

AgeAnno: a knowledgebase of single-cell annotation of aging in human

Kexin Huang^{1,4}, Hoaran Gong^{1,4}, Jingjing Guan², Lingxiao Zhang², Changbao Hu², Weiling Zhao³, Liyu Huang², Wei Zhang^{1,4}, Pora Kim^{3,*} and Xiaobo Zhou^{3,5,6,*}

¹West China Biomedical Big Data Centre, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, P.R. China, ²School of Life Science and Technology, Xidian University, Xi'an, Shaanxi 710071, P.R. China, ³Center for Computational Systems Medicine, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA, ⁴Med-X Center for Informatics, Sichuan University, Chengdu, Sichuan 610041, P.R. China, ⁵McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA and ⁶School of Dentistry, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

Received August 16, 2022; Revised September 14, 2022; Editorial Decision September 18, 2022; Accepted September 22, 2022

ABSTRACT

Aging is a complex process that accompanied by molecular and cellular alterations. The identification of tissue-/cell type-specific biomarkers of aging and elucidation of the detailed biological mechanisms of aging-related genes at the single-cell level can help to understand the heterogeneous aging process and design targeted anti-aging therapeutics. Here, we built AgeAnno (<https://relab.xidian.edu.cn/AgeAnno/#/>), a knowledgebase of single cell annotation of aging in human, aiming to provide comprehensive characterizations for aging-related genes across diverse tissue-cell types in human by using single-cell RNA and ATAC sequencing data (scRNA and scATAC). The current version of AgeAnno houses 1 678 610 cells from 28 healthy tissue samples with ages ranging from 0 to 110 years. We collected 5580 aging-related genes from previous resources and performed dynamic functional annotations of the cellular context. For the scRNA data, we performed analyses include differential gene expression, gene variation coefficient, cell communication network, transcription factor (TF) regulatory network, and immune cell proportion. AgeAnno also provides differential chromatin accessibility analysis, motif/TF enrichment and footprint analysis, and co-accessibility peak analysis for scATAC data. AgeAnno will be a unique resource to systematically characterize aging-related genes across diverse tissue-cell types in human, and it could facilitate antiaging and aging-related disease research.

INTRODUCTION

Aging is defined as the gradual deterioration of functional characteristics in living organisms (1). The aging process is also among the most well-known risk factors for most human diseases (2). Currently, the aging population is growing at an unprecedented pace and there is an urgent need to initiate detailed studies and develop strategies for healthy aging. Abundant evidence indicates that aging is accompanied by complex molecular and cellular alterations, such as dysregulated intercellular communication, epigenetic and transcriptional alterations, genomic instability, and defects in the telomere maintenance machinery (3). For example, forkhead box O3 (FOXO3) is recognized as one of the key regulators of aging-related processes. The expression level of FOXO3 decreases with age (4). The interaction of FOXO3 with NR2E1, SNX3 and CEP5 plays a critical role in oxidative stress response and DNA damage during aging (5). High expression of FOXO3 regulated by the chromatin accessibility of its enhancer region, may have a protective effect on lifespan (6,7). Furthermore, FOXO3 is a potential target of resveratrol, which is a candidate agent for the prevention and treatment of aging and aging-related diseases (8,9). As shown here, a systematic analysis of individual aging genes and the development of a functional reference annotation will help to gain knowledge of the underlying mechanisms of aging and to identify novel therapeutic targets for anti-aging and aging-related diseases.

There are several available aging databases, such as Human Aging Genomic Resources, Aging Atlas, the Digital Ageing Atlas, AGEMAP and Open Genes (<https://open-genes.com/>) (10–13). These databases mainly focused on providing integrated biological datasets or a simple collection of aging-associated genes. In the era of personalized

*To whom correspondence should be addressed. Tel: +1 713 500 3923/3636; Email: Xiaobo.Zhou@uth.tmc.edu
Correspondence may also be addressed to Pora Kim. Email: Pora.Kim@uth.tmc.edu

medicine, with the accumulated genomic data of individuals, an intensive knowledge base annotating the potential mechanisms of individual aging genes is still lacking. Establishing a representative resource on aging will provide a foundation for the individual genomic differences in terms of aging. The high heterogeneity of molecular and cellular changes during aging pushes bulk sequencing techniques to their limits. Thanks to the advancement in single-cell sequencing technologies, single-cell sequencing data can refine our understanding of cellular heterogeneity and molecular alteration at the single-cell level in different tissues during aging. Here, we built AgeAnno, a knowledgebase of single-cell annotation of human aging, with the goal of providing comprehensive and intensive characterizations for human aging-related genes across diverse tissues and cell types by using single-cell RNA and single-cell ATAC sequencing data (scRNA and scATAC).

In this work, we annotated 5580 aging-related genes collected from six public resources based on 1 678 610 cells from 28 healthy tissue samples with ages ranging from 0 to 110 years. For scRNA data, first, we performed differential gene expression analysis for aging-related genes (DEGs) in each cell type and identified tissue- and cell-type-specific DEGs between different age groups. Then, we annotated aging-related DEGs into three aging-related biological categories, including immune genes, telomere maintenance genes, and circadian genes. To better understand the potential mechanisms of aging-related genes, we performed enrichment analysis and transcriptomic variation analysis to determine aging effect on gene transcriptional noise; cell–cell communication analysis based on the coordinated ligand–receptor (L–R) expression; and transcript factor (TF) regulatory network analysis. To explore the immune microenvironment alteration in the aging context, we investigated the immune cell proportions between different age groups. For scATAC data, we identified differential accessible regions (DARs) between age groups, and provided detailed genetic annotation information such as enhancers, promoters and introns. Then, we performed motif/TF enrichment analysis, footprint analysis, and co-accessibility peak analysis to provide functional annotations for aging-related DARs. We also identified gene–drug or gene–chemical interactions and related diseases for aging genes. As described above, AgeAnno provides intensive multiple functional annotations of aging context from single-cell data. We believe that AgeAnno will be a valuable resource that can advance the understanding of the underlying mechanisms of aging and help aging research communities develop anti-aging and aging-related disease treatment strategies.

MATERIALS AND METHODS

Data collection and preprocessing

First, we collected human aging-related genes from six reference aging-related resources, including Genage, AgingAtlas, Opengene, Digital ageing atlas, Gene Ontology, and an aging study by Chatsirisupachai *et al.* based on GTEx data (11,12,14–16). Then, we systematically collected scRNA and scATAC data from previous studies. We searched previous studies from PubMed by using the keywords ‘hu-

man’, ‘single cell RNA sequencing’ and ‘single cell ATAC sequencing’. Only samples from healthy tissues were included. Corresponding age, tissue type, cell-specific marker genes and expression profile of each sample were extracted. Participants were divided into four age categories according to the human development stages by the United Nations and World Health Organization (WHO): childhood and adolescence period included 0- to 19-year-old people (Youth group in the database); adulthood period included 20- to 59-year-old people (Mid group in the database); old age period included 60- to 100-year-old people in scRNA data and 50- to 100-year old people in scATAC data (Old group in the database) and long-lived individuals with > 100 years (Supold group in the database). Overall, total 276 samples were included in the database. In scRNA data, the Youth group contains 32 samples, the Mid group contains 102 samples, the Old group contains 84 samples and the Supold group contains 7 samples. In scATAC data, the Mid group contains 12 samples and the Old group contains 38 samples. The age information of each sample was provided in Supplementary Table S1.

Briefly, for scRNA data, the quality control and preprocessing steps were performed using Seurat (v2.0), and batch effect removal was performed using Harmony (17,18). For scATAC data, we performed the quality control and preprocessing steps using ArchR, and the batch effect removal was performed using Harmony (18,19). The detailed parameters used in preprocessing are provided in Supplementary Table S2.

Cell clustering and annotation in scRNA and scATAC data

Cell clustering for scRNA data was performed using the Seurat package in R based on the normalized gene expression profiles (17). For scATAC data, the gene score matrix was calculated based on the accessibility of regulatory elements close to the genes to estimate gene expression for cell clustering using ArchR (19). Cell type annotation was performed according to the markers provided by the original papers. Cell cluster visualization of different cell types in age groups was performed using both t-distributed stochastic neighbor embedding (t-SNE) and the uniform manifold approximation and projection (UMAP) method.

Differential gene expression analysis for aging-related genes

We performed differential expression analysis to identify aging-related genes across different age groups (i.e. mid versus youth, old versus mid, old versus youth) using the FindMarkers function in the Seurat package for each cell type (17). Aging-related differentially expressed genes (DEGs) between age groups with $P < 0.05$ and $\log_{2}FC > (\text{mean}(\text{abs}(\log_{2}FC)) + 2 \times \text{sd}(\text{abs}(\log_{2}FC)))$ are shown in the database (20). We also identified tissue- and cell-type-specific DEGs. Moreover, we annotated DEGs into three aging mechanism context-based categories, including immune genes, telomere maintenance (TERT) genes, and circadian genes. Immune genes were obtained from InnateDB (<https://www.innatedb.com/>), TERT genes were obtained from TelNet (<http://www.cancertelsys.org/telnet/>) and circadian genes were obtained from the Circadian Gene Database (<http://cgdb.biocuckoo.org/>) (21–23).

Aging-related variation coefficient analysis

Transcriptional noise describes molecular fluctuations that lead to the variability of gene expression within a cell population (24). Prior studies found that transcriptional control showed age-dependent deterioration and instability, which led to increased transcriptional noise (25). To identify the effects of aging on aging-related DEGs in different cell types, we performed variation coefficient analysis based on the expression of aging-related DEGs to measure the transcriptional noise as described in previous studies (26,27). Briefly, for a given gene, the variation coefficient was defined as the ratio of the standard deviation to the mean of expression difference between the old and young groups. The coefficient of variation for each aging-related DEG in multiple cell types are shown using a radar plot. For a gene in a given cell type, a higher variation coefficient represents a higher vulnerability to aging-related stress than those in other cell types.

Enrichment analysis for aging-related DEGs

To explore the potential biological function of aging-related DEGs, we performed Gene Ontology (GO) term enrichment analysis using clusterProfiler v3.15 package for the up- and downregulated DEGs in each cell and tissue type (28). Pathways containing more than five genes with a significance level of $P < 0.05$ are provided on the webpage of AgeAnno.

Identifying cellular communication for aging-related DEGs

To systematically identify the cellular interaction networks, we performed cell–cell communication analysis utilizing CellPhoneDB v2.0 software for aging-related DEGs in each age group (29). CellPhoneDB is a knowledgebase that contains ligands, receptors and their interactions collected from published papers. The L–R interactions with a significant value of $P < 0.05$ were selected. Cell–cell communications are shown using interactive circular network plots.

Identifying the TF–target regulatory network for aging-related DEGs

We used the pySCENIC package to identify key TFs in the aging process according to a ‘three-step’ TF–target regulatory network construction (30). First, TFs and their target genes were extracted using gene inference methods based on the correlation of gene expression across cells. Then, target genes that were not enriched for a corresponding TF were deleted. Finally, the activity of these TFs was calculated via an enrichment score in each cell type (regulon specificity score, RSS). All parameters were set as default values. The information on TFs and their target genes are summarized by table, and the RSSs are shown using interactive radar plots.

Identifying immune cell proportions in the different age groups

Aging leads to profound changes in the immune system. For example, the changes in the proportion and the absolute number of T cells can reduce the immune response in

older individuals (31). The immune cell components were identified using Seurat, and immune proportions between different age groups are shown by using an interactive sunburst plot to analyze the aging impact on the immune microenvironment. All immune cell components are shown in Supplementary Table S3.

Identification of aging-related differentially accessible regions

It is challenging to perform downstream analyses directly since scATAC data are in a binary format (32). To address this, we first performed pseudo-bulk replicates preprocessing using the addGroupCoverages function in ArchR (19). This analysis simulated a bulk ATAC-seq experiment by grouping the data from cells together to generate data at the sample level. The pseudo-bulk replicates were then used to generate a peak matrix using the ArchR function addReproduciblePeakSet. We used MACS2 (v.2.2.7.1) to perform peak calling (33). We then compared peak sets between different age groups using the ArchR function getMarkerFeatures to identify differential accessible regions (DARs). The threshold were set at $P < 0.05$ and $\log_2FC > 1$. After comparison, we used the ChIPseeker (v.1.30.3) package to annotate the differential peaks to the nearest genes (34). Only DARs annotated with aging-related genes will be retained in the database.

Co-accessibility analysis

Co-accessibility was performed to investigate correlated interactions between different genomic regions, such as promoters, introns, and enhancers in aging-related DARs (35). Co-accessibility analysis was performed using the addCoAccessibility function of ArchR. The threshold was set at $\text{corCutOff} > 0.5$.

Motif/TF enrichment analysis and footprint analysis

TFs are highly involved in the aging process, and understanding the accessible variation of TFs will help to advance the knowledge of their roles during aging. scATAC allows for simultaneous analysis of motifs that are associated with the activities of related TFs. Therefore, motif/TF enrichment analysis was performed based on aging-related DARs using the addMotifAnnotations function in ArchR. The threshold was set at $P < 0.05$. Moreover, the activity of motif/TF in the different age groups was measured using the deviation score provided by chromVAR (36). Then, we plotted footprints for each TF in different age groups using the plotFootprints function in ArchR.

Drug and disease information

Disease information for aging-related genes was extracted from the DisGeNet v4.0 database (<https://www.disgenet.org>) (37). Drug/chemical-gene interaction information was derived from The Comparative Toxicogenomics Database (CTD, <http://ctdbase.org>), which contains >10 000 drug/chemical-gene interactions and detailed information on how drugs/chemicals can affect genes (38).

Manual curation of PubMed articles

We searched the publications from PubMed for 5580 aging-related genes. Taking FOS as an example, the keywords used for searching were ‘((FOS [Title/Abstract]) AND aging [Title/Abstract])’. After a manual review of the abstracts, we obtained 142 pieces of literature related to the function of corresponded gene in aging process.

Database construction

AgeAnno is freely available at <https://relab.xidian.edu.cn/AgeAnno/#/>. The front-end of AgeAnno website was developed using Vue 3.0 and Element Plus (<https://element-plus.org/>). The back-end of the website was developed using node.js and express (<https://expressjs.com/>). Data storage and management were performed using MySQL v5.7 (<https://www.mysql.com/>). ECharts v5.3.2 plugin software (<https://echarts.apache.org/zh/index.html>) was used to create interactive tables and results visualization. All upstream and downstream analyses were performed using R 4.1.2 and Python 2.7 based on the Linux system. All data in AgeAnno is available to the users in the ‘Download’ section. AgeAnno website can be visited on popular web browsers, such as Google Chrome, Firefox, Microsoft Edge, and Safari.

DATABASE CONTENT AND FEATURES

Overview of AgeAnno

AgeAnno aims to provide comprehensive characterizations for aging-related genes across diverse tissue-cell types in humans by using scRNA and scATAC data. After cell and gene quality control, 1 678 610 cells across 28 healthy tissue samples with ages ranging from 0 to 110 years were included. Cell counts and sample sizes for each tissue are shown in Supplementary Table S4. Figure 1A–1C shows the content and construction of AgeAnno. For scRNA data, after batch effect removal and quality control, we first identified 152 cell types for 17 tissues based on 432 marker genes. Differential gene expression analysis for aging-related genes identified 20 122 aging-related DEGs. Among them, 2 466 DEGs were tissue-specific and 6 728 genes were cell-type specific per tissue type. We also calculated the variation coefficient for aging-related DEGs in each cell type. Aging-related DEGs include 566 immune genes, 310 telomere maintenance genes and 299 circadian genes. Among aging-related DEGs, scATAC analysis showed that 1230 genes had increased accessibility and 966 genes had decreased accessibility. Enrichment analysis for aging-related DEGs identified 96 189 GO pathways, including 79 358 biological process pathways, 9 243 cellular component pathways and 7 588 molecular function pathways. Cell–cell communication analysis in different age groups identified 36 670 L–R pairs for aging-related DEGs. TF-regulatory network analysis identified 312 enriched TFs with 19 293 target genes. Among them, 105 TFs also showed enriched by using scATAC data and 1 548 target genes showed differential accessibility. Analysis of the changes in the proportion of immune cells found that the proportions of 71 immune cell types changed with aging.

For scATAC data, we identified 124 cell types for 11 tissues based on 179 marker genes. Analysis of aging-related

DARs identified 65 322 genes with increasing accessibility and 36 490 genes with decreasing accessibility. Compared with matched control tissues, 2 196 genes are DEGs. DAR annotation identified 49 438 promoters, 31 716 introns, 2 955 exons, 15 050 distal intergenic regions, 103 downstream regions, 2 316 3’ UTR regions and 234 5’ UTR regions. Co-accessibility analysis for DARs identified 6 844 co-accessible regions for decreasing accessible regions and 6 051 for increasing accessible regions across 11 tissues. Motif/TF enrichment analysis and footprint analysis identified 12 456 TFs. Among them, 7 099 TFs had increased accessible regions and 5 357 had decreased accessible regions following aging. Moreover, disease analysis identified 5 389 disease-related genes and chemicals- and drugs–gene interaction analysis identified 5 532 targetable genes that potentially interact with chemicals and drugs.

For example, ETS proto-oncogene 1 (ETS1) is one of the critical regulators in human aging (39). We found ETS1 was upregulated in immune cells in aged skin (Supplementary Figure S1A). According to scATAC data analysis, this may be due to the higher accessibility of the ETS1 promoter region. Krishnamurthy et al. found that an age-related increase in Ink4a/Arf expression was attributed to the expression of ETS1 (40). Ink4a and Arf are principal mediators of senescence in aging (40). Moreover, ETS1 was identified as a TF enriched in immune cells, and its target genes, such as DUSP4 were also upregulated, as shown in Supplementary Figure S1B and C. High expression of DUSP4 in the elderly impairs T cell-dependent B-cell responses with age (41). ETS1 is also involved in telomere maintenance (42). Chemical & drug analysis showed that ETS1 interacts with Estradiol (Supplementary Figure S1D), which can prevent collagen reduction and has anti-aging potential (43). Therefore, our database can be used to identify potential functions and the underlying mechanisms of aging-related genes and discover chemical-gene interactions, thereby driving the development of anti-aging strategies.

Additionally, AgeAnno also provides tools to manually curate aging-related data from published literature, including aging clock, brain age, anti-aging drug prediction and aging data procession toolkit. All related files in AgeAnno are available for free download on the ‘Download’ page through a github link.

Annotations and utility of AgeAnno

Cell map module. We performed cell clustering and annotation based on scRNA and scATAC data and provided both tSNE and UMAP methods for tissue-specific cell cluster visualization (Supplementary Figure S2A). In this module, AgeAnno also provided visualization of cells from different age groups (Supplementary Figure S2B). Cell numbers and cell-specific marker genes from different age groups are summarized by interactive bar charts and tables (Supplementary Figure S2C and D). In the Cell map module, users can choose the tissue type of interest and view cell proportions, marker genes, and global map of cell clusters. Here, using brain tissue as an example, we annotated six cell types based on marker genes, including excitatory neurons, inhibitory neurons, oligodendrocytes, endothelial cells, oligodendrocyte progenitor cells (OPCs), and astro-

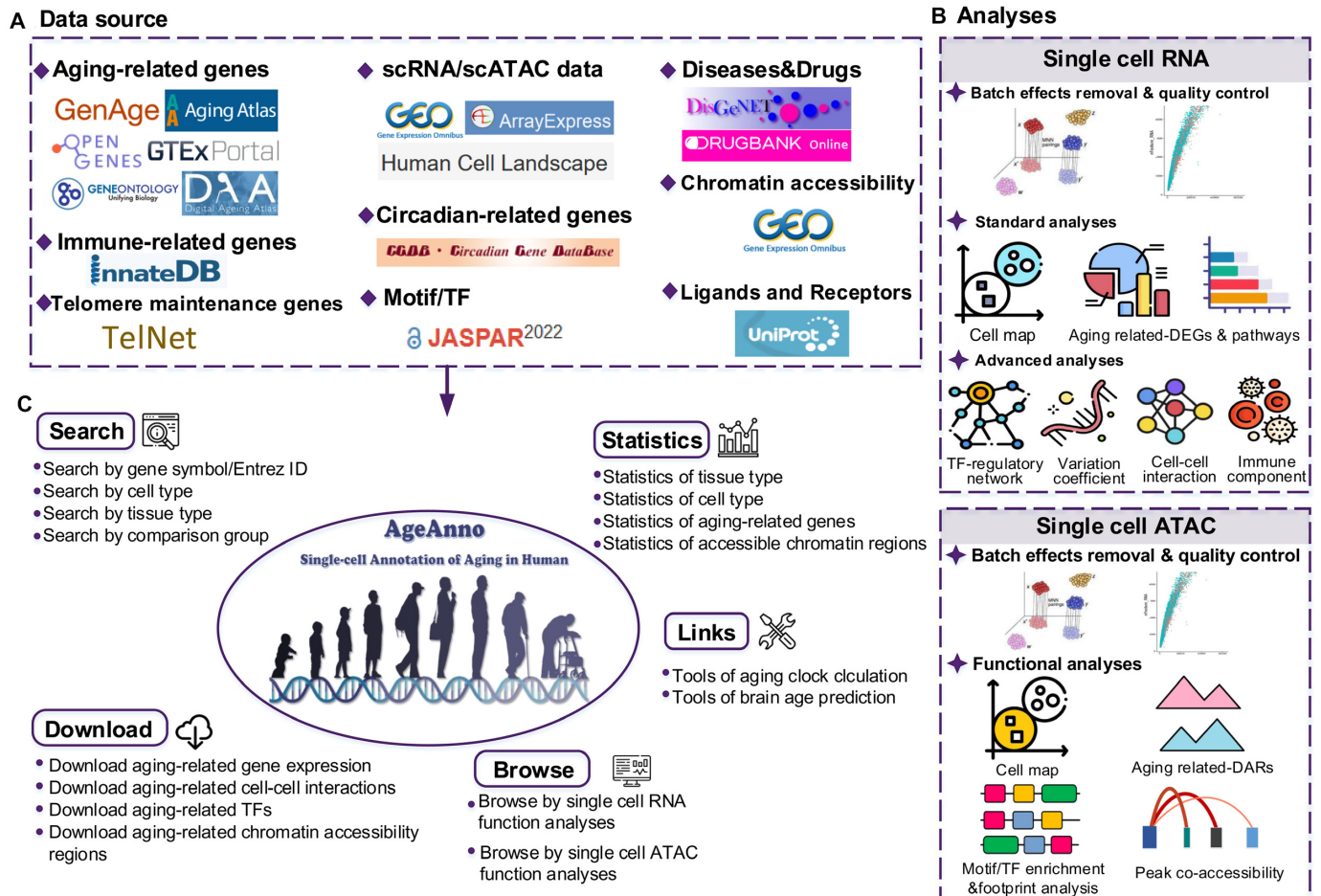


Figure 1. Database content and construction of AgeAnno. (A) Public resources used in AgeAnno. (B) Dynamic functional annotations of aging-related genes in AgeAnno. (C) User interface of AgeAnno. Users can perform data queries and browse through multiple paths. AgeAnno supports browsing, searching, and downloading information on aging-related genes.

cytes. As can be seen from the histograms, there were more excitatory neurons, oligodendrocytes and OPCs in the aged group, and more endothelial cells in the youth group (Supplementary Figure S2C).

Aging-related DEG module. The aging-related DEG module was developed to compare the expression of aging-related genes in different age groups. This module provides four search options, including gene symbol, age group, tissue type, or cell type (Supplementary Figure S3A). Detailed information was summarized by table. AgeAnno also provides the box plot for each DEG. The users can right-click to view the figures in a larger size and save them. This module also provides information on whether this gene is a tissue- or cell-type-specific (Supplementary Figure S3B). We found vasoactive intestinal peptide (VIP) is significantly downregulated in the inhibitory neurons following brain aging (Figure 2A), which is consistent with a previous report (44). VIP is one of the major regulatory peptides in the mammalian brain (45). Age-related down-regulation of VIP mRNA expression in the brain may potentially affect the excitatory-inhibitory balance, leading to learning and memory deficits in the aged population (44,46).

Aging-related variation coefficient module. Aging-dependent increases in transcriptional noise could affect the gene expression. To measure aging-related transcriptional noise, we performed variation coefficient analysis for aging-related genes in each cell type. In this module, users can input gene symbols, and compare groups or tissue types of interest (Supplementary Figure S4A). AgeAnno feeds back the information on the searched gene and radar plots of variation coefficients for individual cell types (Supplementary Figure S4B). Figure 2B shows a radar plot of the variation coefficient of VIP in six cell types of the brain. As shown in Figure 2C, the variation coefficient of VIP increased in all types of cells along with aging. This result suggests that dysregulation of transcriptional regulation may be the potential mechanism responsible for the age-dependent changes in VIP gene expression.

Aging-related pathway module. Aging-related pathway module was designed to identify biological functions for annotated aging-related genes. Users can search by gene symbol to view pathways related to this gene. They can also search by age group, tissue type or cell type of their interest (Supplementary Figure S5). Figure 2D shows the top 5

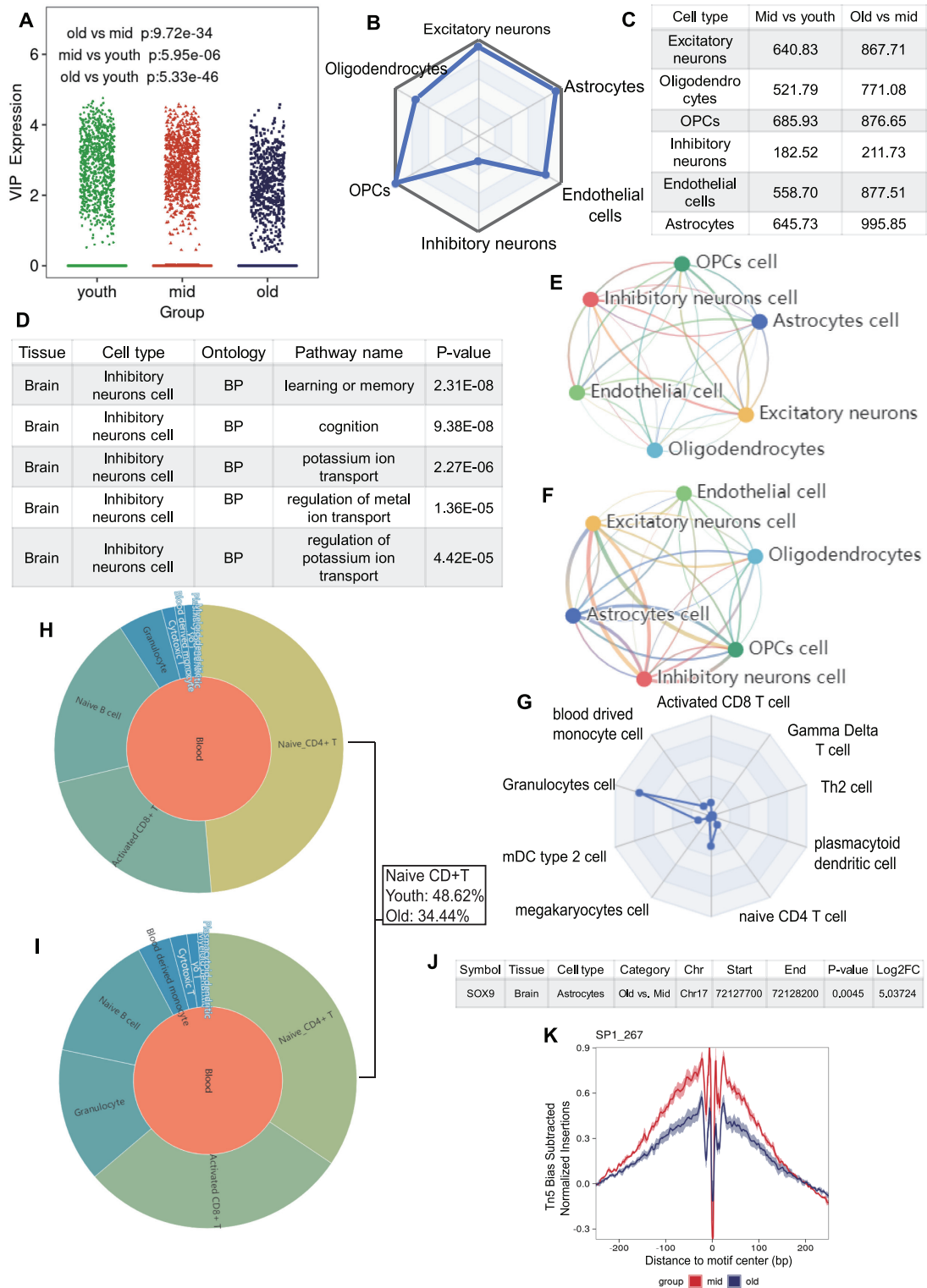


Figure 2. Examples from functional analyses of AgeAnno. **(A)** VIP expression in inhibitory neurons of different age groups. VIP is significantly downregulated with the aging process. **(B)** Radar plot of variation coefficient of VIP in each cell type between mid-group and youth group. **(C)** Comparison of VIP variation coefficient between different age groups (mid versus youth, old versus mid). **(D)** Top 5 enriched biological pathways of VIP. **(E)** Circular network plot of the brain of the mid group. **(F)** Circular network plot of the brain of the old group. Nodes with different colors represent different cell types, while the edges represent L–R interactions between two cell types. **(G)** RSS of FOS in different cell types. **(H)** Sunburst plot of immune cell components and proportion in blood in youth group. **(I)** Sunburst plot of immune cell components and proportion in the blood of the old group. **(J)** SOX9 showed significantly increased chromatin accessibility in astrocytes in aged brain. **(K)** TF footprint plot of different age groups. Red line represents the mid group and blue line represents old group.

enriched pathways related to VIP. We found that VIP exhibits a potential role in ion transportation and memory and cognition functions. Previous studies suggest that dysregulated VIP may have an effect on neuronal excitability through Kv4.2 channels (47) and age-related downregulation of VIP can lead to learning and memory deficits (48). As shown here, this module will help users identify potential biological functions and pathways of aging-related genes.

Cell-cell communication module. To better annotate the function of aging-related genes, we performed cell-cell communication network analysis through L-R pairs across different cell types. AgeAnno provides interactive networks for individual tissues. In the cell-cell communication network module, users can obtain the analysis results on the identified aging-associated L-R pairs in different age groups. Users can choose the tissue type for which they want to obtain the cell interaction networks in different age groups. In the network, nodes with different colors represent different cell types, while edges represent L-R interactions between the two cell types. The more interactions between two cells, the thicker the edges (Figure 2E and 2F). By clicking the nodes, users can browse detailed interactions related to this cell type (Supplementary Figure S6A). By clicking the edges, users can browse detailed interactions between two cell types (Supplementary Figure S6B). We observed the interplay between inhibitory and excitatory neurons via the L-R interaction of VIP-VIPR1. Previous studies showed that VIPR1 mediated the induction of Adenosine 3',5'-cyclic monophosphate (cAMP) signaling (49), and cAMP accumulation in the aged brain can open ion channels and weaken prefrontal neuronal firing (50). Our results indicate that altered cellular interactions might be a potential mechanism of aging.

TF regulatory network module. Growing evidence of transcriptional regulation dysfunction suggests that TFs play an important role in aging and age-related diseases. Therefore, AgeAnno provides TF regulatory network module for identifying TF-target regulatory networks across cell types. In the TF regulatory network module, users can obtain the potential TF regulators driving the age-dependent changes in expression. Users can input gene symbols, and age groups and tissue types of their interests (Supplementary Figure S7A and B). Users can also explore detailed information in AgeAnno by clicking TFs or target genes. For example, FOS is one of the TFs enriched in blood across several age groups, while it is differentially expressed in several cell types. Our analysis indicated that the RSS of FOS in granulocytes was higher than in other cell types, which suggests FOS may be a granulocytes-specific TF in age groups (Figure 2G). S100A9, one of the target genes of FOS, was upregulated in the aged group. Previous studies have shown that FOS is a critical regulator of the aging process (51). Aging is commonly associated with a state of chronic inflammation. Calgranulin B encoded by S100A9 gene could induce chronic inflammation during aging (52).

Immune cell proportion module. Altered proportions of immune cells may reflect immunosenescence in aging. AgeAnno provides an immune cell proportion module for

exploring the aging impact on the immune microenvironment. In this module, users can obtain immune cell proportion in different age groups. We found an age-dependent increase in the proportion of activated CD8+ T cells, whereas the proportion of naïve CD4+ T cells decreased (Figure 2H and 2I). This result is consistent with previous reports (53,54). Studies indicate that the age-dependent decline in naïve T cells weakens proliferative capacity and telomerase activity and leads to the progressive deterioration of the immune responses (55,56).

Aging-related DAR module. AgeAnno provides aging-related DAR module to detect altered chromatin accessibility for exploring underlying epigenetic mechanisms associated with aging-related genes. In the Aging-related DAR module, users can obtain the altered chromatin accessibility regions in aging. Users can input gene symbols or choose tissue or cell types of interest to obtain accessible regions (Supplementary Figure S8A). The search results are shown in Supplementary Figure S8B. As shown in Figure 2J, we found the chromatin accessibility of SOX9 is significantly increased in astrocyte cells of the aged brain. Moreover, the expression of SOX9 was also upregulated in astrocyte cells. This result suggests that the age-related overexpression of SOX9 may be attributed to increased chromatin accessibility. SOX9 has a critical role in aging and aging-related diseases, such as Alzheimer's disease (AD) (57). As an astrocyte-specific nuclear marker, SOX9 maintains astrocyte morphology and interactions with neurons (58,59). Dysregulation of SOX9 may be one of the potential mechanisms leading to astrocyte aging, which contributes to brain aging (60).

Co-accessibility link module. Interactions between genomic elements (i.e. promoter-enhancer interactions) play an important role in transcriptional regulation. AgeAnno provides Co-accessibility link module for co-accessibility analysis of aging-related DARs. In the co-accessibility link module, users can input gene symbols or choose tissue or cell types to obtain interacting genomic region pairs (Supplementary Figure S9A). Supplementary Figure S9B displays information on aging-related DARs and correlated genomic regions, as well as correlation coefficients between these regions (Supplementary Figure S9B).

Motif/TF enrichment module. AgeAnno provides Motif/TF enrichment module for extracting enriched TFs based on aging-related DARs to identify differentially TF accessible activity between age groups. Footprint images are also provided to visualize TF binding activities in different age groups. Moreover, the activity of cis-regulatory elements in different age groups were also estimated to support the TF binding activity. In this module, users can obtain TF accessible activities in different age groups. Users can input gene symbols or choose tissue or cell types to obtain TFs (Supplementary Figure S10). According to our analysis, we found that the DNA binding activity of SP1 is significantly decreased in the brain neurons, and the motif activity in middle age group also higher than old age group (Figure 2K). Previous studies reported altered binding activity of SP1 in the aging process (61). The biological

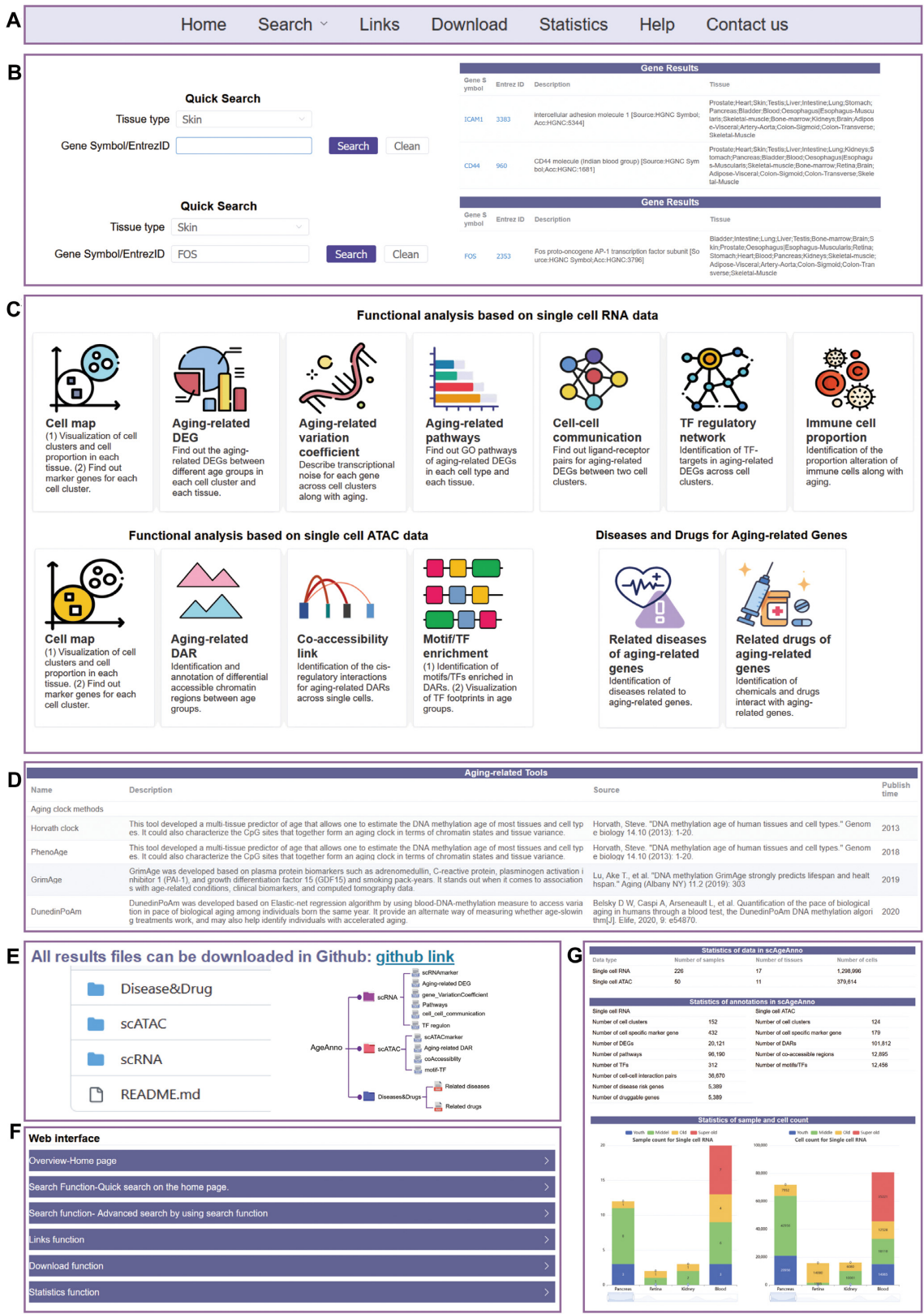


Figure 3. The main functions and usages of AgeAnno. (A) The top navigation bar of the main functions in AgeAnno. (B) On the 'Home' page, users can perform data queries through two paths: 'Search by gene symbol/Entrez ID' and 'Search by tissue type'. (C) The functional analyses included in AgeAnno. Users can query and visualize genes in different functional analysis categories. (D) The 'Link' function includes published aging-related algorithms and tools. (E) Data download function. (F) 'Help' function contains a brief introduction of AgeAnno and its functions. (G) Statistics information of samples and annotations in AgeAnno.

function of SP1 was widely reported in aging studies, such as age-dependent oxidative stress, nucleocytoplasmic trafficking and neuronal survival (62,63). Additionally, SP1 was identified as a pro-inflammatory transcription factor associated with increased AD risk (64).

Related human diseases and chemical/drug module. Disease information for aging-related genes was obtained from DisGeNet. A total of 23 501 diseases were identified to be associated with aging-related genes. These results are shown in the Related diseases of aging-related genes module. Users can search by gene symbol and disease name to obtain related information about diseases (Supplementary Figure S11A). Moreover, AgeAnno also provides gene–chemical/drug interactions extracted from the CTD and stores 12 576 drugs associated with aging-related genes. Users can search by gene symbol and drug name to obtain gene–drug interactions and detailed information of drugs (Supplementary Figure S11B). For example, ETS1 is a druggable target of Estradiol. Estradiol was found to have potential on preventing collagen reduction (43). As shown here, this module will provide potentially druggable targets for anti-aging studies.

Links module curated aging-related tools and methods. AgeAnno provides twelve published methods for estimating the aging clock that use omics data, such as transcriptomic or epigenetic data. Two methods for estimating brain age based on neuroimaging and genetic data, one anti-aging drug prediction method and one bioinformatics toolkit for data-mining (Supplementary Figure S12). Users can browse the basic information of these methods.

DISCUSSION AND FUTURE DEVELOPMENT

Currently, existing databases are mainly focused on providing integrated multi-omics datasets and collecting aging-associated candidate genes. AgeAnno is the first database that systematically characterizes aging-related genes across diverse tissue and cell types in humans by analyzing scRNA and scATAC data (Supplementary Table S5). To date, AgeAnno has cataloged 1 678 610 cells from 28 healthy tissue samples. Aging-related DEGs and DARs were identified and comprehensive functional annotations for these aging genes were obtained. AgeAnno houses 5 580 aging-related genes, including DEGs, DARs, TFs, ligands and receptors, etc. AgeAnno also provides cell clusters and proportions in different age groups. In addition, AgeAnno provides a user-friendly searching and browsing interface (Figure 3A). As shown in Figure 3B, ‘Quick Search’ on the ‘Home’ page allows users to perform data queries through two paths, including ‘search by tissue’ (choose one tissue of interest), and ‘search by gene’ (input gene symbol or gene EntrezID). All functional analyses are also available separately on the home page (Figure 3C). Users can obtain published aging-related tools and algorithms in the ‘Link’ function (Figure 3D). AgeAnno also provides results download, statistics and help functions (Figure 3E–3G).

To better serve the broad biomedical research communities, we will continue to integrate other single-cell sequencing data types, such as single-cell DNA sequencing and

single-cell TCR/BCR sequencing data, to identify the potential functions and mechanisms of aging-related genes. We plan to expand our database with an effective web server toolkit to facilitate cell type annotation, network modeling, anti-aging drug response evaluation and omics data integration. We believe that AgeAnno can continually provide a valuable resource for understanding molecular and cellular changes in aging biology and for identifying targetable biomarkers for aging research.

DATA AVAILABILITY

All data and results can be downloaded on the AgeAnno website.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank all our colleagues and friends who provided advice on this study. We would also like to thank the researchers who selflessly shared their data, and the database staff for their tremendous contributions to the collection and management of the data.

FUNDING

Huang et al., were supported by the 1-3-5 projects for disciplines of excellence—Clinical Research Incubation (2019HXFH022) and Center of Excellence-International Collaboration Initiative Grant (139170052), West China Hospital, Sichuan University. Dr. Zhao and Dr. Zhou were supported by NIHR01GM123037, U01AR069395-01A1, R01CA241930, and NSF 2217515. Dr. Kim was supported by NIHR35GM138184.

Conflict of interest statement. None declared.

REFERENCES

1. Strihler, B. (2012) In: *Times, Cells, and Aging*. Elsevier.
2. Niccoli, T. and Partridge, L. (2012) Ageing as a risk factor for disease. *Curr. Biol.*, **22**, R741–R752.
3. López-Otin, C., Blasco, M.A., Partridge, L., Serrano, M. and Kroemer, G. (2013) The hallmarks of aging. *Cell*, **153**, 1194–1217.
4. Zhao, Y. and Liu, Y.-S. (2021) Longevity factor FOXO3: a key regulator in aging-related vascular diseases. *Front. Cardiovasc. Med.*, **8**, 778674.
5. Donlon, T.A., Morris, B.J., Chen, R., Masaki, K.H., Allsopp, R.C., Willcox, D.C., Elliott, A. and Willcox, B.J. (2017) FOXO 3 longevity interactome on chromosome 6. *Aging Cell*, **16**, 1016–1025.
6. Eijkelenboom, A., Mokry, M., Smits, L.M., Nieuwenhuis, E.E. and Burgering, B.M. (2013) FOXO3 selectively amplifies enhancer activity to establish target gene regulation. *Cell Rep.*, **5**, 1664–1678.
7. Sanese, P., Forte, G., Disciglio, V., Grossi, V. and Simone, C. (2019) FOXO3 on the road to longevity: lessons from SNPs and chromatin hubs. *Comput. Struct. Biotechnol. J.*, **17**, 737–745.
8. Lin, C.-H., Lin, C.-C., Ting, W.-J., Pai, P.-Y., Kuo, C.-H., Ho, T.-J., Kuo, W.-W., Chang, C.-H., Huang, C.-Y. and Lin, W.-T. (2014) Resveratrol enhanced FOXO3 phosphorylation via synergetic activation of SIRT1 and PI3K/Akt signaling to improve the effects of exercise in elderly rat hearts. *Age*, **36**, 9705.
9. Zhou, J., Xue, Z., He, H.-N., Liu, X., Yin, S.-Y., Wu, D.-Y., Zhang, X., Schatten, H. and Miao, Y.-L. (2019) Resveratrol delays postovulatory aging of mouse oocytes through activating mitophagy. *Aging (Albany NY)*, **11**, 11504.

10. Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., Diana, E., Lehmann, G., Toren, D. and Wang, J. (2018) Human ageing genomic resources: new and updated databases. *Nucleic Acids Res.*, **46**, D1083–D1090.
11. Aging Atlas Consortium (2021) Aging atlas: a multi-omics database for aging biology. *Nucleic Acids Res.*, **49**, D825–D830.
12. Craig, T., Smelick, C., Tacutu, R., Wuttke, D., Wood, S.H., Stanley, H., Janssens, G., Savitskaya, E., Moskalev, A. and Arking, R. (2015) The digital ageing atlas: integrating the diversity of age-related changes into a unified resource. *Nucleic Acids Res.*, **43**, D873–D878.
13. Zahn, J.M., Poosala, S., Owen, A.B., Ingram, D.K., Lustig, A., Carter, A., Weeraratna, A.T., Taub, D.D., Gorospe, M. and Mazan-Mamczarz, K. (2007) AGEMAP: a gene expression database for aging in mice. *PLoS Genet.*, **3**, e201.
14. de Magalhaes, J.P. and Toussaint, O. (2004) GenAge: a genomic and proteomic network map of human ageing. *FEBS Lett.*, **571**, 243–247.
15. Gene Ontology Consortium. (2015) Gene ontology consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
16. Chatsirisupachai, K., Palmer, D., Ferreira, S. and de Magalhães, J.P. (2019) A human tissue-specific transcriptomic analysis reveals a complex relationship between aging, cancer, and cellular senescence. *Aging Cell*, **18**, e13041.
17. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.*, **36**, 411–420.
18. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. and Raychaudhuri, S. (2019) Fast, sensitive and accurate integration of single-cell data with harmony. *Nat. Methods*, **16**, 1289–1296.
19. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y. and Greenleaf, W.J. (2021) ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.*, **53**, 403–411.
20. Luo, Y., Zeng, G. and Wu, S. (2019) Identification of microenvironment-related prognostic genes in bladder cancer based on gene expression profile. *Front. Genet.*, **10**, 1187.
21. Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S. and Lynn, D.J. (2013) InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res.*, **41**, D1228–D1233.
22. Braun, D.M., Chung, I., Kepper, N., Deeg, K.I. and Rippe, K. (2018) TelNet—a database for human and yeast genes involved in telomere maintenance. *BMC Genet.*, **19**, 32.
23. Li, S., Shui, K., Zhang, Y., Lv, Y., Deng, W., Ullah, S., Zhang, L. and Xue, Y. (2016) CGDB: a database of circadian genes in eukaryotes. *Nucleic Acids Res.*, D397–D403.
24. Perez-Gomez, A., Buxbaum, J.N. and Petrascheck, M. (2020) The aging transcriptome: read between the lines. *Curr. Opin. Neurobiol.*, **63**, 170–175.
25. Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttill, R.A., Dollé, M.E., Calder, R.B., Chisholm, G.B., Pollock, B.H. and Klein, C.A. (2006) Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature*, **441**, 1011–1014.
26. Salzer, M.C., Lafzi, A., Berenguer-Llargo, A., Youssif, C., Castellanos, A., Solanas, G., Peixoto, F.O., Attolini, C.S.-O., Prats, N. and Aguilera, M. (2018) Identity noise and adipogenic traits characterize dermal fibroblast aging. *Cell*, **175**, 1575–1590.
27. Zou, Z., Long, X., Zhao, Q., Zheng, Y., Song, M., Ma, S., Jing, Y., Wang, S., He, Y. and Esteban, C.R. (2021) A single-cell transcriptomic atlas of human skin aging. *Dev. Cell*, **56**, 383–397.
28. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W. and Zhan, L. (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation*, **2**, 100141.
29. Efremova, M., Vento-Tormo, M., Teichmann, S.A. and Vento-Tormo, R. (2020) CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.*, **15**, 1484–1506.
30. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P. and Aerts, J. (2017) SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods*, **14**, 1083–1086.
31. Li, M., Yao, D., Zeng, X., Kasakovski, D., Zhang, Y., Chen, S., Zha, X., Li, Y. and Xu, L. (2019) Age related human t cell subset evolution and senescence. *Immun. Ageing*, **16**, 24.
32. Baker, S.M., Rogerson, C., Hayes, A., Sharrocks, A.D. and Rattray, M. (2019) Classifying cells with scasat, a single-cell ATAC-seq analysis tool. *Nucleic Acids Res.*, **47**, e10.
33. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
34. Yu, G., Wang, L.-G. and He, Q.-Y. (2015) ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, **31**, 2382–2383.
35. Baek, S. and Lee, I. (2020) Single-cell ATAC sequencing analysis: from data preprocessing to hypothesis generation. *Comput. Struct. Biotechnol. J.*, **18**, 1429–1439.
36. Schep, A.N., Wu, B., Buenrostro, J.D. and Greenleaf, W.J. (2017) chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods*, **14**, 975–978.
37. Piñero, J., Bravo, A., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., García-García, J., Sanz, F. and Furlong, L.I. (2016) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.*, **45**, D833–D839.
38. Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., Wiegiers, J., Wiegiers, T.C. and Mattingly, C.J. (2021) Comparative toxicogenomics database (CTD): update 2021. *Nucleic Acids Res.*, **49**, D1138–D1143.
39. Xiao, F., Yu, Q., Deng, Z., Yang, K., Ye, Y., Ge, M., Yan, D., Wang, H., Chen, X., Yang, L. *et al.* (2022) ETS1 acts as a regulator of human healthy aging via decreasing ribosomal activity. *Sci. Adv.*, **8**, eabf2017.
40. Krishnamurthy, J., Torrice, C., Ramsey, M.R., Kovalev, G.I., Al-Regaiey, K., Su, L. and Sharpless, N.E. (2004) Ink4a/Arf expression is a biomarker of aging. *J. Clin. Invest.*, **114**, 1299–1307.
41. Yu, M., Li, G., Lee, W.-W., Yuan, M., Cui, D., Weyand, C.M. and Goronzy, J.J. (2012) Signal inhibition by the dual-specific phosphatase 4 impairs t cell-dependent B-cell responses with age. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E879–E888.
42. Xu, X., Li, Y., Bharath, S.R., Ozturk, M.B., Bowler, M.W., Loo, B.Z.L., Tergaonkar, V. and Song, H. (2018) Structural basis for reactivating the mutant TERT promoter by cooperative binding of p52 and ETS1. *Nat. Commun.*, **9**, 3183.
43. Verdier-Sévrain, S., Bonté, F. and Gilchrist, B. (2006) Biology of estrogens in skin: implications for skin aging. *Exp. Dermatol.*, **15**, 83–94.
44. Mohan, A., Thalamuthu, A., Mather, K.A., Zhang, Y., Catts, V.S., Weickert, C.S. and Sachdev, P.S. (2018) Differential expression of synaptic and interneuron genes in the aging human prefrontal cortex. *Neurobiol. Aging*, **70**, 194–202.
45. Gozes, I. and Brenneman, D.E. (1989) VIP: molecular biology and neurobiological function. *Mol. Neurobiol.*, **3**, 201–236.
46. Loerch, P.M., Lu, T., Dakin, K.A., Vann, J.M., Isaacs, A., Geula, C., Wang, J., Pan, Y., Gabuzda, D.H. and Li, C. (2008) Evolution of the aging brain transcriptome and synaptic regulation. *PLoS One*, **3**, e3329.
47. Wood, K.C., Blackwell, J.M. and Geffen, M.N. (2017) Cortical inhibitory interneurons control sensory processing. *Curr. Opin. Neurobiol.*, **46**, 200–207.
48. Borbély, É., Scheich, B. and Helyes, Z. (2013) Neuropeptides in learning and memory. *Neuropeptides*, **47**, 439–450.
49. Johnson, G.C., May, V., Parsons, R.L. and Hammack, S.E. (2019) Parallel signaling pathways of pituitary adenylate cyclase activating polypeptide (PACAP) regulate several intrinsic ion channels. *Ann. N.Y. Acad. Sci.*, **1455**, 105–112.
50. Carlyle, B.C., Nairn, A.C., Wang, M., Yang, Y., Jin, L.E., Simen, A.A., Ramos, B.P., Bordner, K.A., Craft, G.E. and Davies, P. (2014) cAMP-PKA phosphorylation of tau confers risk for degeneration in aging association cortex. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 5036–5041.
51. Li, C.-W., Wang, W.-H. and Chen, B.-S. (2016) Investigating the specific core genetic-and-epigenetic networks of cellular mechanisms involved in human aging in peripheral blood mononuclear cells. *Oncotarget*, **7**, 8556.
52. Swindell, W.R., Johnston, A., Xing, X., Little, A., Robichaud, P., Voorhees, J.J., Fisher, G. and Gudjonsson, J.E. (2013) Robust shifts in

- S100a9 expression with aging: a novel mechanism for chronic inflammation. *Sci. Rep.*, **3**, 1215.
53. Dolfi, D.V., Mansfield, K.D., Polley, A.M., Doyle, S.A., Freeman, G.J., Pircher, H., Schmader, K.E. and Wherry, E.J. (2013) Increased T-bet is associated with senescence of influenza virus-specific CD8 t cells in aged humans. *J. Leukocyte Biol.*, **93**, 825–836.
 54. Linton, P.-J., Haynes, L., Klinman, N.R. and Swain, S.L. (1996) Antigen-independent changes in naive CD4 t cells with aging. *J. Exp. Med.*, **184**, 1891–1900.
 55. Moro-García, M.A., Alonso-Arias, R. and López-Larrea, C. (2013) When aging reaches CD4+ T-cells: phenotypic and functional changes. *Front. Immunol.*, **4**, 107.
 56. Nikolich-Zugich, J. (2005) T cell aging: naive but not young. *J. Exp. Med.*, **201**, 837–840.
 57. Morabito, S., Miyoshi, E., Michael, N., Shahin, S., Martini, A.C., Head, E., Silva, J., Leavy, K., Perez-Rosendahl, M. and Swarup, V. (2021) Single-nucleus chromatin accessibility and transcriptomic characterization of alzheimer's disease. *Nat. Genet.*, **53**, 1143–1155.
 58. Ung, K., Huang, T.-W., Lozzi, B., Woo, J., Hanson, E., Pekarek, B., Tepe, B., Sardar, D., Cheng, Y.-T. and Liu, G. (2021) Olfactory bulb astrocytes mediate sensory circuit processing through sox9 in the mouse brain. *Nat. Commun.*, **12**, 5230.
 59. Sun, W., Cornwell, A., Li, J., Peng, S., Osorio, M.J., Aalling, N., Wang, S., Benraiss, A., Lou, N. and Goldman, S.A. (2017) SOX9 is an astrocyte-specific nuclear marker in the adult brain outside the neurogenic regions. *J. Neurosci.*, **37**, 4493–4507.
 60. Neyrinck, K., Van Den Daele, J., Vervliet, T., De Smedt, J., Wierda, K., Nijs, M., Vanbokhoven, T., D'hondt, A., Planque, M. and Fendt, S.-M. (2021) SOX9-induced generation of functional astrocytes supporting neuronal maturation in an all-human system. *Stem Cell Rev. Rep.*, **17**, 1855–1873.
 61. Kumazaki, T. and Mitsui, Y. (1996) Alterations in transcription factor-binding activities to fibronectin promoter during aging of vascular endothelial cells. *Mech. Ageing Dev.*, **88**, 111–124.
 62. Kim, S.Y., Kang, H.T., Han, J.A. and Park, S.C. (2012) The transcription factor Sp1 is responsible for aging-dependent altered nucleocytoplasmic trafficking. *Ageing Cell*, **11**, 1102–1109.
 63. Dasari, A., Bartholomew, J.N., Volonte, D. and Galbiati, F. (2006) Oxidative stress induces premature senescence by stimulating caveolin-1 gene transcription through p38 mitogen-activated protein kinase/Sp1-mediated activation of two GC-rich promoter elements. *Cancer Res.*, **66**, 10805–10814.
 64. Citron, B.A., Saykally, J.N., Cao, C., Dennis, J.S., Runfeldt, M. and Arendash, G.W. (2015) Transcription factor Sp1 inhibition, memory, and cytokines in a mouse model of alzheimer's disease. *Am. J. Neurodegen. Dis.*, **4**, 40.