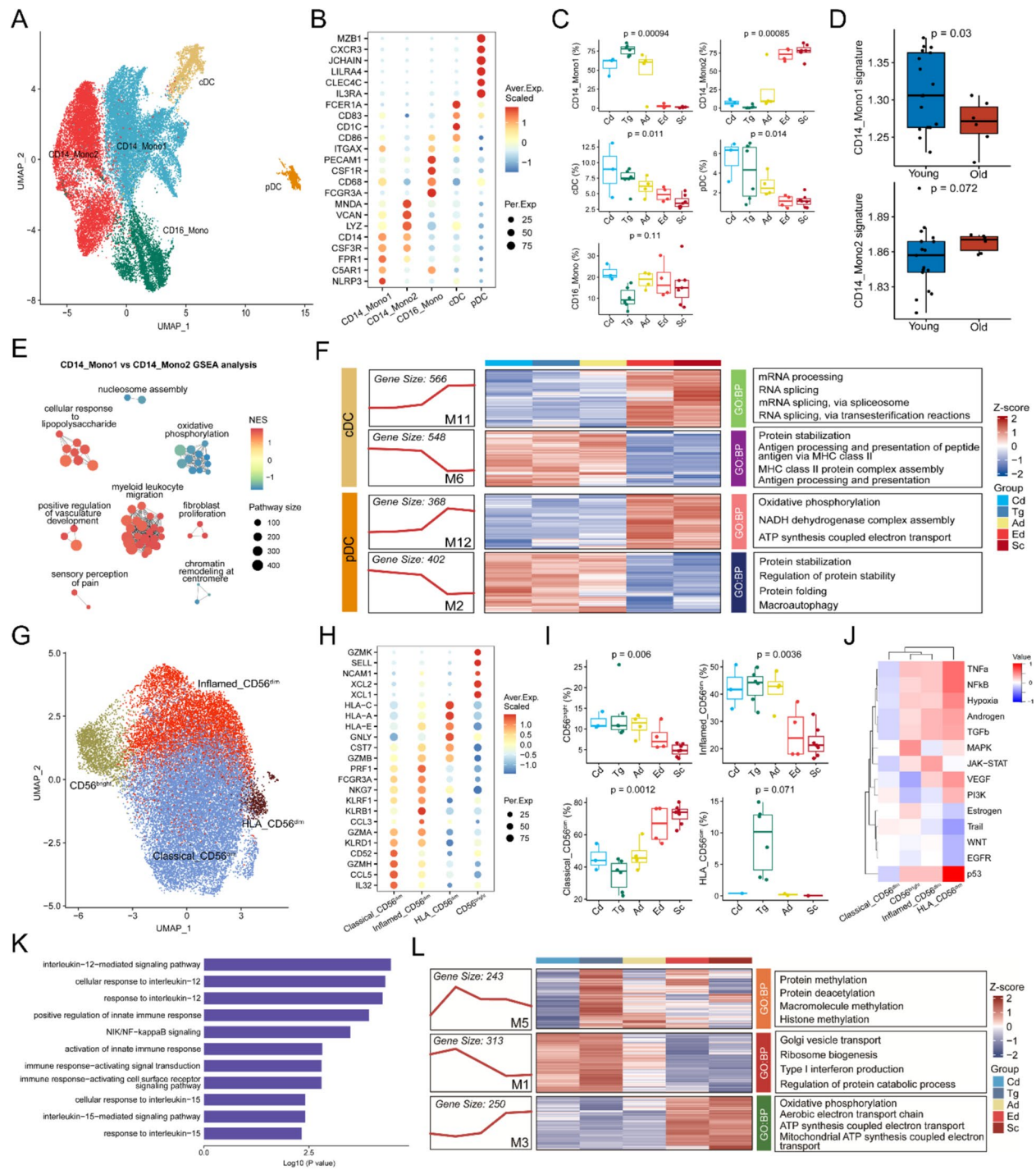
Fig. 2 (See legend on previous page.)

available perturbation experiments, to evaluate 14 cell global functional pathways of NK subsets [28]. The result showed HLA_CD56[dim] NK had strong inflammatory signatures and the p53 pathway was activated (Fig. 2J). Considering that HLA_CD56[dim] NK specifically presented in the Tg group, we wanted to further explore its features.

Pathway enrichment analysis of especially upregulated in HLA_CD56[dim] NK cells revealed enrichment of terms associated with "response to IL-12", "response to IL-15" and "NF-kB signaling", which were consistent with the inflammatory phenotype (Fig. 2K). Consistent with the literature [29], the CD56[bright] NK population responsible

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 6 of 21

for cytokine production also decreased with aging in our data (Fig. 2I). Additionally, we extracted the highly expressed genes in CD56^{bright} NK cells and performed gene pattern analysis across different age groups. We found the decrease in cytokine and chemokine secretion ability of this type of NK cell with aging was not due to cytokine regulation dysfunction but the decrease in golgi function and epigenetic alterations (Fig. 2L and Supplementary Table 6). Similar to CD56^{bright} NK cells, inflammatory CD56^{dim} NK cells also decreased with age, which may relate to cell cycle arrest (Figures S3F). Collectively, we revealed aging-related changes in NK cell composition and function.

Age-group-specific adaptive subsets match distinct system immune function

Although the steps underlying the activation and differentiation of Age-engaged B cells have been extensively characterised [30, 31], the cellular and molecular mechanisms underlying this process remain unclear. Therefore, we extracted B cells to further reveal seven cell clusters in PBMCs of all age groups (Figures S4A). Analysis of differentially expressed genes across these clusters revealed six B cell types with unique transcriptional features (Fig. 3A and Figures S4B): Naïve_B (*TCL1A*$^+$), CD83_B (*CD83*$^+$), Memory_B (*TNFRSF13B*$^+$), Activated_memory_B (*CD86*$^+$), Plasma_cell (*XBP1*$^+$*MZB1*$^+$), and Transitional_B cells (*CD5*$^+$) (Fig. 3B). Unlike innate immune cells, B-cell subsets varied erratically across age groups except Plasma_cell (Fig. 3C). However, although the number of plasma cells gradually increased with age, the function of protein secretio declined (Fig. 3D). Notably, CD83_B cells were specifically enriched in the Tg group, and each patient in this group had a higher proportion of these type B cells (Figures S4C and 4D). To reveal the features of each B cell subset, we evaluated the module scores of functional pathways based on the

signatures collected from previous studies. As depicted in Figures S4E, plasma cells showed higher protein secretion ability and lower antigen processing features among the six cell types, whereas Activated_memory B cells had the highest inflammatory score. Besides, CD83_B cells exhibited the second highest inflammatory features, also characterized by high expression of MHC class II molecules. PROGENy analysis illustrated that CD83_B cells had high TGF-β and MAPK pathway scores (Figures S4F). Thereafter, we compared CD83_B cells (Tg-enriched) with naïve B cells and found the former upregulated activation-related molecules (*IGFBP4* and *CD69*) and inflammatory molecules (*CCL4* and *CCL4L2*) (Fig. 3E and Supplementary Table 7). Furthermore, gene set enrichment analysis (GSEA) similarly indicated that CD83_B cells were transcriptionally poised to IL-2 and IL-15 stimulated B cells in an inflammatory state (Fig. 3F). Aging is associated with decreased efficacy of vaccination in both humans and mice, with reduced B cell memory formation [32]. In this regard, we found that BCR signaling was significantly impaired in the elderly Memory_B cells, which was also related to the reduced efficacy of vaccination and increased susceptibility to infection in the elderly (Figures S4G). Activated memory B cells were enriched in the Ad group. To reveal this type of cell function, we compared activated memory B cells with memory B cells and found that the former upregulated many cytotoxic molecules, such as *NKG7, GNLY, CCL5,* and *GZMB* (Figures S4H). Taken together, our data suggest that B cells did not adhere strictly to age variation, and group-specific enriched subsets played special roles with distinct functional features.

In contrast to B cells, there is a clear and well-accepted understanding that age-related aberrant T cell-driven cytokine and cytotoxic responses lead to the failure of immune tolerance and sensitivity to infectious diseases in older people [33]. Therefore, we extracted T cells to

(See figure on next page.)

Fig. 3 Age-group-specific adaptive subsets match distinct system immune function. **A** UMAP projection of B cell profiles in PBMC of different age groups. **B** Dot plots showing expression profiles of marker genes in different B cell subsets. **C** Box plot showing distribution of B cell subsets across different age groups. The *P* values are calculated with kruskal.test. **D** Densityheatmap showing indicates pathway score of plasma cells based on AUCELL algorithm (left), and representative genes dot plot (right). **E** Differential gene expression analysis using the log-fold change expression versus the difference in the percentage of cells expressing the gene comparing CD83_B cells versus Naïve_B cells (Δ Percentage Difference). **F** GSEA plots depict the enriched gene sets identified between the CD83_B cells and naïve_B cell subsets associated with B cell activation. **G** UMAP projection of T cell profiles in PBMC of different age groups. **H** Dot plots showing expression profiles of marker genes in different T cell subsets. **I** Box plot showing distribution of T cell subsets across different age groups. The *P* values are calculated with kruskal.test. **J** Densityheatmap showing indicates pathway score of CD8_CTL based on AUCELL algorithm (left), and representative genes dot plot (right). **K** Densityheatmap showing indicate pathway score of Treg based on AUCELL algorithm. **L** Heatmap showing transcriptomic similarity of T cell subsets. **M** Spearman correlation between percentage of CD4_CTL and unknown_T cells. **N** The distribution of T cell subtypes during the transition, along with the pseudo-time (upper). Subtypes are labeled by colors (lower). **O** Two-dimensional plots showing the dynamic expression of feature genes. **P** Heatmap showing the dynamic changes in gene expression along the different branches (left) and pathway enrichment results in each gene module (right). In figure **D**, **J** and **K**, the upper and lower curves of heatmap represent 75% and 25% of density, respectively. Point size of dot plot shows the fraction of cells with non-zero expression. The *P* values in Figure D, J and K are calculated with kruskal.test

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 7 of 21

Fig. 3 (See legend on previous page.)

explore the timing and mechanisms behind the decline in T cell function. First, all T cells were clustered into 18 subsets (Figures S5A). We then identified four CD4[+] T subsets (CD4_TNa, CD4_TEM, CD4_CTL, and Treg), three CD8[+] T subsets (CD8_TNa, CD8_TEM, and CD8_CTL), NKT, proliferating_T, and unknown_T cells (Fig. 3G). Subsequent marker gene expression analysis confirmed the accuracy of these annotations (Fig. 3H),

and the most up- and down-regulated genes were also calculated (Figures S5B). Among the T cell subsets, we found unprecedented percentage variation across age groups (Fig. 3I). Notably, CD4_CTL and unknown_T cells were especially enriched in the Sc group, although there was variation between patients (Figures S5C and 5D). To characterize various T cell subsets, we evaluated the functional pathways of all T cell subsets using module

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 8 of 21

scores. Our results showed that, except CD4_CTL, CD4$^+$ T cell subsets exhibited lower activities of HLA_signature and chemokine pathways than CD8$^+$ T cell subsets (Figures S5E). Interestingly, CD4_CTL exhibited comparable cytotoxic and inflamed scores, yet they did not express the same exhausted features as CD8_CTL (Figures S5F). Furthermore, the PROGENy score also distinctively clustered CD4$^+$ and CD8$^+$ T cells (Figures S5G). CD8_CTL cells, which serve as the body's robust defenders, were found to be most affected by aging [33]. We, therefore, examined the antimicrobial function of this T cell type across different age groups and found a decline in the elderly (Fig. 3J). Dot plots of related genes revealed that CD8_CTL cells in old people gradually lost the CD8 molecule (*CD8A* and *CD8B*) expression, whereas nature killer receptors (*KLRB1* and *KLRD1*) were upregulated. We also assessed the suppressive function of Tregs by analyzing the IL2-STAT5 and TGF-β pathways across different age groups. Results showed that although children had a higher percentage of Tregs, the Tregs in age group showed lower IL2 and TGF-β signaling, possibly due to limited antigenic acclimation (Fig. 3K). Effector memory T cells (TEM), which indicate the body's response to secondary immunity. As shown in Figures S5H, the TCR signaling in CD4_TEM and CD8_TEM were found to be lower in children, peaked in adults, and declined in the elderly. A similar pattern was observed in TCR and inflammatory signaling in NKT cells, while chemokine capacity was affected only by advanced age (Figures S5I). Recent research has highlighted the significant role of CD4_CTL cells in autoimmune diseases and tumor immunity, and an increase in these cells has been observed in supercentenarians [17, 34, 35]. However, the precursor cells for CD4 CTL remain unclear. As shown in Fig. 3I, CD4_CTL and unknown_T cells were both enriched in the Sc group, prompting us to investigate a potential correlation between these two T cell types. Correlation analysis suggested that other T cells were most similar to CD4_CTL at the transcriptome level (Fig. 3L). Further correlation analysis revealed a significant strong correlation between the proportion of CD4_CTL and unknown_T cells (Pearson correlation value 0.88, Fig. 3M). According to previous study [36], we discovered that unknown_T cells highly expressed CD4_CTL precursors' molecules (*CD27, TCF7*, and *LTB*) compared to CD4_CTL cells (Figures S5J and Supplementary Table 8). Therefore, we extracted CD4_CTL and unknown_T cells for pseudotime analysis to construct a developmental trajectory (Fig. 3N). We found that the expression levels of naive and stemness features were down-regulated, while killing molecules were gradually up-regulated along with pseudotime (Fig. 3O). We next investigated the transcriptional changes associated with two branches,

and explored the functional features of two distinct CD4_CTL subsets. We identified one subset characterized by a response to chemokines, and another expressing adhesion molecules indicative of direct contact with target cell; both subsets acquired cell-killing abilities during differentiation (Fig. 3P and Supplementary Table 9). Together, we identified a new CD4 T cell subset that was enriched in Sc group, may related to a higher percentage of CD4_CTL cells in the blood supercentenarians.

Characterizing developmental hierarchies and quiescent immune cells using CytoTRACE

Considering that not all immune processes are uniformly sensitive to aging, various immune cell populations are heterogeneously influenced by the increase in age [37]. To reveal cellular states with intrinsic differences in differentiation state across different age groups, we used CytoTRACE to construct a cell differentiation atlas among PBMCs (Methods). As depicted in Figures S6A, all myeloid cells were divided into different cell clusters, which were highly consistent with the age groups according to their CytoTRACE scores. Additionally, the box plot showed that myeloid cells in the Cd, Ad, and Tg groups were less differentiated, while those in the elderly showed higher levels of differentiation. Surprisingly, myeloid cells in the Sc group exhibited a higher potential for differentiation capacity. The conclusion that aging markedly affects NK function remains controversial and is related to the diverse health statuses of the study subjects [38]. Our results demonstrated that NK cells in the Ad group had the highest differentiation capacity, suggesting that their differentiation function was well maintained into adult stage (Figures S6B). In contrast to innate immune cells, previous studies have indicated that adaptive immune function declined with age [30, 39]. However, our findings illustrated that the differentiation state of B/T cells did not decline in a straight line but gradually improved before the adulthood, peaked in adults, and then declined (Figures S6C and 6D). Unlike other immune cells, T cells in the Sc group were the most differentiated, in contrast to Ed, which strictly followed a trend of gradual decline with increasing age. Conversely, these results revealed that lymphoid and myeloid cells had different differentiation trajectories; the former had a well-established differentiation potential early on, while the latter developed this ability until adulthood.

Analysis of the metabolic pathways reveal a shift in energy supply during immune cell aging

Metabolism, being central to all biological processes, is crucial for the proper regulation of immune cells. Perturbations in metabolism can cause immune dysfunction and disease progression [40]. Therefore, we aimed

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 9 of 21

to determine the metabolic activity of all immune cells across different age groups at single cell resolution. Following the instructions recommended by KEGG's official website, we divided the extracted specific metabolic pathways into six categories (Methods). Lipid and carbohydrate metabolism declined the most dramatically with increasing age (Figures S7A). Notably, some lipid-related metabolic pathways showed an inverse trend, such as arachidonic acid metabolism in myeloid and NK cells, alpha-linolenic acid metabolism, steroid hormones, and fatty acid biosynthesis in B and T cells. As shown in Figures S7B, oxidative phosphorylation activity was heightened in elderly groups across all immune cells, with other energy metabolism pathways also displaying increased activity in advanced age, excluding myeloid cells. Amino acid metabolism is essential for metabolic rewiring, supporting various immune cell functions beyond increased ATP generation, nucleotide synthesis, and redox balance [41]. Notable amino acid metabolism-specific enhancements were observed in aged T cells. In particular, taurine and hypotaurine metabolism pathways were detected exclusively in T cells and were more active in the elderly. Intrigued by these findings, we further explored this metabolic pathway to differentiate between CD4$^+$ and CD8$^+$ T cell subsets. Surprisingly, it was not the CD8$^+$ T cells (Figures S7C), but the CD4$^+$ T cells (Figures S7D) that showed an increasing trend with age. Collectively, these results highlighted that besides the uniformly active metabolites found in various immune cells of the elderly, there were also specific metabolites that were elevated exclusively in certain immune cells, which deserves further study.

Decline of infection-fighting immune function with aging

To delineate the characteristics of entire immune cell subsets, we integrated all immune cells (Panage_data) and calculated the percentages of each cell type within all PBMCs across samples (Supplementary Table 3). Unsupervised hierarchical clustering, based on cellular composition, showed that the patients formed a remarkable divergence in different age groups (Fig. 4A). Odds ratio (OR) analysis revealed the cell distribution preferences of each age group, such as HLA_CD56^{dim} NK and CD83_B cells enriched in Tg and unknown_T cells, and CD4_CTL enriched in Sc (Fig. 4B). Furthermore, to confirm the preferences of HLA_CD56^{dim} and CD83_B cells in Teenagers and verify their widespread distribution, we incorporated another recently published scRNA-seq dataset from a study that included seven patients with multisystem inflammatory syndrome (MIS-C), eight patients with coronavirus disease 2019 (COVID-19), and seven age- and sex-matched healthy controls (HC) [42]. We mapped 124,448 immune cells from all patients to Panage_data,

assigning cell annotation and age group labels based on transcriptome similarity (Figures S8A and 8B) (Method). From the OR analysis results, the HLA_CD56^{dim} and CD83_B cells also displayed a pronounced preference for the Tg group (Fig. 4C), while CD83_B and Treg cells were predominantly found in HC (Fig. 4D). These results suggested that CD83_B cells were depleted, while reduced Treg content was associated with systemic inflammation in COVID-19 and MIS-C patients. Considering the actual age of all included COVID-19 and MIS-C patients were below 18 years (Cd and Tg group), we defined the immune cells that mapped to Ad group (>18y) as 'shift-immune' cells. We then systemically assessed cell-type-specific functional changes among 'shift-immune' and 'preserve-immune' cells in COVID-19 and MIS-C patients (Method). We found that ribosome-related pathways were up-regulated in shift-immune cells, while TLR signaling, microbial and vaccine response-related pathways were significantly down-regulated (Fig. 4E). Besides, CD4_TNa cells underwent more significant remodeling in the COVID-19 group than in the MIS-C group, while CD8_TNa cells were more profoundly affected in MIS-C patients.

Based on the differences in the composition of immune cells, we conducted a comprehensive assessment of the anti-infection ability and sepsis score for each age group based on all immune cells (Methods). As depicted in Fig. 4F, except for the sepsis score, anti-bacterial, anti-viral, and anti-antimicrobial scores all declined progressively with increasing age. The cell types primarily contributing to these scores varied within each age group (Figures S8C). Considering the critical role of IFN signaling in anti-bacterial [43], anti-viral [44], and anti-antimicrobial [45] abilities, we further evaluated two types of IFN signaling across different age groups. These signaling pathways consistently decreased in the two elderly groups (Figures S8D). In summary, we revealed that the immune cell composition of each age group was ever-changing, and aging remodeled the anti-infection functions of immune cells.

Functional shifting of inflammatory myeloid cells is associated with complication progression in Kawasaki disease (KD)

KD, a self-limiting vasculitis, predominantly affects children under 5 years old and has become the leading cause of acquired heart disease in children from developed countries [46]. Besides the primary febrile symptoms, KD can lead to serious arterial lesions, and the prompt diagnosis of KD remains challenging [47]. Moreover, a recent study has indicated that KD can accelerate immune cell senescence through the generation of reactive oxygen species [48]. Therefore, we examined whether
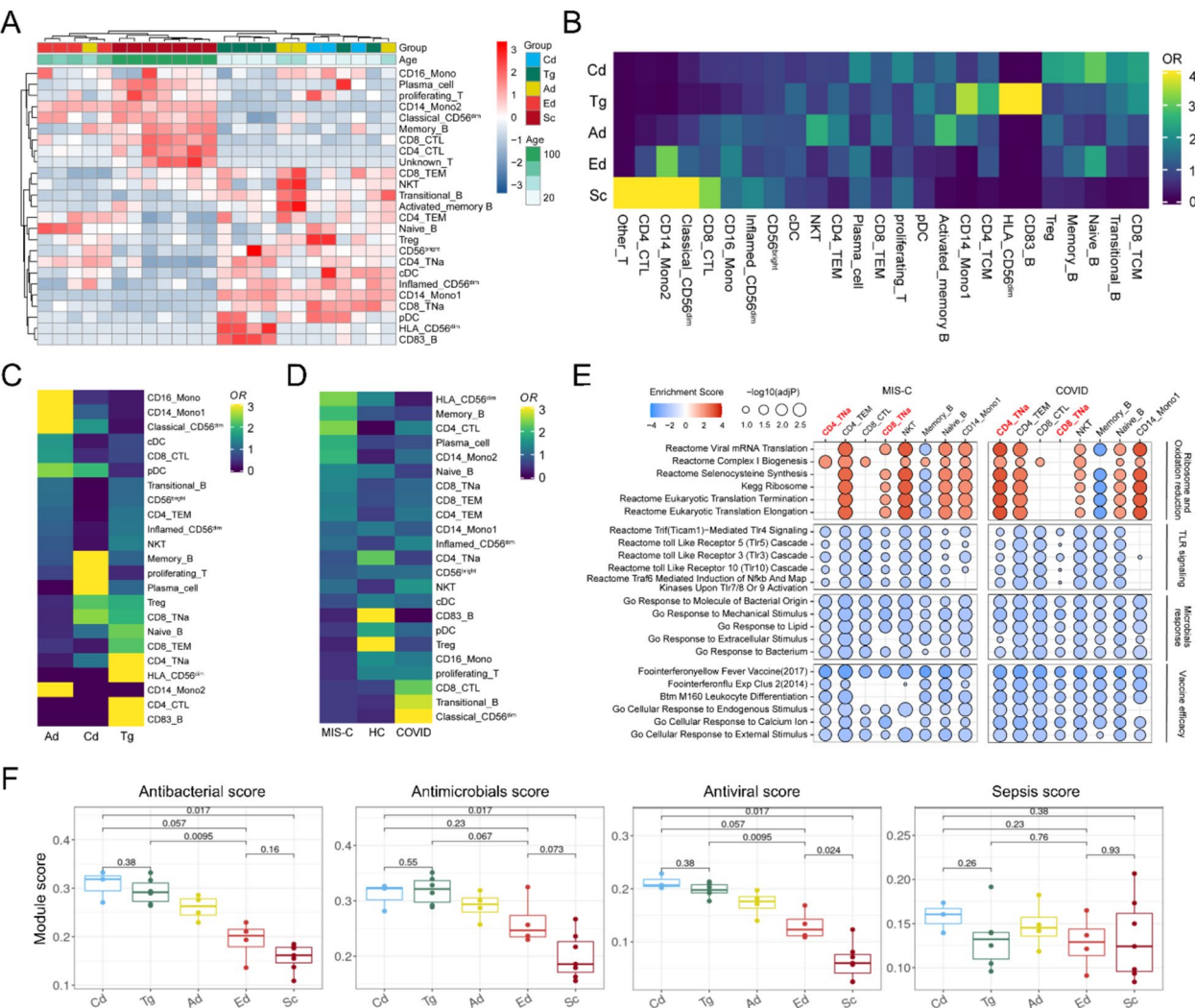
Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 10 of 21

Fig. 4 The anti-infectious functions gradually decline among age increase. **A** Heatmap showing clustering results based on the proportion of PBMC immune cell clusters, color-coded by age groups, patients indicated, Zscore means scaled proportion of cells. **B** Heatmap showing the ORs of meta clusters occurring in different age groups based on TAIA data. Heatmap showing the ORs of predicted cell types occurring in different age groups (**C**) and clinical groups (**D**) based on independent COVID cohort. **E** GSEA of shifting versus preserving immune cells in MIS-C (left) and COVID-19 (right) patients. Selected gene sets are grouped into functional/pathway categories. Dot color denotes normalized gene set enrichment score, and size indicates log10(adjusted *P* value). **F** Comparing anti-infectious related module scores with different age groups based on each patient's contribution. Statistical significance between groups is computed using a two-sided non-parametric Wilcoxon test

scRNA-seq data from KD, mapped to Panage_data, could capture KD-induced cellular senesence. We reanalyzed scRNA-seq data from a previous study that collected seven patients with acute KD before and after IVIG therapy [47]. After data quality control, we obtained 35,336 high-quality cells for further analysis before IVIG therapy and mapping to Panage_data (Figures S9A). Based on the aforementioned marker genes, six immune cell types were finely defined, including myeloid, NK cells, T/B cells, platelets, and Hpc (Fig. 5A and Figures S9B). Furthermore, we found that many immune cells in the Cd age group were mapped to older age groups (Fig. 5B).

This result suggested a broad influence of KD accelerated immune cell senescence process. Considering that all KD patients were children (Cd), we defined the change from Cd to Cd (Cd-Cd) as the 'preserve-state' and all others (Cd-Tg, Cd-Ad, Cd-Ed, and Cd-Sc) as 'shift-state' according to the mapping result (Figures S9C). We noted a remarkable divergence between the preserve- and shift- immune cell states, indicating transcriptomic difference between the two cell states (Figures S9D). To further verify the differential states of the shifting cells, all the immune cells were transferred to CytoTRACE. The result showed immune cells in the Cd group (preserve)

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 11 of 21

possessed the highest degree of differentiation potential compared to the other groups (Fig. 5C). As depicted in Fig. 5D, myeloid cells had higher shifting ratios (47.84%) than T (13.84%) and B cells (20.14%). Next, we assessed the functional differences between preserve- and shift-myeloid cells. GO enrichment analysis revealed that shifting myeloid cells increased the expression of genes involved in migration, cellular stress, and the inflammatory response (Fig. 5E). KEGG enrichment analysis identified pathways associated with severe KD complications, such as atherosclerosis, rheumatoid arthritis, and myocarditis in the highly expressed genes of shifting-myeloid cells (Fig. 5F). These findings suggested that the abnormal increase in shifting myeloid cells in KD patients may be associated with adverse clinical outcomes. In addition, we also found that shifting-T/B cells all possess a more active state than preserve-cells (Figures S9E and 9F). Cell metabolism analyses showed that immune cells in the shifting state had higher lipid and vitamin-related metabolic activities in contrast to preserving cells, indicating that shifting cells presented an overwhelming immune response (Fig. 5G and Figures S9G).

Intravenous immunoglobulin (IVIG) within the first 10 days after fever onset is the standard therapy for KD and remarkably reduces the rate of complications [49]. In this regard, we aimed to investigate the changes in KD PBMCs' shifting ratios after IVIG therapy. We analyzed the scRNA-seq data of KD after IVIG therapy and mapped it to Panage_data (Figures S9H). The same immune cell types were identified based on marker genes (Fig. 5H and Figures S9I). CytoTRACE analysis further confirmed that shift-cells exhibited a lower differentiation potential (Figures S9J). After IVIG therapy, each immune cell type was also found to shift-state, and the total shift- ratio was comparable (23.44% to 25.23%) (Fig. 5I). It is well established that monocyte abundance decreases and plasma cell numbers increase following IVIG therapy [47, 50, 51]. Our results further showed that the number of shift- B cells increased (20.14% to 36.10%), while shift- myeloid (47.84% to 36.62%) and NK cells (30.20% to 23.66%) decreased (Fig. 5J). Notably, shift- myeloid and NK signature remarkably down-regulated after IVIG therapy (Fig. 5K). Thereafter, we asked whether the shift-myeloid signature could be considered as a robust measure, and further evaluated changes after therapy using independent bulk RNA-seq datasets. We found shift-myeloid signature was significantly reduced after IVIG/IFX treatment (Fig. 5L). Besides, patients resistant to IVIG treatment had a more pronounced shift-myeloid signature compared to responsive patients (Fig. 5M). Correlation analysis revealed that shift-myeloid signature may also induce T cell senescence (Fig. 5N). It is known that KD progresses through different stages, but there is no diagnostic test for exact staging. We found that shift- myeloid signature was significantly higher in the acute phase than in the convalescent phase (Fig. 5O), which may provide new diagnostic insights for KD. Collectively, our results showed that the emergent pathogenic (shift) myeloid cells in PBMCs of KD patients may be responsible for complications; therefore, treatments that attenuate its proportion and pathogenicity could alleviate symptoms.

Constructing a machine learning predictor model to reveal the aging process

To build a transcriptome-based physiological age predictor model (PHARE), we compiled scRNA-seq datasets from PBMCs of healthy individuals spanning one lifespan, which contained more than one million cells (Method). All cells received cell annotation based on Panage_data, and then the proportion of all immune cell types was calculated to train the PHARE model (Fig. 6A). PHARE utilized two different pipelines for scRNA-seq and bulk RNA-seq data to predict physiological age from new datasets (Fig. 6B). We firstly evaluated the predictive performance of our model in the 177 healthy individuals

(See figure on next page.)

Fig. 5 Functional-shifting inflammatory myeloid cells are associated with complications emerge in KD. UMAP projection of immune cell profiles from KD PBMCs, (**A**) group by cell types (**B**) and predicted groups. **C** Boxplots showing CytoTRACE values for different predicted groups of KD PBMCs. **D** Alluvial diagram showed the different shifting rates from KD PBMCs mapping to TAIA. Axis 1, 2, and 3 represented real age, predicted groups, and cell types. Dotchart showed GO-BP (**E**) and KEGG (**F**) based enrichment results of shift- versus preserve- myeloid upregulated genes. Color indicates $\log_{10}P$value, and size indicates gene counts. (**G**) lipid-metabolic signature pathways in the main immune cell types of different age groups. Both color and size indicate the VISION score (**H**) UMAP projection of immune cell profiles from KD PBMCs after IVIG treatment, group by cell types. **I** Alluvial diagram showed the different shifting rates from KD PBMCs after IVIG treatment mapping to TAIA. Axis 1, 2, 3 represented real age, predicted groups and cell types. **J** Percentages of different myeloid and NK cell shift- subsets between before and after IVIG treatment, color-coded by cell subsets. **K** Violin plot showing shift- myeloid (upper) and NK (lower) signature score before and after IVIG treatment (GSE168732). The *P* values were calculated with wilcox.test. **L** Box plot showing shift- myeloid signature scores before and after IVIG treatment (left) or IFX treatment (right) based on ssGSEA algorithm (GSE48498). **M** Box plot showing shift- myeloid signature scores before and after IVIG treatment (left) or differential response groups (right) based on ssGSEA algorithm (GSE16797). **N** Spearman correlation analysis between shift- myeloid and T cell signature score (GSE16797). **O** Box plot showing shift- myeloid signature scores in different periods of onset based on ssGSEA algorithm (GSE73463). In figure L, M and O, the *P* values are calculated with t-tests (two-sided). The *P* values in Figure C is calculated with kruskal.test
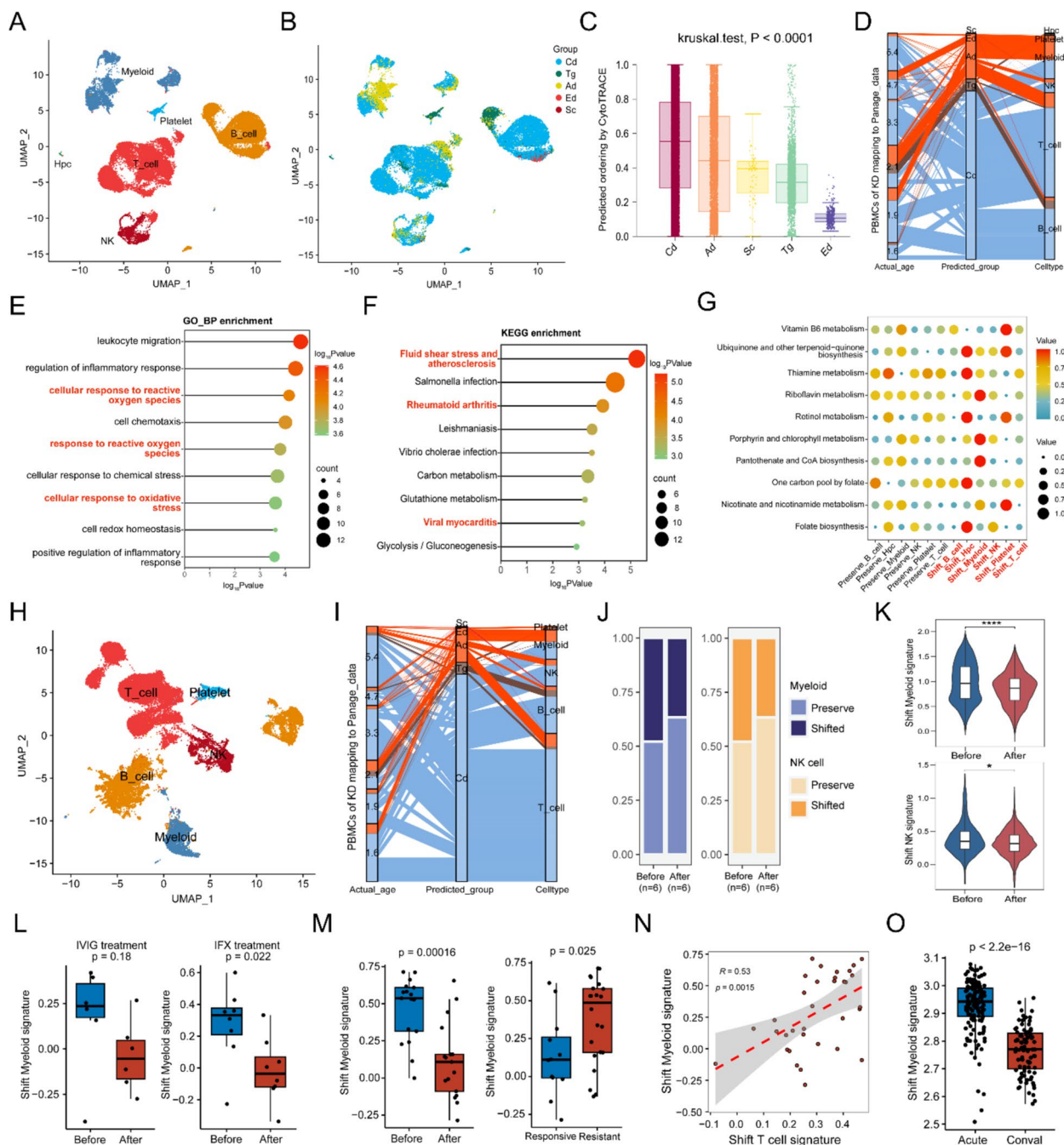
Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 12 of 21

Fig. 5 (See legend on previous page.)

using tenfold cross-validation. Specifically, the individuals were divided into 10 roughly equal parts. In each iteration, the model was trained on 9 parts, with predictions made on the remaining part, and the mean Mean Absolute Deviation (MAD) between chronological age and predicted age was calculated as a performance metric. After completing 10 rounds of training and generating predictions for all individuals, the final average MAD all

validations was 9.43 years (Fig. 6C). We noted there were no age differences were found when comparisons were made on each dataset, this result indicated the model was robust and independent of the different datasets (Fig. 6D). To further validate the accuracy and transferability of PHARE, we tested another PBMC scRNA-seq dataset [52] from a new cohort of healthy individuals (syn49637038, *n* = 166). We found the MAD was 9.35

Zhang *et al. Immunity & Ageing* (2024) 21:74
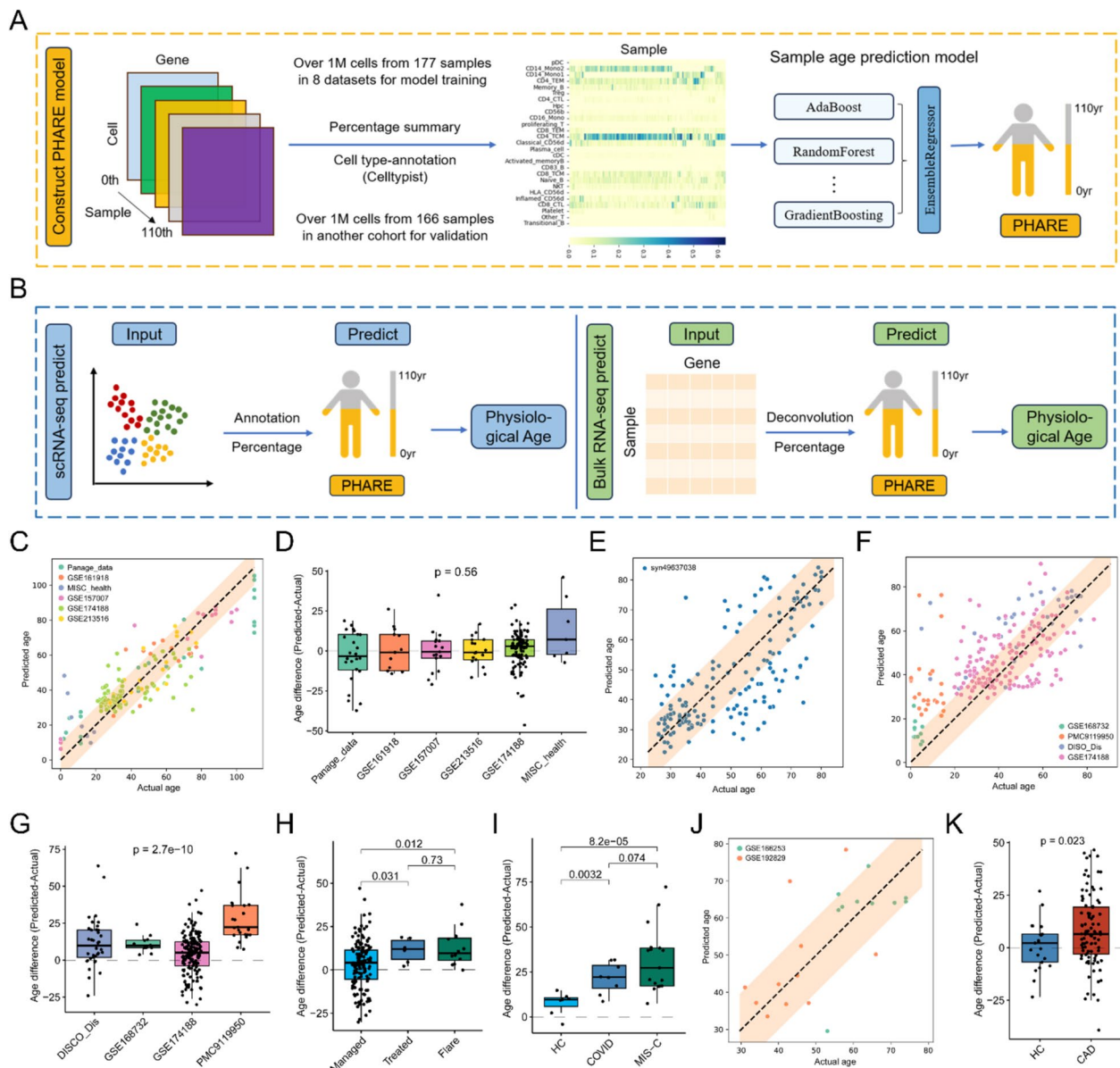
Page 13 of 21

Fig. 6 Constructing an age predictor model to explore human aging process. Workflow summary of PHARE construction (**A**) and applied PHARE to age prediction based on scRNA-seq and bulk RNA-seq data (**B**). **C** Spearman correlation analysis between predicted and actual age in discovery scRNA-seq datasets from healthy people. **D** Box plot showing age difference was not different between different healthy scRNA-seq cohorts. **E** Spearman correlation analysis between predicted and actual age in validation scRNA-seq datasets from healthy people. **F** Spearman correlation analysis between predicted and actual age in scRNA-seq datasets from diseased people. **G** Box plot showing age difference was significantly different between different disease cohorts. **H** Box plot showing age difference was significantly different between different stages of SLE. **I** Box plots showing age difference were significantly different between different clinical groups. **J** Spearman correlation analysis between predicted and actual age in bulk RNA-seq datasets from healthy people. **K** Box plot showing age difference was significantly different between healthy and CAD samples based on bulk RNA-seq data (GSE180081), the *P* values are calculated with t-tests (two-sided). In figure C, D and F, the *P* values are calculated with kruskal.test. In Figure G, H and J, the *P* values are calculated with t-tests (two-sided)

years in this dataset similar to the results of internal validation, indicating that PHARE had good accuracy for age prediction (Fig. 6E). To assess whether PHARE can detect disease-induced aging changes, we then applied PHARE to scRNA-seq datasets from various disease samples. As hypothesized, we found the predicted age difference for disease-derived datasets to be 13.61 years, higher than that for the healthy sample (Fig. 6F). Besides, we found age difference differed significantly across datasets from different disease sources, which indicated

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 14 of 21

diseases had a variable impact on the aging process (Fig. 6G). Further analysis in systemic lupus erythematosus (SLE) patients revealed flare symptoms and those treated with no response had significantly higher age difference than those well-managed (Fig. 6H). This result suggested PHARE captured aging-accelerated features, which can help assess the impact of treatment. When applied to a COVID-19 dataset, PHARE revealed that patients with multisystem inflammatory syndrome in children (MIS-C) appeared older than ordinary COVID-19 patients and HC (Fig. 6I). This finding was consistent with the notion that hyper-inflammation can accelerate the aging process. Although PHARE was constructed based on scRNA-seq data, we wanted to further test whether it was also applicable for age prediction from bulk RNA-seq data. We then evaluated the accuracy of PHARE in two bulk RNA-seq datasets from healthy samples, which had age difference between chronological age and predicted age of 9.09 years (Fig. 6J). This result indicated performance of PHARE on the bulk RNA-seq data remains robust. We further performed PHARE on another bulk RNA-seq data from CAD, and found CAD patients also had higher age difference than HC sample (Fig. 6K). Overall, these results illustrate that PHARE is a robust aging clock predictor, which can successfully capture aging-accelerated and aging-delayed features across various diseases.

Discussion

The aging process is complex, influenced by genetics, the environment, and their interplay. Given these complexities, traditional molecular biology approaches often fall short of elucidating the multimodal processes of aging. Single-cell techniques have emerged as powerful tools in high-throughput biology and systems immunology analyses [37]. By integrating single-cell datasets from previous studies [5, 17, 47], our study illustrated immune cell signaling, metabolism, differentiation, and function changes of PBMCs, ranging from newborns to supercentenarians. We revealed several period-specific enriched cell types and reproduced these findings in independent datasets. Notably, we discovered a new population of T cell subset (unknown_T) in supercentenarians, whose gene expression were similar to CD4_CTL and proportions are positively correlated with CD4_CTL. Besides, we found this subset had the potential differentiated into two distinct CD4_CTL subsets. By mapping immune cells from autoimmune and infectious diseases to healthy immune cells across human lifespan, we accurately captured aging-accelerated features at the cellular level. Collectively, our study constructed the first immune cell functional profile completely spanning people's lifelong, and developed a machine learning model capable of predicting the BA of

patients, suitable for both scRNA-seq and bulk RNA-seq datasets.

With aging, innate immune cells exhibit heterogeneous aging phenotypes, which correlate with their impaired capability for performing effective immune responses to newly encountered pathogens or vaccine antigens [29]. Consistent with a previous study, our results also showed a bias towards myeloid over lymphoid cell differentiation [53]. While the total percentage of myeloid cells was higher in the elderly, the proportions of both cDC and pDC decreased with age. Additionally, monocytes and dendritic cells progressively lose their immune stimulation function as they age. This contributes to why advanced age is associated with progressive immune deficiency, increased susceptibility to infections, and an impaired response to vaccination [54]. NK cells are another vital component of the innate immune system and serve as the first line of defense against infections and emerging malignancies. Our findings showed an increased percentage of NK cells in the elderly; however, a large proportion of these were senescent NK cells with weakened functional pathway signaling. Additionally, CD56^{bright} NK cell subsets, known for rapid proliferation and extensive cytokine and chemokine production upon activation, gradually decrease with aging [38]. Notably, we identified a specific NK cell subset named HLA_NK cell, which was particularly enriched in the Tg group. Through pathway enrichment analysis, we found that HLA_NK cells showed high expression of IL-12 and IL-15 signaling genes. Previous studies have demonstrated that IL12 and IL15 sustain the effector function of NK cells in established tumors and infections [55, 56]. Therefore, this NK cell subset may contribute to the lower cancer rates and robust anti-infection abilities observed in teenagers.

Unlike the innate immune system, the functions of adaptive immune cells undergo drastic remodeling, often described as a shift from naïve to memory phenotypes with increasing age [2]. The representation of B cells is considerably altered with aging, and age-associated B cells, termed *CD19$^+$CD21$^-$CD23$^-$* B cells in old mice and *CD19$^+$CD38$^-$CD24$^-$* B cells in humans, were found to increase [57, 58]. In addition to the change in composition, B cell functions also become dramatically remodeled, gradually losing their antibody production capacity and acquiring inflammatory characteristics. We found a potential mechanism: a decreased protein secretion ability in plasma cells, marked by reduced expression of genes involved in protein trafficking (*ARF1* and *ARFGAP3*), protein sorting (AP-3 complex), and protein channel (*TMED10*) in the elderly [59, 60]. Unlike normal B cells, Age-Associated B Cells (ABCs) can rapidly respond to innate stimuli (toll-like receptors) and require minimal B-cell receptor (BCR) stimulation for

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 15 of 21

activation [61]. Mechanistically, we found that BCR signaling in memory B cells, which are prevalent in the elderly, diminished with age. Additionally, CD83_B cells were specifically enriched in the Tg group, a result validated by an independent dataset. GSEA also indicated that CD83_B cells were transcriptionally predisposed to differentiate into light zone B cells. Considering that light zone B cells are activated and selected based on the affinity of their immunoglobulin, this finding implies that vaccination during adolescence might be more effective [62]. Owing to thymic involution in early adulthood, the maintenance of the naïve T-cell pool from the adult stage relies heavily on peripheral homeostatic proliferation [61, 63]. Moreover, because aging hematopoietic cells tend to favor myeloid differentiation, both our study and others have observed a gradual decline in total T cells [64]. Age most notably affects CD8$^+$ T cells, characterized by a loss of naïve CD8$^+$ T cells and gain of terminally differentiated CD8$^+$ TEMRA cells (*CD45RA$^+$KLRG1$^+$CCR7$^-$*) [65]. Contrary to a previous study, our data showed that CD8_CTL in the elderly did not highly express KLRG1 but rather KLRD1 and KLRB1, both of which acted as immune inhibitory receptors on NK cells. Peripheral homeostatic proliferation of T cells in humans effectively maintains the naïve CD4$^+$ T cell pool, ensuring that CD4$^+$ T cells are only moderately affected by age [61]. The mechanism by which CD4$^+$ and CD8$^+$ T cells are affected differently by aging remains unclear. Recently, CD4$^+$ T cells with cytotoxic functions (CD4 CTL) have been recognized for their significant roles in viral infections, autoimmune diseases, and cancers [35, 66, 67]. Hashimoto et al. illustrated that CD4 CTL dramatically increased in supercentenarians, a cell type also found in elderly patients with bladder cancer [17, 34]. Consequently, it is increasingly acknowledged that CD4 CTL have diverse functions and that their percentage is significantly associated with advanced age. On this basis, our results showed that CD4_CTL was detected in the peripheral blood of some adults, with a significant increase in the elders (60–110 yr), and the highest levels in supercentenarians (>100 yr). We further discovered that the emergence of unknown_T subset in supercentenarians may be contributing to a sustained high generation of CD4_CTL in this group of people.

Biological aging is a complex process characterized by progressive deterioration occurring simultaneously at multiple levels. This process is tissue-specific and can even vary from one cell to another [9]. Accurate measurements to assess BA can be beneficial in capturing age-related physiological changes. Previous studies have developed numerous BA predictors through telomeres, epigenetic clocks, frailty, and other clinical biomarkers [68–71]. However, all these measurements remained

at the individual level, which significantly obscures the differences at single cell level. Therefore, we proposed a novel single-cell level BA prediction model (PHARE) based on the changes in every immune cell. Similar to other age prediction models [11, 13, 14, 72], our approach also exhibits age difference between the predicted ages and the patients' actual ages. Upon analyzing KD (representing autoimmune diseases) and COVID-19 (representing infectious diseases) patient data, we consider the existence of age difference to be reasonable.

Collectively, our study has uncovered changes in immune cell function throughout the human lifespan, establishing a valuable reference for further research on aging. Our comprehensive analysis of the cell signaling, metabolism, differentiation, and functions of PBMCs provides an invaluable resource for the study of the aging process research, which will facilitate future precision treatments for diverse populations. Ultimately, we have developed PHARE into a user-friendly web tool (https://xiazlab.org/phare), which will greatly assist users without programming skills to use it for research on large-scale age-related diseases. In the future, we will continuously update PHARE and include additional samples from healthy individuals for training to enhance the accuracy of age prediction even further.

Limitations of the Study
Our study has several limitations. One limitation of our study was the use of a cross-sectional dataset in which associations can be identified rather than causalities. Further follow-up studies are required to validate these findings. The healthy samples we included rely on the clinical information provided by the original studies. Although these samples were disease-free at the time of collection, their medical history and genetic background could still impact the function of their immune systems. Future studies should ideally establish a standardized definition of health to uniformly select samples for analysis. In addition, sex-based differences in the immune system vary with age, and future studies need to include more samples to separate genders. Although we included data from multiple studies, we still did not define neutrophils. A possible reason is that the 10×platform does not perform well on the capture rate of these two types of cells, and other single-cell sequencing platforms are required to explore these two types of cells in the future.

Materials and methods
Panage_data collection and integration
To fully cover PBMC data across human lifespan, we included three separate studies. We extracted PBMC data from healthy children [47] from Zhen Wang et al.' study (GSE168732), PBMC data from healthy teenagers

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 16 of 21

and adults [5] from Anjali Ramaswamy et al.' study (GSE166489), and PBMC data from healthy elders and supercentenarians [17] from Kosuke Hashimoto et al.' study (http://gerg.gsc.riken.jp/SC2018/) (Supplementary Table 1). Furthermore, we integrated cells into a shared space from different datasets for unsupervised clustering and used the harmony algorithm to correct the batch effect [73]. To detect the most variable genes used for harmony algorithm, we selected high-variable genes separately for each patient by "FindVariableFeatures" function from Seurat package (version 4.0.6). A consensus list of 2,000 high-variable genes was then formed by selecting the genes with the greatest recovery rates across patients, with ties broken by random sampling. Then, we ran Harmony with default parameters and integrated all datasets by using the "RunHarmony" function.

After dataset integration, the gene expression matrixes of all PBMCs were imported into Seurat v4 for subsequent analyses [74]. The following filtering steps were carried out to exclude low-quality cells: cells with fewer than 300 and more than 8000 detected genes were discarded; cells with a high fraction of mitochondrial genes (>10%) and hemoglobin genes (>1%) were removed. As a result, a total of 159,671 cells (Cd: 16,025, Tg: 57,945, Ad: 31,759, Ed: 15,716, Sc: 38,226) with a median of 1,087 genes were included in the further analyses. We applied SCTransform workflow (https://satijalab.org/seurat/) to analyze the scRNA-seq data with default parameters, which replaces the "NormalizeData", "ScaleData" and "FindVariableFeatures" functions. Then, we performed a principal component analysis (PCA) dimensionality reduction (RunPCA) and selected the first 30 PCs to construct a shared nearest neighbor (SNN) graph (FindNeighbors). To visualize the clustering results, the non-linear dimensional reduction was performed with the Uniform Manifold Approximation and Projection (UMAP) method, and cluster biomarkers were found by the "FindAllMarkers" function from the Seurat package.

External datasets mapping to Panage_data

To verify whether the cell types enriched in Tg group we identified in our origin datasets exist in other teenager PBMC samples, we included another published teenager PBMC scRNA dataset in our analyses [42]. For using Panage_data as a reference tool to interpret other datasets, PBMCs from KD patients [47] and COVID-19 patients [42] were included for further analysis. After the same data quality control, highly-quality immune cells were prepared for mapping to Panage_data. Then, marker genes common across a reference (Panage_data) and query (external datasets) were found by running FindAllMarkers, and the top 30 PCs were selected for cell clustering. Furthermore, we followed the tutorial (https://satijalab.org/seurat/v4.0/reference_mapping.html) to map each donor dataset from the query individually. We used the FindTransferAnchors function with reduction="pcaproject" and MapQuery function as the previous study described [75], and setting reduction="pca" (as the documentation recommends for unimodal analysis). Finally, all PBMCs from external datasets were annotated according to our defined cell type.

Scores quantification of function in scRNA-seq data

To score individual cells for Hallmark, KEGG, or other previously published functional pathway activities, we used multiple previously described methods to analyze different immune cell subsets. Firstly, the used human genes were from msigdbr (version 7.4.1) package, and gene sets were then used to score each cell. To eliminate the bias of sample background, we selected gene set enrichment analysis methods based on single-cell gene expression ranking AUCell [76], UCell [77], singscore [78], and ssGSEA [79]. Of note, ssGSEA cancels the final standardization step, making it closer to the gene set enrichment analysis of a single cell. In addition, to evaluate whether the gene set is enriched in a certain cell subpopulation, we calculated the differential gene sets in the enrichment score matrix by Wilcox test (the filter criterion for differential genes is that the *P*-value after correction is less than 0.05). Finally, we used the rank aggregation algorithm (RRA) in the RobustRankAggreg package [80] (version 1.1.0) to comprehensively evaluate the results of the different analyses, and screen out the genes that are significantly enriched in most gene set enrichment analysis methods Set (the filter criterion for comprehensive evaluation is *P* value less than 0.05).

Anti-infection module score calculation

Module scores were calculated using the "AddModuleScore" function of Seurat with the default parameters. The anti-viral score consists of *SIGLEC1, RABGAP1L, IFI27, CADM1, RSAD2, MX1*, and *SERPING1* [81]. The anti-bacterial score consists of *SMPD1, CD44, SERPING1, SPI1, HERC1, MCTP1, FOLR3, CFAP45, PRF1, CTBP1, HLA-DRB1, ARL1, OAS3, ZER1, CHI3L1, IFIT2*, and *IFITM1* [81]. The anti-microbial score consists of 463 genes that come from ImmPort database [82]. The sepsis score consists of *PLAC8, CLU, RETN, CD63, ALOX5AP, SEC61G, TXN,* and *MT1X* [83]. Besides, to reveal the anti-infectious mechanism, Type_I_IFN and Type_II_IFN related module scores were also calculated based on Rooney et al.' study [84]. In these four anti-infectious features, each patient has 25 cell types and each cell type has its specific score. For *i*-th patient, his/her patient's score is calculated as follows:

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 17 of 21

$$\sum_{j=1}^{25} s_j \cdot \alpha_{i,j}$$

where s_j is j-th cell type's specific score, and $\alpha_{i,j}$ denotes the proportion of the j-th cell of the i-th patient, satisfying:

$$\alpha_{i,j} \geq 0, \sum_{j=1}^{25} \alpha_{i,j} = 1$$

Then the differences between groups are compared for each patient of groups, the differential features in the established score matrix by Wilcox test. Other indicated functional gene sets used for cell annotation were extracted from IOBR (version 0.99.9) package [27].

Depicting cell global function using PROGENy
Given that many previous methods placed less emphasis on integrating responses from various cell types, we selected PROGENy to more accurately infer pathway activity from gene expression in heterogeneous samples [28]. This method can overcome both limitations by leveraging a large compendium of publicly available perturbation experiments to yield a common core of Pathway RespOnsive GENes. Specifically, PROGENy evaluated 14 cell global functional pathways, which contained cell death, cell proliferation, cell metabolism, hormone signaling regulation and several immune signaling. Therefore, we used this method to evaluate the indicated immune cell function of different age groups.

GO and KEGG enrichment analysis
Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed by ClusterProfiler (version 3.18.0) package, using genes specifically upregulated in indicated cell types as the input gene list. GO/KEGG biological process terms with Bonferroni-corrected P values (FDR) < 0.05 were considered significantly enriched terms. The aPEAR (version 1.0.0) package was used for clustering of enriched pathways and visualization. To cluster the differentially expressed genes, ClusterGVis (version 0.1.1) package was employed. Integration of the heatmap and functional enrichment analysis was also performed using the ClusterGVis package.

Geneset and pathway scores assessment
To assign pathway activity estimates to individual cells, we applied GSVA using standard settings, as implemented in the GSVA package (version 1.42.0). For bulk RNA-seq data, we applied ssGSEA, as implemented in the IOBR package (version 1.0.0), to assess signature scores using the top 50 upregulated genes of indicated immune cell subsets (Supplementary Table 10). Gene Set

Enrichment Analysis (GSEA) analysis was performed for each cell subpopulation using genesets (Hallmark) available at Molecular Signatures Database (MSigDB, https://www.gseamsigdb.org/gsea/downloads.jsp) or previously reported functional genesets [42] with default parameters. P values were adjusted using the Benjamini–Hochberg method for the whole geneset list. Selected pathways shown in figure were manually curated to select genesets relevant to metabolism, infection and vaccine across the various differential expression comparisons.

Detection of differentiation states based scRNA-seq data
Determining both the state and direction of cell differentiation is tricky, while scRNA-seq offers a powerful method for reconstructing cellular differentiation trajectories. Notably, CytoTRACE infers stem cell properties by scoring each cell based on its transcriptional diversity, establishing a novel RNA-based marker of developmental potential and providing a platform for delineating cellular hierarchies. [85]. Therefore, we used CytoTRACE to reveal the differentiation degree of each cell type, which indicated the ability to respond to new antigens. The raw expression matrix (counts) for different immune cell subtypes was extracted and imported to the CytoTRACE package (version 0.3.3), and analyzed with default parameters. Then, the output CytoTRACE score for each cell was plotted with the "plotCytoTRACE" function, which visualized all cells in a low-dimensional embedding.

Mining metabolic activity using scMetabolism
To estimate the metabolic activity of individual immune cells among different age groups, the scMetabolism (version 0.2.1) package was used developed by Yingcheng Wu et al [86]. We chose the KEGG database (https://www.genome.jp/kegg/pathway.html#metabolism) as the reference to quantify metabolic pathways in different cell types. In essence, selected KEGG metabolism pathways were generally divided into 6 categories (Amino acids, Lipids, Carbohydrates, Glycan, Cofactor/vitamin and Energy) according to the database recommendation. Considering VISION to be most appropriate for single-cell metabolism quantification among all methods, we evaluated the metabolism activity based on VISION score. Then, each metabolism activity was visualized using a modified "DotPlot.metabolism" function.

Similarity and correlation analysis
To measure the transcriptional similarity of T cell subsets, we evaluated the expression values of all genes in the T cell. Briefly, ward.D2 method was applied for hierarchical cluster analysis following Euclidean distance estimated by using the average expression of all genes. Then, spearman correlations were calculated between

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 18 of 21

the different T cell subsets and visualization was performed by heatmap. For correlation analysis, the proportions of different immune-cell subsets or pathway scores were first calculated. Then, the correlation coefficients between them were estimated based on Spearman.

Developmental trajectory inference of CD4_CTL cell

To characterize the developmental origins of CD4_CTL cells, we extracted unknown_T and CD4_CTL cells and imported them to the Monocle (version 2.14.0) software [87] to construct developmental trajectory. The signature genes were calculated by dispersionTable function, and further filtered out low-expressed genes based on their expression values. Then, the CD4_CTL cell differentiation trajectory was inferred with the default parameters of Monocle after dimension reduction and cell ordering. After the cell trajectories were constructed, differentially expressed genes along the pseudotime were detected using the "differentialGeneTest" function.

The odds ratio for cell distribution preferences

To characterize the cell distribution preferences, the odds ratio (OR) was calculated and used to indicate this feature as previously described [88]. OR is a measure of how strongly an event is associated with exposure and a ratio of two sets of odds. We first assumed that the cell type was *Ct* and the group was *g*, and calculated the OR to evaluate the preferences between *Ct* and *g*. The Fisher's exact test was applied to this contingency table, thus OR and corresponding *p*-value could be obtained. Specifically, we defined a higher OR with a value > 2 indicating that indicated cell type *Ct* was more preferred to distribute in group *g*. *P*-values were adjusted using the BH method using the R function "p.adjust".

Diversity assessment based on immune cell composition

To compare the immune cell composition among different age groups, the frequency of immune cell metaclusters was calculated separately. Then a Shannon equitability index (normalized Shannon diversity index) was calculated, and the details were described as previously reported [88]. The high index indicated composition of the various immune cells is more diverse.

Physiological age prediction model construction (PHARE)

We developed a novel physiological age prediction model (PHARE) using PBMC single-cell transcriptome data from healthy individuals, leveraging a machine learning approach. The specific steps are as follows: First, we used a single-cell dataset Panage_data that had been pre-annotated with various immune cell subtypes as a reference to train the Celltypist model [89]. The training process followed the standard procedure

of Celltypist, which involves proportional sampling of each cell type from the reference dataset to enhance the ability to label rare cell types. Then, we used the trained model to label cell types across all other normal samples' single-cell data. Ultimately, for each sample within all normal peripheral blood datasets (including the reference dataset), we calculated the proportion of each cell type and used these proportions as features to construct a sample age prediction model. This prediction model is an ensemble of three machine learning techniques: 'Random Forest', 'Gradient Boosting', and 'AdaBoost', developed using the standard process of Sklearn [90].

Subsequently, we developed a process to use the aforementioned trained model for physiological age prediction in samples at single-cell resolution and bulk resolution. For single-cell resolution transcriptome data samples, after cell type labeling using our trained Celltypist model, the ensemble model can directly provide age predictions (Fig. 6B). For bulk-resolution transcriptome data samples, we first employ cell type deconvolution technology, still using the pre-annotated single-cell dataset (Panage_data) as a reference, to parse out the cell type proportions in bulk samples. Specifically: We first use the *scanpy.tl.rank_genes_groups* function to calculate the up and down markers of the reference dataset; then, we retain markers unique to each cell type and use them with the intersection of genes contained in the bulk dataset as the reference genes for deconvolution; moreover, from the single-cell data, we calculate the average expression level of reference genes across each cell type to create a signature matrix; finally, we perform cell type deconvolution using the non-negative matrix factorization technique (NNLS), while providing NuSVR and Linear [91] methods as alternative options for deconvolution algorithms. Through this process, we input the deconvoluted cell type proportions into the ensemble model, thereby providing age predictions for bulk resolution samples (Fig. 6B).

Statistical analysis

Data are presented as means ± SEM, with the sample size per group provided in the respective figure legends. For bulk RNA-seq data, statistical significance was determined using a two-tailed Student's t-test, while the Wilcoxon test (wilcox.test) was applied for comparisons based on scRNA-seq data. Differences among multiple groups were evaluated using the Kruskal Wallis test (kruskal.test). All bar graphs depict means with error bars representing data distribution. Significance levels are indicated as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. *P* values < 0.05 were considered statistically significant.

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 19 of 21

Supplementary Information

Supplementary Material 1: Figure S1. Quality control and cell types annotation. Figure S2. Single-cell transcriptomics revealed distinct myeloid subsets of different age groups. Figure S3. Single-cell transcriptomics revealed distinct NK subsets of different age groups. Figure S4. Single-cell transcriptomics revealed distinct B cell subsets of different age groups. Figure S5. Single-cell transcriptomics revealed distinct T cell subsets of different age groups. Figure S6. Characterization of developmental hierarchies and quiescent immune cells using CytoTRACE. Figure S7. Analysis of the metabolic pathways reveal a shift in energy supply during immune cell aging. Figure S8. External validation of cell subsets enriched in teenager group. Figure S9. Reanalyzing previous KD PBMCs data based on TAIA as reference.

Supplementary Material 2.

Code availability
All original code used to replicate this study has been deposited on GitHub repository (https://github.com/Immugent/Analysis-of-pan-age-PBMC-scRNAseq), and the PHARE model generated in this study are available at the GitHub repository (https://github.com/cliffren/PHARE).

Authors' contributions
C.Z., Z.X., and B.Z. conceptualized the study and designed the article. C.Z., T.R., X.Z., and Y.S. analysed the scRNA-seq data. Q.W., T.Z., B.H., Y.C., and L.-Y.W. provided insights into the analytic strategies. L.S., B.Z., and Z.X. supervised the study. C.Z., T.R., B.Z., and Z.X wrote the manuscript with feedback from all other authors. All the authors read and approved the final manuscript.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
No animal experiments were involved in this study.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Pathogenic Microbiology and Immunology, School of Basic Medical Sciences, Xi'an Jiaotong University, Xi'an, Shaanxi, China. [2]Institute of Infection and Immunity, Translational Medicine Institute, Xi'an Jiaotong University Health Science Center, Xi'an, Shaanxi, China. [3]Key Laboratory of Environment and Genes Related to Diseases (Xi'an Jiaotong University), Ministry of Education, Xi'an, Shaanxi, China. [4]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. [5]School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, China. [6]Department of Biomedical Engineering, Oregon Health & Science University, Portland, OR, USA. [7]State Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan, China. [8]Department of Pathology and Laboratory Medicine, Oregon Health & Science University, Portland, OR, USA. [9]Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA. [10]Center for Biomedical Data Science, Oregon Health & Science University, Portland, OR, USA.

References
1. Meyer-Arndt L, Schwarz T, Loyal L, et al. Cutting edge: serum but not mucosal antibody responses are associated with pre-existing SARS-CoV-2 Spike Cross-Reactive CD4(+) T Cells following BNT162b2 Vaccination in the elderly. J Immunol. 2022;208(5):1001–5 (Baltimore, Md : 1950).
2. Nikolich-Zugich J. The twilight of immunity: emerging concepts in aging of the immune system. Nat Immunol. 2018;19(1):10–9.
3. Lopez-Otin C, Pietrocola F, Roiz-Valle D, Galluzzi L, Kroemer G. Meta-hallmarks of aging and cancer. Cell Metab. 2023;35(1):12–35.
4. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: an expanding universe. Cell. 2023;186(2):243–78.
5. Ramaswamy A, Brodsky NN, Sumida TS, et al. Immune dysregulation and autoreactivity correlate with disease severity in SARS-CoV-2-associated multisystem inflammatory syndrome in children. Immunity. 2021;54(5):1083-95 e7.
6. Castagnoli R, Votto M, Licari A, et al. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection in children and adolescents: a systematic review. JAMA Pediatr. 2020;174(9):882–9.
7. Crooke SN, Ovsyannikova IG, Poland GA, Kennedy RB. Immunosenescence and human vaccine immune responses. Immun Ageing. 2019;16:25.
8. Park JA, Cheung NV. Limitations and opportunities for immune checkpoint inhibitors in pediatric malignancies. Cancer Treat Rev. 2017;58:22–33.
9. Robinson JL, Kocabas P, Wang H, et al. An atlas of human metabolism. Sci Signal. 2020;13(624):eaaz1482.
10. Arrojo EDR, Lev-Ram V, Tyagi S, et al. Age mosaicism across multiple scales in adult tissues. Cell Metab. 2019;30(2):343-51 e3.
11. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell. 2013;49(2):359–67.
12. Klemera P, Doubal S. A new approach to the concept and computation of biological age. Mech Ageing Dev. 2006;127(3):240–8.
13. Zhu H, Chen J, Liu K, 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. 2023;9(26):eabq7599.
14. Peters MJ, Joehanes R, Pilling LC, et al. The transcriptional landscape of age in human peripheral blood. Nat Commun. 2015;6:8570.
15. McInnes IB, Gravallese EM. Immune-mediated inflammatory disease therapeutics: past, present and future. Nat Rev Immunol. 2021;21(10):680–6.
16. Huang Z, Chen B, Liu X, et al. Effects of sex and aging on the immune cell landscape as assessed by single-cell transcriptomic analysis. Proc Natl Acad Sci U S A. 2021;118(33):e2023216118.
17. Hashimoto K, Kouno T, Ikawa T, et al. Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. Proc Natl Acad Sci U S A. 2019;116(48):24242–51.
18. Luo OJ, Lei W, Zhu G, et al. Multidimensional single-cell analysis of human peripheral blood reveals characteristic features of the immune system landscape in aging and frailty. Nature Aging. 2022;2(4):348–64.
19. Tuo Y, Zhang Z, Tian C, et al. Anti-inflammatory and metabolic reprogramming effects of MENK produce antitumor response in CT26 tumor-bearing mice. J Leukoc Biol. 2020;108(1):215–28.
20. Vaena S, Chakraborty P, Lee HG, et al. Aging-dependent mitochondrial dysfunction mediated by ceramide signaling inhibits antitumor T cell response. Cell Rep. 2021;35(5):109076.
21. Zhang L, Dong X, Lee M, Maslov AY, Wang T, Vijg J. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. Proc Natl Acad Sci U S A. 2019;116(18):9014–9.
22. Blacher E, Tsai C, Litichevskiy L, et al. Aging disrupts circadian gene regulation and function in macrophages. Nat Immunol. 2022;23(2):229–36.

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 20 of 21

23. Paula C, Motta A, Schmitz C, Nunes CP, Souza AP, Bonorino C. Alterations in dendritic cell function in aged mice: potential implications for immunotherapy design. Biogerontology. 2009;10(1):13–25.

24. O'Neill DE, Forman DE. Cardiovascular care of older adults. BMJ. 2021;374:n1593.

25. Orsini G, Legitimo A, Failli A, Massei F, Biver P, Consolini R. Enumeration of human peripheral blood dendritic cells throughout the life. Int Immunol. 2012;24(6):347–56.

26. Shehata HM, Hoebe K, Chougnet CA. The aged nonhematopoietic environment impairs natural killer cell maturation and function. Aging Cell. 2015;14(2):191–9.

27. Zeng D, Ye Z, Shen R, et al. IOBR: multi-omics immuno-oncology biological research to decode tumor microenvironment and signatures. Front Immunol. 2021;12:687975.

28. Schubert M, Klinger B, Klunemann M, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun. 2018;9(1):20.

29. Shaw AC, Goldstein DR, Montgomery RR. Age-dependent dysregulation of innate immunity. Nat Rev Immunol. 2013;13(12):875–87.

30. Frasca D, Blomberg BB. Aging affects human B cell responses. J Clin Immunol. 2011;31(3):430–5.

31. Cancro MP, Hao Y, Scholz JL, et al. B cells and aging: molecules and mechanisms. Trends Immunol. 2009;30(7):313–8.

32. Guerrettaz LM, Johnson SA, Cambier JC. Acquired hematopoietic stem cell defects determine B-cell repertoire changes associated with aging. Proc Natl Acad Sci U S A. 2008;105(33):11898–902.

33. Carrasco E, Gómez de Las Heras MM, Gabandé-Rodríguez E, Desdín-Micó G, Aranda JF, Mittelbrunn M. The role of T cells in age-related diseases. Nat Rev Immunol. 2022;22(2):97–111.

34. Oh DY, Kwek SS, Raju SS, et al. Intratumoral CD4(+) T Cells mediate antitumor cytotoxicity in human bladder cancer. Cell. 2020;181(7):1612–25 e13.

35. Wang Y, Chen Z, Wang T, et al. A novel CD4+ CTL subtype characterized by chemotaxis and inflammation is involved in the pathogenesis of Graves' orbitopathy. Cell Mol Immunol. 2021;18(3):735–45.

36. Patil VS, Madrigal A, Schmiedel BJ, Clarke J, O'Rourke P, de Silva AD, Harris E, Peters B, Seumois G, Weiskopf D, Sette A, Vijayanand P. Precursors of human CD4+ cytotoxic T lymphocytes identified by single-cell transcriptome analysis. Sci Immunol. 2018;3(19):eaan8664.

37. Mogilenko DA, Shchukina I, Artyomov MN. Immune ageing at single-cell resolution. Nat Rev Immunol. 2022;22(8):484–98.

38. Le Garff-Tavernier M, Beziat V, Decocq J, et al. Human NK cells display major phenotypic and functional changes over the life span. Aging Cell. 2010;9(4):527–35.

39. Goronzy JJ, Weyand CM. Mechanisms underlying T cell ageing. Nat Rev Immunol. 2019;19(9):573–83.

40. Artyomov MN, Van den Bossche J. Immunometabolism in the Single-Cell Era. Cell Metab. 2020;32(5):710–25.

41. Kelly B, Pearce EL. Amino Assets: How Amino Acids Support Immunity. Cell Metab. 2020;32(2):154–75.

42. Sacco K, Castagnoli R, Vakkilainen S, Liu C, Delmonte OM, Oguz C, et al. Immunopathological signatures in multisystem inflammatory syndrome in children and pediatric COVID-19. Nat Med. 2022;28(5):1050–62.

43. Boxx GM, Cheng G. The Roles of Type I Interferon in Bacterial Infection. Cell Host Microbe. 2016;19(6):760–9.

44. Sadler AJ, Williams BR. Interferon-inducible antiviral effectors. Nat Rev Immunol. 2008;8(7):559–68.

45. Coccia EM, Battistini A. Early IFN type I response: Learning from microbial evasion strategies. Semin Immunol. 2015;27(2):85–101.

46. McCrindle BW, Rowley AH, Newburger JW, et al. Diagnosis, treatment, and long-term management of Kawasaki Disease: a scientific statement for health professionals From the American Heart Association. Circulation. 2017;135(17):e927–99.

47. Wang Z, Xie L, Ding G, et al. Single-cell RNA sequencing of peripheral blood mononuclear cells from acute Kawasaki disease patients. Nat Commun. 2021;12(1):5444.

48. Zhu Q, Dong Q, Wang X, et al. Palmitic acid, a critical metabolite, aggravates cellular senescence through reactive oxygen species generation in Kawasaki Disease. Front Pharmacol. 2022;13:809157.

49. Makino N, Nakamura Y, Yashiro M, et al. Epidemiological observations of Kawasaki disease in Japan, 2013–2014. Pediatr Int. 2018;60(6):581–7.

50. Furukawa S, Matsubara T, Jujoh K, et al. Reduction of peripheral blood macrophages/monocytes in Kawasaki disease by intravenous gamma-globulin. Eur J Pediatr. 1990;150(1):43–7.

51. Katayama K, Matsubara T, Fujiwara M, Koga M, Furukawa S. CD14+CD16+ monocyte subpopulation in Kawasaki disease. Clin Exp Immunol. 2000;121(3):566–70.

52. Terekhova M, Swain A, Bohacova P, et al. Single-cell atlas of healthy human blood unveils age-related loss of NKG2C+GZMB–CD8+ memory T cells and accumulation of type 2 memory T cells. Immunity. 2023;56(12):2836-54.e9.

53. Rossi DJ, Jamieson CH, Weissman IL. Stems cells and the pathways to aging and cancer. Cell. 2008;132(4):681–96.

54. Nyugen J, Agrawal S, Gollapudi S, Gupta S. Impaired functions of peripheral blood monocyte subpopulations in aged humans. J Clin Immunol. 2010;30(6):806–13.

55. Ni J, Miller M, Stojanovic A, Garbi N, Cerwenka A. Sustained effector function of IL-12/15/18-preactivated NK cells against established tumors. J Exp Med. 2012;209(13):2351–65.

56. Mahmoudzadeh S, Nozad Charoudeh H, Marques CS, Bahadory S, Ahmadpour E. The role of IL-12 in stimulating NK cells against Toxoplasma gondii infection: a mini-review. Parasitol Res. 2021;120(7):2303–9.

57. Ratliff M, Alter S, Frasca D, Blomberg BB, Riley RL. In senescence, age-associated B cells secrete TNFalpha and inhibit survival of B-cell precursors. Aging Cell. 2013;12(2):303–11.

58. Buffa S, Pellicano M, Bulati M, et al. A novel B cell population revealed by a CD38/CD24 gating strategy: CD38(-)CD24 (-) B cells in centenarian offspring and elderly people. Age (Dordr). 2013;35(5):2009–24.

59. Kartberg F, Asp L, Dejgaard SY, et al. ARFGAP2 and ARFGAP3 are essential for COPI coat assembly on the Golgi membrane of living cells. J Biol Chem. 2010;285(47):36709–20.

60. Nguyen TA, Debnath J. Unconventional secretion: cargo channeling by TMED10. Cell Res. 2020;30(9):713–4.

61. Yanes RE, Gustafson CE, Weyand CM, Goronzy JJ. Lymphocyte generation and population homeostasis throughout life. Semin Hematol. 2017;54(1):33–8.

62. Basso K, Dalla-Favera R. Germinal centres and B cell lymphomagenesis. Nat Rev Immunol. 2015;15(3):172–84.

63. Palmer DB. The effect of age on thymic function. Front Immunol. 2013;4:316.

64. Berent-Maoz B, Montecino-Rodriguez E, Dorshkind K. Genetic regulation of thymocyte progenitor aging. Semin Immunol. 2012;24(5):303–8.

65. Goronzy JJ, Weyand CM. Successful and maladaptive T cell aging. Immunity. 2017;46(3):364–78.

66. Cachot A, Bilous M, Liu YC, et al. Tumor-specific cytolytic CD4 T cells mediate immunity against human cancer. Sci Adv. 2021;7(9):eabe348.

67. Weiskopf D, Bangs DJ, Sidney J, et al. Dengue virus infection elicits highly polarized CX3CR1+ cytotoxic CD4+ T cells associated with protective immunity. Proc Natl Acad Sci U S A. 2015;112(31):E4256–63.

68. Belsky DW, Moffitt TE, Cohen AA, et al. Eleven telomere, epigenetic clock, and biomarker-composite quantifications of biological aging: do they measure the same thing? Am J Epidemiol. 2018;187(6):1220–30.

69. Kim S, Myers L, Wyckoff J, Cherry KE, Jazwinski SM. The frailty index outperforms DNA methylation age and its derivatives as an indicator of biological age. Geroscience. 2017;39(1):83–92.

70. Zhang Y, Saum KU, Schottker B, Holleczek B, Brenner H. Methylomic survival predictors, frailty, and mortality. Aging (Albany NY). 2018;10(3):339–57.

71. Vetter VM, Meyer A, Karbasiyan M, Steinhagen-Thiessen E, Hopfenmuller W, Demuth I. Epigenetic clock and relative telomere length represent largely different aspects of aging in the Berlin Aging Study II (BASE-II). J Gerontol A Biol Sci Med Sci. 2019;74(1):27–32.

72. Sayed N, Huang Y, Nguyen K, et al. An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging. Nat Aging. 2021;1:598–615.

73. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289–96.

74. Stuart T, Butler A, Hoffman P, et al. Comprehensive integration of single-cell data. Cell. 2019;177(7):1888-902 e21.

75. Kang JB, Nathan A, Weinand K, et al. Efficient and precise single-cell reference atlas mapping with Symphony. Nat Commun. 2021;12(1):5890.

Zhang *et al. Immunity & Ageing* (2024) 21:74

Page 21 of 21

76. Aibar S, Gonzalez-Blas CB, Moerman T, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. 2017;14(11):1083–6.
77. Andreatta M, Carmona SJ. UCell: Robust and scalable single-cell gene signature scoring. Comput Struct Biotechnol J. 2021;19:3796–8.
78. Foroutan M, Bhuva DD, Lyu R, Horan K, Cursons J, Davis MJ. Single sample scoring of molecular phenotypes. BMC Bioinformatics. 2018;19(1):404.
79. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14: 7.
80. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics. 2012;28(4):573–80.
81. Lydon EC, Henao R, Burke TW, et al. Validation of a host response test to distinguish bacterial and viral respiratory infection. EBioMedicine. 2019;48:453–61.
82. Bhattacharya S, Dunn P, Thomas CG, et al. ImmPort, toward repurposing of open access immunological assay data for translational and clinical research. Sci Data. 2018;5(1):1.
83. Reyes M, Filbin MR, Bhattacharyya RP, et al. An immune-cell signature of bacterial sepsis. Nat Med. 2020;26(3):333–40.
84. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell. 2015;160(1–2):48–61.
85. Gulati GS, Sikandar SS, Wesche DJ, et al. Single-cell transcriptional diversity is a hallmark of developmental potential. Science. 2020;367(6476):405–11.
86. Wu Y, Yang S, Ma J, et al. Spatiotemporal Immune Landscape of Colorectal Cancer Liver Metastasis at Single-Cell Level. Cancer Discov. 2022;12(1):134–53.
87. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14(10):979–82.
88. Zheng L, Qin S, Si W, et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. Science. 2021;374(6574):abe6474.
89. Domínguez Conde C, Xu C, Jarvis LB, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science. 2022;376(6594):eabl5197.
90. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
91. Aliee H, Theis FJ. AutoGeneS: Automatic gene selection using multi-objective optimization for RNA-seq deconvolution. Cell Syst. 2021;12(7):706–15 e4.

Publisher's Note