# Gene Expression Analysis Reveals Age and Ethnicity Signatures Between Young and Old Adults in Human PBMC

**Yang Hu** ( yanghu0325@163.com )
Jinan University School of Medicine https://orcid.org/0000-0002-8291-3182
**Yudai Xu**
Jinan University School of Medicine
**Lipeng Mao**
Jinan University School of Medicine
**Wen Lei**
Jinan University School of Medicine
**Jan Jian Xiang**
Jinan University School of Medicine
**Guodong Zhu**
South China University of Technology School of Medicine
**Yutian Hu**
meng yi centre limited
**Haitao Niu**
Jinan University School of Medicine
**Feng Gao**
Jinan University School of Medicine
**Lijuan Gao**
Jinan University School of Medicine
**Li`an Huang**
Jinan University First Affiliated Hospital
**Oscar Junhong Luo**
Jinan University School of Medicine
**Jinhai Duan**
Guangdong General Hospital: Guangdong Provincial People's Hospital
**Guobing Chen**
Jinan University School of Medicine

**Research Article**

**Keywords:** Immune aging, race and ethnicity, PBMC,WGCNA, biomarkers

**Posted Date:** August 3rd, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-746314/v1



---

**Version of Record:** A version of this preprint was published at Frontiers in Aging on February 3rd, 2022.
See the published version at https://doi.org/10.3389/fragi.2021.797040.

# Gene Expression Analysis Reveals Age and Ethnicity Signatures Between Young and Old Adults in Human PBMC

Yang Hu[1,2], Yudai Xu[1], Lipeng Mao[1,4], Wen Lei[1], Jan Jian Xiang[1], Guodong Zhu[5], Yutian Hu[6], Haitao Niu[7], Feng Gao[8], Lijuan Gao[1], Li`an Huang[2], Oscar Junhong Luo[1,4]*, Jinhai Duan[3]*, Guobing Chen[1,2]*

1, Institute of Geriatric Immunology, Department of Microbiology and Immunology, School of Medicine, Jinan University, Guangzhou, Guangdong, China.

2, Department of Neurology, the First Affiliated Hospital, Jinan University, Guangzhou, Guangdong, China.

3, Eastern Department of Neurology of Guangdong General Hospital， Guangdong Academy of Medical Sciences, Guandong, China.

4, Department of Systems Biomedical Sciences, School of Medicine, Jinan University, Guangzhou, Guangdong, China.

5, Department of Geriatrics, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, Guangdong, China.

6, Meng Yi Center Limited, Macau, China.

7, School of Medicine & Institute of Laboratory Animal Sciences, Jinan University, Guangzhou, Guangdong, China.

8, School of Medicine, Jinan University, Guangzhou, Guangdong, China.

*Correspondence should be addressed to: Drs. Oscar Junhong Luo, Jinhai Duan or Guobing Chen, Institute of Geriatric Immunology, School of Medicine, Jinan University, 601 Huangpu Avenue West, Guangzhou 510632, Guangdong Province, China. E-mails: luojh@jnu.edu.cn and guobingchen@jnu.edu.cn.

## Abstract

**Background:** Human immune system functions over an entire lifetime, yet how and why the immune system becomes less effective with age are not well understood. Therefore, the aim of this study is to exploit a large-scale population-based strategy to systematically identify genes and pathways differentially expressed as a function of chronological age. Despite the importance of age and race in shaping immune cell numbers and functions, it is unclear whether Asian and Caucasian immune systems go through similar gene expression changes throughout their lifespan, and to what extent these aging-associated variations are shared among ethnicities.

**Results:** Here, we characterize peripheral blood mononuclear cells transcriptome from 19 healthy adults of RNA-seq data and 153 healthy subjects of micoarray data with 21~90 years of age using the weighted gene correlation network analyses (WGCNA). These data reveal a set of insightful gene expression modules and representative gene biomarkers for human immune system aging from Asian and Caucasian ancestry, respectively. Among them, the aging-specific modules may show an age-related gene expression variation spike around early-seventies. In addition, we find the top hub genes including NUDT7, CLPB, OXNAD1 and MLLT3 are shared between Asian and Caucasian aging related modules and further validated in human PBMCs from different age groups.

**Conclusion:** Overall, our findings reveal how age and race differentially affect the immune systems between Asian and Caucasian, as well as discovered a common genetic variant that greatly impacts normal PBMC aging between Asian and Caucasian.

**Key words: Immune aging, race and ethnicity, PBMC, WGCNA, biomarkers**

## Introduction

Aging is a multifaceted process, involving numerous molecular and cellular mechanisms in the context of different organ systems [1]. A crucial component of aging is a set of functional and structural alterations in the immune system that can diminish the effectiveness of vaccinations, increase disease susceptibility, and contribute to mortality in older adults [2]. In addition to alterations in the stromal microenvironment in primary and secondary lymphoid organs, cell-intrinsic changes like cell numbers, ratio and function in both innate and adaptive immune cells play an important role in age-associated immune dysfunction. These alterations and transformations manifest themselves in increased morbidity and mortality of older organisms. However, the interplay between the PBMC age-related gene expression changes that affect the immune aging remains incompletely elucidated, and there is no clear understanding of which gene changes are primary, arising as a consequence of aging, and which might be secondary, adaptive or compensatory to the primary changes. Thus, a transcriptome analyses might lend greater insight than a static genetic investigation. In contrast to the relatively invariable genome sequence, the transcriptome is highly dynamic and changes in response to stimuli. Therefore, the aim of this study was to exploit a large-scale population-based strategy to systematically identify genes and pathways differentially expressed as a function of chronological age.

More importantly, analyses of human blood samples from different race and ethnicity uncovered significant aging-related variations in various subsets of PBMCs [3]. For example, a study on PBMC subsets characterized not only the presence of benign ethnic neutropenia among African Americans but further discovered a higher proportion of CD19+ cells and a lower proportion of CD3+ cells than in Caucasian population [4]. Moreover, the proportions of PBMCs' subpopulation in Asian cohorts were also different. Choong and colleagues observed that there were differences in cell counts for T, NK, and CD4+ cells as well as in the CD4/CD8 ratio among healthy Malaysians, Chinese, and Indians across the life span (18~71 years) [5]. In addition, Indians were significantly different from Malays and Chinese. Indians had higher T cells, higher CD4 cells, higher CD4/CD8 ratio, and lower NK cells; Chinese donors had lower B-cell levels than Malays and Indians [5]. Recent studies also have shown that there were differences in gene expression among European-derived and Asian-derived populations due to the common genetic variants in Epstein-Barr virus

(EBV)-transformed lymphoblastoid cell lines [6]. Despite the importance of age and race in shaping immune cell numbers and functions, it is not known whether some similarities of gene expression changes exist in Asian and Caucasian immune systems across the life span, and to what extent these aging-associated alternations are shared among Asian and Caucasian.

To study this, we compared the PBMC transcriptomes of healthy Asian and Caucasian adults with matching aging-stage by WGCNA method. Computational WGCNA analyses, differentially expressed genes (DEGs) and functional enrichments analyses revealed distinct immune system aging signatures in Asian and Caucasian. Our findings revealed how the aging differentially affects the immune systems between Asian and Caucasian, as well as discovered a common genetic variant that greatly impacts normal PBMC aging.

## Results

### Aging and Race Cause Transcriptomic Variations over Human Adult Lifespan

To identify main factors of variation in PBMC aging transcriptomic data, the microarray-based gene expression from 10 KIP provided by Lu *et al*. [8], was downloaded from the 10,000 immunomes project (10KIP, https://comphealth.ucsf.edu/app/10kimmunomes). This dataset contained quantile normalized genome-wide expression profiles of 153 adult human PBMC samples from young to old adults, including 65 young (ages 21~40: 24 men, 41 women), 40 middle-aged (ages 41~64: 18 men, 22 women), and 48 older subjects (65+: 20 men, 28 women) and containing three races including 19 Asian, 113 Caucasian and 21 Black or African American (Supplementary Fig. S1a, b; Supplementary Table 1). Then, we performed WGCNA analyses using expressed genes (n = 19,089) from 153 adult human PBMC samples. First, a soft thresholding power (β=6) with a scale-free topology fitting index of $R^2 > 0.80$ was defined to establish an adjacency matrix (Supplementary Fig. S2a). We used a tree-cutting algorithm to calculate the average linkage clustering, obtain gene co-expression modules, merge similar modules with a module eigengene over 0.75 (Supplementary Fig. S2b), and gradually build a co-expression network (Fig. 1a). Ultimately, the gene modules were identified. After that, the co-expression genes were clustered into 22 modules and labeled with different colors (Fig. 1a). To further quantify the correlation of genes in different modules, we calculated their pearson's correlation of the entire modules. Each module

showed independent validation to each other, and higher correlation indicated higher co-expression interconnectedness (Fig. 1b). Genes within the same module exhibited higher correlation than the genes between different modules.

In order to identify which factor mainly affected the variation in PBMC transcriptomes, the Pearson's correlation coefficients between modules and the trait of age, gender and race were calculated, respectively (Fig. 1c, Supplementary Table 2). Notably, among them, green module significantly correlated with age positively (Pearson's r=0.39, $p$=8.30e-07; Fig. 1c), while darkolivegreen module showed the negatively result (Pearson's r=0.45, $p$=2.02e-08; Fig. 1c). Moreover, sex and race individually showed a significant correlation in terms of the transcriptomic aging signatures. To investigate whether PBMC transcriptome changed gradually throughout the lifespan or rapidly at some stages, we analyzed various age brackets. Within each aging related module, we used the spline nonlinear regression for the transcriptomic profiles of the 153 samples in both green and darkolivegreen modules. Finally, our analyses indicated that PBMC transcriptome rapidly changed at two periods in adult lifespan: (i) a timepoint in early-forties, and (ii) a later timepoint after 70 years of age (Fig. 1d, e, Supplementary Table 3).

Besides, racial/ethnic differences in PBMC aging among adult lifespan were also important with profound effect on health. To determine the main causes of variation in transcriptomic data, we performed the principal component analyses (PCA) using expressed genes (n = 19,089) from high-quality samples. The first principal component (PC1) captured 14.9% of the variation in 153 microarray data and associated to age groups (Fig. 1f, Supplementary Table 4). The variants of PC1 score between Asian and Caucasian samples were more significantly different (Fig. 1g), whereas the difference in Asian and Black or Africa American is not significant (Fig. 1G). Besides, we used the genes in age-related modules (Pearson's $|r| \geqq 0.19$, $p \leqq 0.05$, modules including green, darkolivegreen, sienna3, darkmagenta, greenyellow, brown, turquoise and red) to performed the PCA analysis. Similarly, the variants of PC1 score between Asian and Caucasian samples were also significantly different (Supplementary Fig. S3a, b). Together, these results suggested that aging and race had both influenced the variation in PBMC transcriptomes, where it was not clear to what extent these aging-associated alternations were shared in different races, such as Asian and Caucasian.

## Aging-related PBMC Transcriptome Dynamics in Asian (Chinese)

We recruited 19 community dwelling healthy volunteers (10 female, 9 male) whose ages span 21~93 years old (Supplementary Table 1): 9 young (age 21~30: 5 female, 9 male), and 10 old donors (age 74~93: 5 female, 5 male), from Guangdong Province of China. Samples from different sexes in each assay were comparable in terms of age (males: ~81.4 vs. females: ~85.2; t-test p-value = 0.41) (Supplementary Table 5). Then, PBMCs transcriptome of these 19 donors were profiled using RNA-seq. As a result, 29,367 genes were selected after normalization of raw expression counts, excluding the genes with no expression in all samples. PCA of the 19 PBMC transcriptomes of Asian donors (Chinese) revealed that young and old samples were divided into two parts, and females had larger variants than males, especially in the old group (Fig. 2a). To further identify the related genes of PBMC aging, WGCNA was conducted using FPKM values of 29,367genes (FPKM > 1 of all sequenced transcript) and the trait of age and sex (Fig. 2b). Genes with the similar expression pattern were clustered into the same module to generate a cluster dendrogram (Fig. 2b). The sample dendrogram and trait heatmap were visualized to understand the relationship between the corresponding gene expression data and biological traits (Fig. 2c). Forty modules were obtained, of which four ME-modules (cyan, darkturquoise, orange, brown) showed significant correlations with age, with the absolute Pearson's correlation coefficient $|r| \geq 0.70$ ( $p < 0.01$) (Fig. 2c, Supplementary Table 6). Further, a consensus clustering also confirmed the four main modules were clearly separated between young to old (Fig. 2d). Similarly, based on the module eigengene (ME) expression profile and the ages of the donors, these four significant modules all showed sharp changes at the age of 74 (Fig. 2e, Supplementary Table 7). These results suggested these four gene modules were highly associated with chronological age in Asian (Chinese), especially for the brown (r=0.85, $p$=3.54E-06) and darkturquoise modules (r=0.77, $p$=9.90E-05).

## Novel and Known Age-associated Genes and Pathways Associated with PBMC Aging in Asian (Chinese)

WGCNA analyses defined that the module eigengene (ME) was the first principal component of a given module and could be considered as a representative of the module's gene expression profile. Based on ME expression profile of the four significant modules, the expression of cyan, darkturquoise and orange modules were

downregulated in young donors, while brown module showed the opposite results (Fig. 3a). To further explore the biological functions of the most closely age-related modules (brown module Pearson's r=0.85, $p$=3.54E-06; darkturquoise module Pearson's r=0.77, $p$=9.90E-05), we performed Gene Ontology (GO) enrichment analyses, as well as pathway ontology analyses by using clusterProfiler R package [13] (Supplementary Table 8). The enrichment analyses revealed that in the brown module, the top two enriched terms in GO ontology were "Cellular amino acid metabolic process" (FDR=5.74E-04) and "Negative regulation of neuron apoptotic process" (FDR=8.43E-04) (Fig. 3b); for the KEGG pathway analyses, the top enriched terms were "Herpes simplex virus 1 infection" (FDR= 9.19E-09) and "Valine, leucine and isoleucine degradation" (FDR=1.33E-03) (Fig. 3c). Meanwhile, functional annotations of darkturquoise module genes showed the top enriched terms in the GO databases were "Protein-DNA complex subunit organization" (FDR=7.68E-07) and "ncRNA processing" (FDR=1.33E-06) (Fig. 3b). Moreover, genes in darkturquoise module were found to be significantly enriched in protein export and lysine degradation signaling pathway (Fig. 3c). These findings together with previous research, which found persistent virus infections and metabolic dysregulation were closely related with immune aging [16], implied that the above signaling pathways might play an important role in aging.

To identify key genes associated with chronological age, we performed a more detailed analyses of the brown and darkturquoise modules. First, a total of 924 differentially expressed genes (DEGs) in the 19 Chinese PBMC transcriptomic data were found to be dysregulated in old individuals ($|\log FC| \geq 1$ and adjusted $p < 0.05$). Then, as shown in Venn diagram supplementary Fig. S4, 278 and 19 overlapping genes were extracted from the brown module and darkturquoise module in 19 Chinese PBMC DEG dataset, respectively. Subsequently, the protein protein interaction (PPI) network among the overlapping genes was established to identify potential aging related hub genes by using the STRING database. And then, based on the Maximal Clique Centrality (MCC) scores calculated by cytoscape, the top six highest-scored genes, including probable ATP-dependent RNA helicase (DDX27), Signal recognition particle subunit (SRP68) , E3 ubiquitin-protein ligase (RNF25), Transmembrane protein 131-like (TMEM131L), UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit (OGT), Early endosome antigen 1 (EEA1) , exhibiting the highest connections with other genes were identified for

further investigation (Fig. 3d, e). Strikingly, the mRNA abundance of these hub genes was significantly associated with chronological age (Fig. 3f, g). It was previously demonstrated TMEM131L could regulate immature single-positive thymocyte proliferation arrest by acting through mixed Wnt-dependent and -independent mechanisms [17]. Reports also demonstrated O-GlcNAc transferase (OGT) level was decreased in multiple aged tissues and suggested that dysregulation of OGT related O-GlcNAc formation might play an important role in the development of age-related diseases [18]. Researchers also reported the abundance of EEA1 proteins was altered in the brains of aged mice [19]. Moreover, SRP68 has been reported for its association with cellular senescence, while the ubiquitination-related genes RNF25 is not clear in immune aging. These data supported the notion that TMEM131L, OGT, EEA1, DDX27, SRP68 and RNF25 played important roles during PBMC aging, which might function as the novel candidate biomarkers of aging for Chinese individuals.

**Age-related Transcriptional Variation of Caucasians.**

Similarly, to investigate the aging-related gene modules in PBMC transcriptomes in Caucasian individuals, we performed WGCNA on microarray data from 113 Caucasian individuals, including 48 young (<40 years), 25 mid-age (40–65 years), and 41 health elderly (65–90 years). Then, a total of 16,376 genes from these transcriptomic data were used for this computation. Twenty major gene modules (Each module containing ≥ 160 genes) were identified. Similarly, we plotted the heatmap of module-trait relationships to evaluate the association between each module and the trait of age and sex (Fig. 4a; Supplementary Table 9). The results revealed that the brown and turquoise module were found to have the highest association with chronological age (brown module: r = 0.52, $p$ = 2.45e-09; turquoise module: r = -0.47, $p$ = 2.05e−07). More interestingly, these two aging-related modules showed two periods in the human lifespan during which the PBMC gene expression underwent abrupt changes: (i) a timepoint in early thirties, and (ii) a later timepoint after 65 years of age (Fig. 4b, c). GO functional enrichment analyses suggested that the brown and turquoise modules were mainly enriched in hormone transport and postsynaptic specialization, respectively (Fig. 4d, Supplementary table 10). Moreover, KEGG pathway enrichment analyses showed that the genes of brown module were mainly categorized into long-term depression and gap junction, while the turquoise module was mainly enriched in phototransduction and hedgehog signaling pathway (Fig. 4e). Next, we focused on the core genes of the brown and turquoise modules. By

using the differential expression analyses, we identified 1185 genes differentially expressed with chronological age in Caucasian, and 50 and 177 of these DEG genes were members of the brown and turquoise module, respectively (Fig. 4f). Subsequently, the 50 and 177 genes from brown and turquoise module were subject to hub gene identification using the STRING database, respectively (Fig. 4g). The results showed that the top two hub genes (Adenylate Cyclase 4, ADCY4; Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform, PIK3CA) in turquoise module were significantly down-regulated in PBMCs of old adults (Fig. 4h), whereas immunoglobulin superfamily DCC subclass member 2 (NEO1) from brown module showed the opposite result in the Caucasian cohorts (Fig. 4h). From the aging atlas website (https://bigd.big.ac.cn/aging/age_related_genes), ATP Pyrophosphate-Lyase 4 (ADCY4) and Serine/Threonine Protein Kinase (PIK3CA) have both involved in Longevity regulating pathway. As reported, ADCY4 catalyzed the formation of the signaling molecule cAMP in response to G-protein signaling [20], and PIK3CA participated in cellular signaling in response to various growth factors, which also involved in the activation of AKT1 upon stimulation by receptor tyrosine kinases ligands such as EGF, insulin, IGF1, VEGFA and PDGF [21]. Besides, Gulati *et al*. reported neogenin-1 (NEO1) was associated with the long-term HSCs (LT-HSCs) expand during age [22]. Taken together, these data also revealed that ADCY4, PIK3CA, and NEO1 were critical in aging, which might serve as the novel candidate biomarkers in Caucasian individuals during the PBMC aging.

**Shared Transcriptomic Signatures of Aging Between Asian (Chinese) and Caucasian.**

As age-expectation is ethnicity dependent [23], we sought to test whether gene expression in PBMC of aging individuals differed across racial/ethnic groups. The brown module from Asian (Chinese) and the turquoise module from Caucasian were both negatively correlated with chronological age. This naturally led us to compared these two modules for common expressed genes. Ninety-five genes were shared between Asian (Chinese) and Caucasian, despite thousands of race-specific genes associated with aging (2623 and 1688 genes in Asian and Caucasian, respectively; Fig. 5a). Functional annotation of these 95 shared genes revealed that they were highly enriched in the GO biological process of hindbrain development and coenzymeA metabolic process, as well as in the KEGG pathway of TGF-beta signaling (Fig. 5b, c). To uncover potential regulators of common transcriptomic variation in Asian (Chinese)

and Caucasian, we identified hub genes by using the STRING database. Accordingly, the top-scored genes, including peroxisomal coenzyme A diphosphatase (NUDT7) and caseinolytic peptidase B protein homolog (CLPB) were selected as the hub genes (Fig. 5d). Meanwhile, two genes (OXNAD1 and MLLT3) that shared between Caucasian and Asian (Chinese) aging-related modules showed common differential expression (Fig. 5e). These two potential aging-specific markers (OXNAD1 and MLLT3) were both downregulated in old Asian (Chinese) and Caucasian (Fig. 5f). These data uncovered that despite the stark contrast between races in aging-related gene expression pattern, our analyses were able to highlight shared aging biomarkers with common functional enrichment in Asian (Chinese) and Caucasian.

**Validated Shared Ggenes Involved in PBMC Aging**

After the 4 hub genes (NUDT7, CLPB, OXNAD1, MLLT3) together shared in Asian (Chinese) and Caucasian, we verified the expression levels of the hub genes among young and older individuals using the RNA-seq data and qPCR assay. Interestingly, these 4 hub genes were significantly down-regulated in old individuals compared with the youth in Asian (Chinese) and Caucasian (Fig. 6a, b). Meawhile, they were also found to be down-regulated in women during their lifespan in both Caucasian and Asian (Chinese) (Fig. 6c). To further investigate whether these 4 hub genes expressed differentially during PBMC aging, we collected samples from another 12 healthy volunteers residing in the Guangzhou, China, including 7 young adult (ages 21~30), and 5 aged healthy adults (ages 74+). We measured mRNA levels of these four hub genes (NUDT7, CLPB, OXNAD1 and MLLT3) in extracts of PBMC from the 12 subjects. Similarly, the mRNA level of NUDT7, CLPB, OXNAD1 and MLLT3 were both remarkably down-regulated in the elderlies (Fig. 6d). All the above-mentioned observations confirmed down-expression of NUDT7, CLPB, OXNAD1 and MLLT3 is associated with PBMC aging in Asian (Chinese) and Caucasian.

**Discussion**

Investigating how genes jointly preserve or change in different races during human PBMC aging is important, yet challenging. Recently, Peters *et al*. identified 1,497 differentially expressed genes with chronological age by meta-analysis in 14,983 individuals' whole blood of European ancestry [24]. However, gene expression difference between Asian and Caucasian during PBMC aging is still unclear. Therefore, our study aims to fill this gap in Asian and Caucasian subjects. Besides,

WGCNA is an integrated bioinformatic analyses, which is characterized effectively and systematically to find modules and gene signatures highly related with the clinical trait, and provides a comprehensive characterization of the transcriptomic changes for disease's functional interpretation [9]. Thus, in our study, several significant gene modules with the same expression trends were identified by using WGCNA integrated bioinformatic analyses in 19 Asian (Chinese) and 113 Caucasian populations. As suggested in functional annotation analyses by the clusterProfiler package, these module genes were mainly enriched in virus infection, amino acid metabolic and differentiation, which were basic processes in aging mechanisms including dysregulation of herpes simplex virus 1 infection, valine, leucine and isoleucine degradation, long-term depression, gap junction, and hedgehog signaling pathway. Furthermore, according to MCC scores from the CytoHubba plugin in Cytoscape, the top chronological age-related genes were screened out (namely TMEM131L, OGT, EEA1, DDX27, SRP68 and RNF25 in Asian; ADCY4, PIK3CA and NEO1 in Caucasian). According to reports in the literature, all of these genes were more or less closely associated with aging. Consistent with these reports, the expression of these genes was also found be significantly regulated among young and old individuals in our study, supporting these genes might play a causal role in human PBMC aging. More importantly, our study revealed that although aging related transcriptomic alternations is a cumulative process throughout adult life, while there might existe two periods in the human lifespan during which the immune system underwent abrupt changes. The two breakpoints (30 and 65~70 years old) were much similar in Asian and Caucasian during their whole lifespan. A potential limitation of the two breakpoints is that we relied on a linear regression model to identify the time point which the immune system underwent abrupt changes. A recent study demonstrated that a quadratic regression model has a higher statistical fit to identify age-regulated expression trends in cross-sectional gene expression datasets [25]. So more complex models may be used to investigate the two breakpoints in future studie.

Despite well-characterized race differences in immune responses, disease susceptibility, and lifespan, it was unclear to what extend aging differentially affected peripheral blood cells of European and Asian ancestry. To fill this gap, we generated RNA-seq data in PBMCs from 19 age-matched healthy adults in Guangdong province of China and downloaded microarray data of 113 Caucasian PBMCs from 10 KIP (http://10kimmunomes.org/). By using WGCNA integrated bioinformatic analyses,

we discovered a gene expression signature of aging that was shared between Asian (Chinese) and Caucasian ancestry including: (1) 95 age-associated genes shared in 132 individuals. (2) four hub genes (NUDT7, CLPB, OXNAD1 and MLLT3) all decreased in old ages. According to reports in the literature about these four hub genes, NUDT7, acted as a coenzyme A (CoA) diphosphatase, which mediated the cleavage of CoA. NUDT7 functioned as a house-keeping enzyme by eliminating potentially toxic nucleotide metabolites, such as oxidized CoA from β-oxidation in the peroxisome, as well as nucleotide diphosphate derivatives, including NAD+, NADH, and ADP-ribose [26]. Furthermore downregulation of NUDT7 in mice accelerating senescence [27], was observed in the liver of starved mice [28]. Interestingly, OXNAD1 also known as oxidoreductase NAD-binding domain-containing protein, had been reported differentially expressed with chronological age [24]. And according to the uniprot anotation for CLPB, it might function as a regulatory ATPase and be related to secretion/protein trafficking process, involved in mitochondrial-mediated antiviral innate immunity, and activated RIG-I-mediated signal transduction and production of IFNB1 and proinflammatory cytokine IL6 [29]. Moreover, the hub gene of MLLT3 was a component of the superelongation complex and co-operated with DOT1L, which di/trimethylates H3K79 to promoted transcription [30, 31]. Recently, Vincenzo Calvanese, et.al, found MLLT3 could govern human haematopoietic stem-cell self-renewal and engraftment [32]. From above, NUDT7 and OXNAD1 both had an important role in cellular metabolism and aging, which was consistent with our finding of PBMC aging analyses, while the role of CLPB and MLLT3 in immune aging or senescence was unclear. Thus, by using co-expression networks, we identified new genes that were likely important in PBMC aging in Asian and Caucasian ancestry, opening new avenues of enquiry for future studies.

By WGCNA analyses, aging-specific regulatory modules and hub genes were identified in bulk PBMCs in Caucasian and Asian. Although this approach was effective in annotating the aging signatures, it was prone to biases in the differences of data quality and formats. Besides, as we had much smaller sample sizes for both PBMCs in Asian and the other ancestry groups, we used a nominal *P*-value threshold ($p<0.05$) in these specific sub-analyses. Larger sample sizes will ultimately be needed to fully understand the transferability of the aging-transcriptome signatures. And more importantly, further studies were needed to verify the important molecules, identified here (NUDT7, CLPB, OXNAD1 and MLLT3) as aging specific biomarkers of

immune system aging. Future studies might be needed to describe these race differences at single-cell resolution and in sorted cells and to establish their functional implications. Taken together, these findings indicated that aging played a critical role in human immune system aging and should be taken into consideration while searching for molecular targets and time frames for interventions/therapies to target aging and age-related diseases.

## Conclusion

Through WGCNA analysis of 153 cases of transcriptomic expression profile data, we find that there are two module gene sets that are significantly related to human immune aging. And interestingly, these two modules shows that the first peak of age-related gene expression changes occurs around 40 years old, and the second peak occurs around 70 years old. In addition, the influence of race on the expression of immune aging genes has clinical application value for the diagnosis and interpretation of immune aging. Thus, by using WGCNA analysis, we also focus on the similarities and differences in the changes in gene expression profiles of 19 Asian and 113 Caucasian PBMCs during aging. Finally, the hub genes NUDT7, CLPB, OXNAD1 and MLLT3, which are common shared in the aging process of Asian and Caucasian PBMCs. Finally, in this study, we focus on the effects of age and race on the gene expression of human peripheral blood mononuclear cells, which may provide important insights for future research on how to deal with immune decline.

## Methods

### Human Subjects

All studies were conducted following approval by the Ethics Committee of Jinan University (Approval #:KY-2020-027). Following informed consent, blood samples were obtained from 31 healthy volunteers residing in the Guangzhou, China region recruited by the First affiliated Hospital of Jinan University and Guangzhou First People's Hospital. For selecting the older adults 65 years and older, the eligibility criteria were in line with the 2019 NIH Policy on Inclusion Across the Lifespan (NOT-OD-18-116) [7]. Subjects were carefully screened and excluded if undergoing potentially a history of diseases and medications, as well as frailty according to the eligibility criteria [7]. Besides, smoking history data were also excluded in our study

and all donors were asked for consent for genetic research. Fresh blood was collected into heparin tubes and PBMCs were isolated by density gradient centrifugation using Ficoll-Paque Plus (GE) and washed with Ca/Mg-free PBS twice.

**RNA-seq Library Generation and Processing**

A total amount of 1 μg RNA per sample, isolated from PBMCs using the TRIzol (Invitrogen, United States), was used as input material for the RNA-seq sample preparations. Sequencing libraries were constructed using NEBNext® UltraTM RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's protocols. Final libraries were assessed on a Bioanalyzer DNA High Sensitivity Chip (Agilent Technologies). Paired-end sequencing (2 × 150 bp) of stranded total RNA libraries was carried out in Illumina NovaSeq 6000. The FASTQC tool was used to assessed the quality of the raw sequencing data , which computed read quality using summary of per-base quality defined using the probability of an incorrect base call. According to our quality criteria, reads with more than 30% of their nucleotides with a Phred score under 30 were removed, whereas samples with more than 20% of such low-quality reads were dropped from analyses. Reads from samples that passed the quality criteria were quality-trimmed and filtered using trimmomatic. High-quality reads were then used to estimate transcript abundance using RSEM. Finally, the estimate transcript abundance (read counts) were renormalized to include only protein-coding genes and all the downstream analyses were based on high quality data.

**Microarray Data Acquisition**

The microarray-based expression from 10 KIP provided by Lu *et al.* [8], was downloaded from the 10,000 immunomes project (10KIP, https://comphealth.ucsf.edu/app/10kimmunomes). This dataset contained quantile normalized genome-wide expression profiles of 153 adult human PBMC samples from young to old adults, including samples from 65 young (ages 21~40: 24 men, 41 women), 40 middle-aged (ages 41~64: 18 men, 22 women), and 48 older subjects (65+: 20 men, 28 women) and containing three races including 19 Asian, 113 Caucasian and 21 Black or Africa American.

## Identification of Key Co-expression Modules Using WGCNA

WGCNA R software package was applied to identify the co-expression modules of highly correlated genes among samples and related modules to external sample traits [9]. And a more detailed description of the WGCNA method was described in our previous study [10] .In brief, sample was clustered to recognize and remove outlier samples by the average linkage method. And then the optimum soft thresholding power (β) was selected to obtain a scale-free topology fitting index of >0.8. The soft thresholding power β=6 was used in the analysis of 153 microarray data set. Similarly, the soft thresholding power β=5 was in 113 Caucasian data set and the power β=6 in 19 Asian (Chinese). By using the soft thresholding power (β), the topological overlap matrix (TOM) and the corresponding dissimilarity matrix (1-TOM) were calculated, which was further used to classify the similar gene expressions into different gene co-expression modules [9]. Afterwards, highly similar dynamic modules were merged into larger ones at the cutline of 0.2. The module eigengenes (ME) was the first principal component of the expression matrix which represented the gene expression profile of the entire module. Afterwards the correlation between MEs and previously sample traits were assessed to identify the most relevant clinically significant module by Pearson's correlation analysis. Meanwhile, the most significant module was also validated by calculating the gene significance (GS) and module membership (MM) [9]. In our study, the gene expression data profiles of microarray data and RNA-seq profile were collected seperately to consturct gene co-expression networks by using the WGCNA package.

## Screening the Differentially Expressed Genes and Comparing with the Gene Modules of Interest

In order to find the differentially expressed genes (DEGs) between Asian and Caucasian, "limma" R package was applied in the Asian RNA-seq and Caucasian dataset to screen out DEGs, with the cut-off criteria of $|\log FC| \geq 0.50$ and adjusted $p < 0.05$ [11]. The DEGs of the Asian and Caucasian dataset were visualized by a volcano plot. Subsequently, the overlapping genes between DEGs and co-expression modules,

listed by the Venn diagram using the R package VennDiagram, were used to identify potential prognostic genes [12]. Benjamini-Hochberg *p*-value correction was used to select differentially expressed genes at an FDR adjusted *p*-value of 5%.

**Functional Annotation for the Modules of Interest**

For genes in each module, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted to analyze the biological functions of gene modules. The *clusterProfiler* package offered a gene classification method, namely groupGO, to classify genes based on their projection at a specific level of the GO corpus, and provided functions, *enrichGO* and *enrichKEGG*, to calculate enrichment test for GO terms and KEGG pathways based on hypergeometric distribution [13]. In our study, *clusterProfile*r package was used to perform the gene ontology (GO) and KEGG pathway enrichment analysis. Significant GO terms were defined with a adjusted $p < 0.05$ and count > 6. For the KEGG pathway analysis, the *enrichKEGG* function was utilized and adjusted $p < 0.01$ was set as a cutoff. We used the Benjamini-Hochberg FDR multiple test correction to assess significance of hypergeometric *p*-values.

**Network Analysis of Module Genes/Hub Genes**

The hub genes were identified based on protein interaction evidence from the STRING database (version 11.0; http://string-db.org/) [14]. The evidence of protein interaction network for key modules from the STRING database were retrieved by an interaction score with highest confidence (0.9). The Cytoscape plugin cytoHubba was used to rank nodes (genes) based on the Maxial clique centrality (MCC) topological method [15], and the top-5 genes were selected as hub genes for verification. MCC assumes that the node network is an undirected network; given a node $v$, S($v$) is the set of the maximal cliques containing $v$, and (| C | −1)! is the product of all positive integers less than | C |. The calculation is as follows:

$$MCC(V)=\sum C \in S(v)(|C|-1)!$$

**Validation of the Hub Genes**

In order to confirm the reliability of the hub genes, we tested the expression patterns of the hub genes from healthy individuals including 7 young (ages: 23~30) and 5 old (ages: ≧74). The expression level of each hub gene between young and old

individuals was plotted as a violin graph. Total RNA from PBMCs was extracted by TRIzol (Invitrogen, United States). Synthesis of cDNA was performed by using 1μg of total RNA with PrimeScriptTM Reverse Transcriptase (Takara) according to the manufacturer's instructions. Specific primers used for qPCR were listed in the supplementary table 11. ACTB (NM_001101.5) was used as a reference gene for normalization. Quantitative real-time PCR was performed using the SYBR® Premix Ex Taq Kit (Takara) in a CFX96 Real Time PCR System (Bio-Rad Laboratories, Hercules, CA, United States) for at least three independent experiments. The relative gene expression levels was normalized to ACTB (NM_001101.5), and quantified using the $2^{-\Delta\Delta CT}$ method.

## List of abbreviations

WGCNA: weighted gene correlation network analyses; PBMC: peripheral blood mononuclear cells; 10KIP: the 10,000 immunomes project; PCA: principal component analyses; PC1: the first principal component; FPKM: fragments per kilobase of exon model per million mapped fragments; ME: module eigengene; GO: Gene Ontology; KEGG: kyoto encyclopedia of genes and genomes; FDR: false discovery rate; DEGs: differentially expressed genes; PPI: the protein protein interaction; MCC: the Maximal Clique Centrality.

## Ethics approval and consent to participate

The studies involving human participants were reviewed and approved by the local ethics committee of the First Affiliated Hospital of Jinan University (Approval#:KY-2020-027). All study participants provided their written informed consent to participate using institutional review board approved forms.

## Consent for publication

Not applicable.

## Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files. The raw sequencing data presented in the study has been deposited in the Sequence Read Archive (SRA) database, which was hosted by the NCBI, under accession number (SRA: PRJNA703752).

## Competing interests

Author Yutian Hu is employed by Meng Yi Center Limited, Macau, China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.


## Funding

This work was supported by grants from the National Key Research and Development Program of China (No.2018YFC2002003), the Natural Science Foundation of China (U1801285, 81971301), Guangzhou Planned Project of Science and Technology (201904010111, 202002020039), Milstein Medical Asian American Partnership (MMAAP) Foundation Irma and Paul Milstein Program for Senior Health Research Project Award (to G.C.), the China Post-doctoral Science Foundation (No.55350399), the Science and Technology Program of Guangdong Province (No.2020A1414040027), the Science and Technology Program of Guangzhou (201904010066), and the Postdoctoral Start-up Fund from the First Affiliated Hospital of Jinan University (No.809026).


## Authors' contributions

The concept of the study was planned by G.C., J.D., O.J.L. and Y.H.. Experiments were conducted, analyzed, and interpreted by Y.H., Y.X., L.M., W.L., G.Z., Y.H.(Yutian Hu), H.N., F.G., L.H., and L.G.. Sample preparation for mRNA, and sequencing were done by Y.X., J.X. and W.L.. Y.H. drafted the manuscript. O.J.L. and G.C. edited the manuscript and provided advice. All authors contributed to the article and approved the submitted version.


## Acknowledgements

We would like to thank all donors as well as the Guangzhou First People's Hospital for provision of the samples. We sincerely thank the contributors of 10KIP (the 10,000 immunomes project), as well as the Beijing Novogene Biotechnology Co., Ltd. for assistance with RNA-sequencing.

## References


1. Lopez-Otin C, Blasco M A, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. Cell.2013;153(6):1194-1217.
2. Nikolich-Zugich J. The twilight of immunity: emerging concepts in aging of the immune system. Nat Immunol.2018;19(1):10-19.
3. Noren H N, Longo D L, Evans M K. Age- and Race-Related Changes in Subpopulations of Peripheral Blood Lymphocytes in Humans. In: Fulop T, Franceschi C, Hirokawa K, Pawelec G, editors. Handbook of Immunosenescence. Springer: Cham; 2018. p. 1-30.
4. Tollerud D J, Clark J W, Brown L M, Neuland C Y, Pankiw-Trost L K, Blattner W A, et al. The influence of age, race, and gender on peripheral blood mononuclear-cell subsets in healthy nonsmokers. J Clin Immunol.1989;9(3):214-222.
5. Choong M L, Ton S H, Cheong S K. Influence of race, age and sex on the lymphocyte subsets in peripheral blood of healthy Malaysian adults. Ann Clin Biochem.1995;32:532-539.
6. Spielman R S, Bastone L A, Burdick J T, Morley M, Ewens W J, Cheung V G. Common genetic variants account for differences in gene expression among ethnic groups. Nat Genet.2007;39(2):226-231.
7. Kuchel G A. Inclusion of Older Adults in Research: Ensuring Relevance, Feasibility, and Rigor. J Am Geriatr Soc.2019;67(2):203-204.
8. Zalocusky K A, Kan M J, Hu Z, Dunn P, Thomson E, Wiser J, et al. The 10,000 Immunomes Project: Building a Resource for Human Immunology. Cell Rep.2018;25(2):513-522 e513.
9. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics.2008;9:559.
10. Hu Y, Pan J, Xin Y, Mi X, Wang J, Gao Q, et al. Gene Expression Analysis Reveals Novel Gene Signatures Between Young and Old Adults in Human Prefrontal Cortex. Front Aging Neurosci.2018;10:259.
11. Ritchie M E, Phipson B, Wu D, Hu Y, Law C W, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res.2015;43(7):e47.
12. Chen H, Boutros P C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics.2011;12:35.
13. Yu G, Wang L G, Han Y, He Q Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS.2012;16(5):284-287.
14. Szklarczyk D, Gable A L, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res.2019;47(D1):D607-D613.
15. Chin C H, Chen S H, Wu H H, Ho C W, Ko M T, Lin C Y. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol.2014;8 Suppl 4:S11.

16. Hamrick M W, Stranahan A M. Metabolic regulation of aging and age-related disease. Ageing Res Rev.2020;64:101175.
17. Maharzi N, Parietti V, Nelson E, Denti S, Robledo-Sarmiento M, Setterblad N, et al. Identification of TMEM131L as a novel regulator of thymocyte proliferation in humans. J Immunol.2013;190(12):6187-6197.
18. Fulop N, Feng W, Xing D, He K, Not L G, Brocks C A, et al. Aging leads to increased levels of protein O-linked N-acetylglucosamine in heart, aorta, brain and skeletal muscle in Brown-Norway rats. Biogerontology.2008;9(3):139.
19. Ve H, Cabana V C, Gouspillou G, Lussier M P. Quantitative Immunoblotting Analyses Reveal that the Abundance of Actin, Tubulin, Synaptophysin and EEA1 Proteins is Altered in the Brains of Aged Mice. Neuroscience.2020;442:100-113.
20. Ludwig M G, Seuwen K. Characterization of the human adenylyl cyclase gene family: cDNA, gene structure, and tissue distribution of the nine isoforms. J Recept Signal Transduct Res.2002;22(1-4):79-110.
21. Yamaguchi H, Yoshida S, Muroi E, Yoshida N, Kawamura M, Kouchi Z, et al. Phosphoinositide 3-kinase signaling pathway mediated by p110alpha regulates invadopodia formation. J Cell Biol.2011;193(7):1275-1288.
22. Gulati G S, Zukowska M, Noh J J, Zhang A, Wesche D J, Sinha R, et al. Neogenin-1 distinguishes between myeloid-biased and balanced Hoxb5 (+) mouse long-term hematopoietic stem cells. Proc Natl Acad Sci U S A.2019;116(50):25115-25125.
23. Menkin J A, Guan S A, Araiza D, Reyes C E, Trejo L, Choi S E, et al. Racial/Ethnic Differences in Expectations Regarding Aging Among Older Adults. Gerontologist.2017;57(suppl_2):S138-S148.
24. Peters M J, Joehanes R, Pilling L C, Schurmann C, Conneely K N, Powell J, et al. The transcriptional landscape of age in human peripheral blood. Nat Commun.2015;6:8570.
25. Gheorghe M, Snoeck M, Emmerich M, Back T, Goeman J J, Raz V. Major aging-associated RNA expressions change at two distinct age-positions. BMC Genomics.2014;15:132.
26. Song J, Baek I J, Chun C H, Jin E J. Dysregulation of the NUDT7-PGAM1 axis is responsible for chondrocyte death during osteoarthritis pathogenesis. Nat Commun.2018;9(1):3427.
27. Cho Y M, Bae S H, Choi B K, Cho S Y, Song C W, Yoo J K, et al. Differential expression of the liver proteome in senescence accelerated mice. Proteomics.2003;3(10):1883-1894.
28. Bauer M, Hamm A C, Bonaus M, Jacob A, Jaekel J, Schorle H, et al. Starvation response in mouse liver shows strong correlation with life-span-prolonging processes. Physiol Genomics.2004;17(2):230-244.
29. Yoshinaka T, Kosako H, Yoshizumi T, Furukawa R, Hirano Y, Kuge O, et al. Structural Basis of Mitochondrial Scaffolds by Prohibitin Complexes: Insight into a Role of the Coiled-Coil Region. iScience.2019;19:1065-1078.
30. Steger D J, Lefterova M I, Ying L, Stonestrom A J, Schupp M, Zhuo D, et al. DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled

with gene transcription in mammalian cells. Mol Cell Biol.2008;28(8):2825-2839.
31. Li Y, Wen H, Xi Y, Tanaka K, Wang H, Peng D, et al. AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. Cell.2014;159(3):558-571.
32. Calvanese V, Nguyen A T, Bolan T J, Vavilina A, Su T, Lee L K, et al. MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. Nature.2019;576(7786):281-286.

**Supplementary information:**

**Supplementary Figures**

Supplementary Figure S1. The age and race distribution of 153 individuals from 10 KIP dataset.Supplementary Tables.

Supplementary Figure S2. Power values and module clustering of WGCNA for the transcriptome data of 153 healthy human subjects in 10KIP.

Supplementary Figure S3. PCA analysis for age related gene modules in 153 individuals.

Supplementary Figure S4. The venn diagram of genes among DEG lists and co-expression module in 19 Asian (Chinese) RNA-seq data.

**Supplementary tables**

Supplementary Table 1: Basic information for 153 individuals from 10 KIP.

Supplementary Table 2: Trait-module relationships of the 153 individuals.

Supplementary Table 3: Module eigengene (ME) of the 153 individuals.

Supplementary Table 4: PCA analysis for the transcriptome of 153 individuals.

Supplementary Table 5: Basic information for 19 Chinese healthy individuals.

Supplementary Table 6: Trait-module relationships for the 19 Asian (Chinese).

Supplementary Table 7: Module eigengene (ME) of the 19 Asian (Chinese).

Supplementary Table 8: GO and KEGG enrichment for brown and Darkturquoise modules from 19 Chinese.

Supplementary Table 9: Trait-module relationships in the 113 Caucasian.

693 Supplementary Table 10: GO and KEGG enrichment for brown and turquoise
694 modules from 113 Caucasian.

695 Supplementary Table 11: The qPCR primers for homo sapiens.

696

697

## Figure legend

**Fig. 1** Age and race influenced the transcriptomic variations over human adult lifespan. WGCNA approach was applied for gene module consturction for the transcriptome data of 153 healthy human subjects in 10KIP. Principal component analyses (PCA) were calculated individually. **a** Cluster dendrogram. Each branch represented one gene and each color below denoted one co-expression gene module. The two colored rows below the dendrogram represented the original and merged modules, respectively. **b** Eigengene adjacency heatmap of different modules. Each module showed independent validation to each other, and higher correlation indicated higher co-expression interconnectedness. **c** Heatmap of the Pearson's correlation coefficient between trait (age, race and gender) and module eigengenes (ME, n=22). The column and row corresponded to ME and trait, respectively. Each cell contained the value of Pearson's correlation coefficient. The table was color-coded by correlation according to the color legend. The $p$-value < 0.05 represented statistical significance. **d-e** The characteristic gene expression changed during PBMC aging. The left and right-hand Y-axis represented the eigengene expression of each module, and trend line for each individuals, respectively. **f-g** Principal component1 scores (PC1) were calculated for each individual from principal component analyses (PCA). PC1 scores from transcriptomic data were differentially expressed among different races. Wilcoxon rank-sum test was used to compare data from Asian (n = 19) and Caucasian (n = 113) or African American subjects. Dot plot represented median and IQR values; ****$p$ < 0.0001, ***$p$ < 0.001, **$p$ < 0.01, *$p$ < 0.05, n.s.: non-significant.

**Fig. 2** The characteristic gene expression of PBMC aging in Asian (Chinese). Transcriptome data of 19 healthy human subjects in Guangdong China were analyzed, and gene modules were constructed by WGCNA. **a** Principal component analyses. Young and old individuals were largely separated according to the principal component1 scores (PC1). **b** Cluster dendrogram. Each branch represents one gene and each color below denotes one co-expression gene module. The two colored rows below the dendrogram, represented the original and merged modules, respectively. **c** Heatmap of the Pearson’s correlation between trait (age and gender) and module eigengenes (ME, n=40). The column and row corresponded to trait and ME, respectively. The color in each cell represented corresponding correlation, and scaled in the color legend. **d** Hierarchical cluster analyses of four interested modules. Based

on the module-trait's Pearson's r and $p$ value (absolute r > 0.5, $p$< 0.05), three modules (cyan, darkturquoise, and orange) showed relatively lower expression in young PBMC groups and high expression in the aged, while the brown modules showed the opposite result. Each dot represented an individual. **e** The histograms of the eigengene expression in the four age-related modules from young to old.

**Fig. 3** Novel and known age-associated genes and pathways associated with PBMC aging in Asian (Chinese). Gene Ontology (GO) and KEGG pathway enrichment analyses were conducted to analyze the biological functions of modules. Module eigengene (ME) was defined as the first principal component of the expression matrix of the corresponding module and was considered as a representation of the gene expression profiles in a module. **a** The transcriptomic expression of age related modules changed significantly among young and old individuals. Based on ME expression profile of the four interesting modules, the expression of cyan, darkturquoise and orange modules were downregulated, while brown module showed the opposite results. Box plots represent mean ± standard deviation (SD). The $p$-value was calculated by the student's t-test, n=19, *$p$ < 0.05, **$p$ < 0.01, ***$p$ < 0.001, ****$p$ < 0.0001. **b-c** Top 10 GO biological process functional annotation (B) and KEGG pathway enrichment (C) analyses for the brown and darkturquoise modules. The Benjamini-Hochberg FDR multiple test correction was also applied to assess significance of hypergeometric p-values at a false discovery rate (FDR) of 5%. The color represented the adjusted $p$-values. **d-e** Hub gene detection for the darkturquoise (D) and brown (E) modules. PPI network of the brown and darkturquoise modules were based on the STRING database. And each node represented a protein-coding gene and the size of each node was mapped to its connectivity (also known as degree). **f-g** The verification of the hub genes. The top of three genes in brown and darkturquoise modules were selected and its mRNA abundance of these hub genes were detected in young and old individuals. Box plots represent mea ± standard deviation (SD). The $p$-value was calculated by the student's t-test, young=9, old=10, *$p$ < 0.05, **$p$ < 0.01, ***$p$ < 0.001, ****$p$ < 0.0001.

**Fig. 4** Aging-related changes in PBMC transcriptomes in Caucasian and its hub gene detection. Transcriptome data of 113 healthy Caucasian subjects in 10KIP were analyzed and gene modules were constructed by WGCNA. **a** Module-trait relationships. Row and column corresponded to module eigengenes and clinical trait (age and gender), respectively. Each cell contained the corresponding Pearson's

correlation coefficient and *p*-value. **b** The histograms presentation of the eigengene expression in the brown and turquoise modules from young to old. **c** Comparison of eigengene expression of the brown and turquoise modules between young and older. The transcriptomic expression of age related modules changed significantly among young and old individuals. Box plots represent mean ± standard deviation (SD). The *p*-value was calculated by the student's t-test, n=113, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. **d-e** Gene Ontology (GO) and KEGG enrichment analyses for the genes in the brown and turquoise modules. The top 10 of the GO enriched biological process and enriched KEGG pathway were shown. The Benjamini-Hochberg FDR multiple test correction was also applied to assess significance of hypergeometric p-values at a false discovery rate (FDR) of 5% in GO and KEGG enrichment analyses. The color represented the adjusted *p*-values, and the size of the bars represented the gene number. **f** The Venn diagram showed the overlapping genes among differential expression genes (DEG) and co-expression modules. In total, 50 and 177 overlapping genes were listed in the intersection of DEG lists and two co-expression modules, respectively. **g** Hub genes detection for the brown and turquoise modules by using the STRING database. **h** The verification of the hub genes. The top of genes in brown and turquoise modules were selected and mRNA abundance of these hub genes were detected in young and old individuals. Box plots represent mean ± standard deviation (SD). The *p*-value was calculated by the student's t-test, young=33, old=41, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

**Fig. 5** Shared transcriptomic signatures of PBMC aging between Caucasian and Asian (Chinese). The brown module from Asian (Chinese), the turquoise module from Caucasian and its differentially expressed genes (DEGs) were compared and listed. **a** The Venn diagram of genes among turquoise module from Caucasian and brown module from Asian (Chinese). Despite thousands of race-specific gene associated with aging corresponding to 2623 and 1688 genes in Asian (Chinese) and Caucasian, 95 genes in Asian (Chinese) and Caucasian significantly overlapped. **b-c** GO and KEGG enrichment analyses for the 95 common shared genes. **d** Hub genes detection for 95 genes by using the STRING database, and visualized by the cytoscape. **e** The Venn diagram of differentially expressed genes (DEGs) with the age-related modules in White and Caucasian revealed two aging-specific gene markers. **f** The two overlapping genes (OXNAD1 and MLLT3) were both downregulated, as shown in the volcano diagram for the DEG genes in Asian and Caucasian dataset.

**Fig. 6** Validation of expression levels of the common hub genes involved in PBMC aging. The validation in Asian (Chinese) and Caucasian were performed using additional samples and 10KIP data, respectively. **a-b** Gene expression value of the hub genes among young and old samples in 19 Asian (Chinese) (A) and 153 Caucasian (B). The data were expressed as the mean ± standard deviation (SD). Student's t-test was used for statistical analyses. In Asian (Chinese), young n=9, old n=10; in Caucasian, young n=33, old n=40, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. **c** Gene expression of hub genes among samples of man and woman during their lifespan, n=124, including 19 Asian and 105 caucasian. **d** Quantification of the four hub genes was confirmed and presented by the qPCR assay. The data were expressed as the mean ± standard deviation (SD). The *p*-value was calculated by the student's t-test. Young n=7; old n=5, *$p < 0.05$, **$p < 0.01$.

# Figures

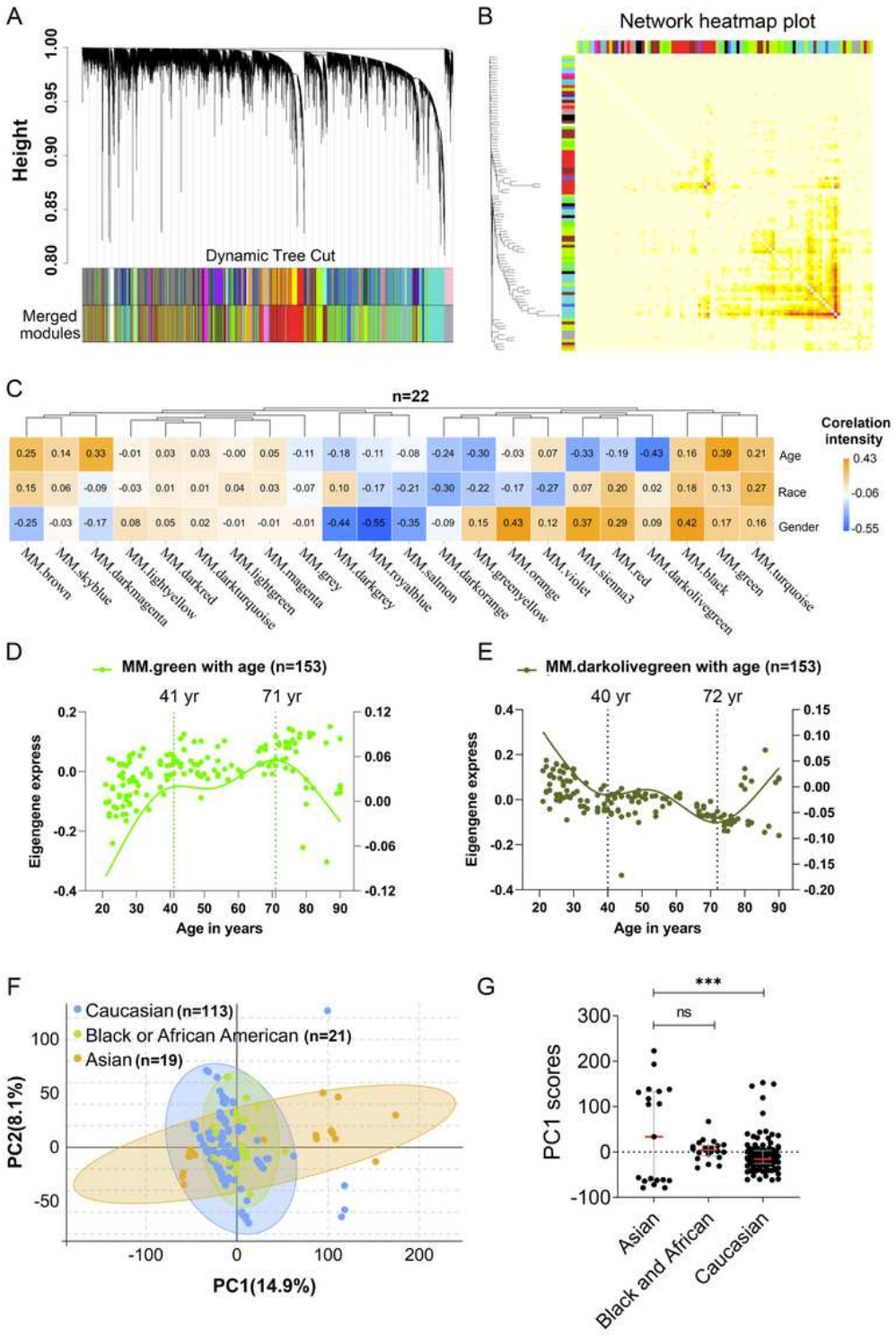
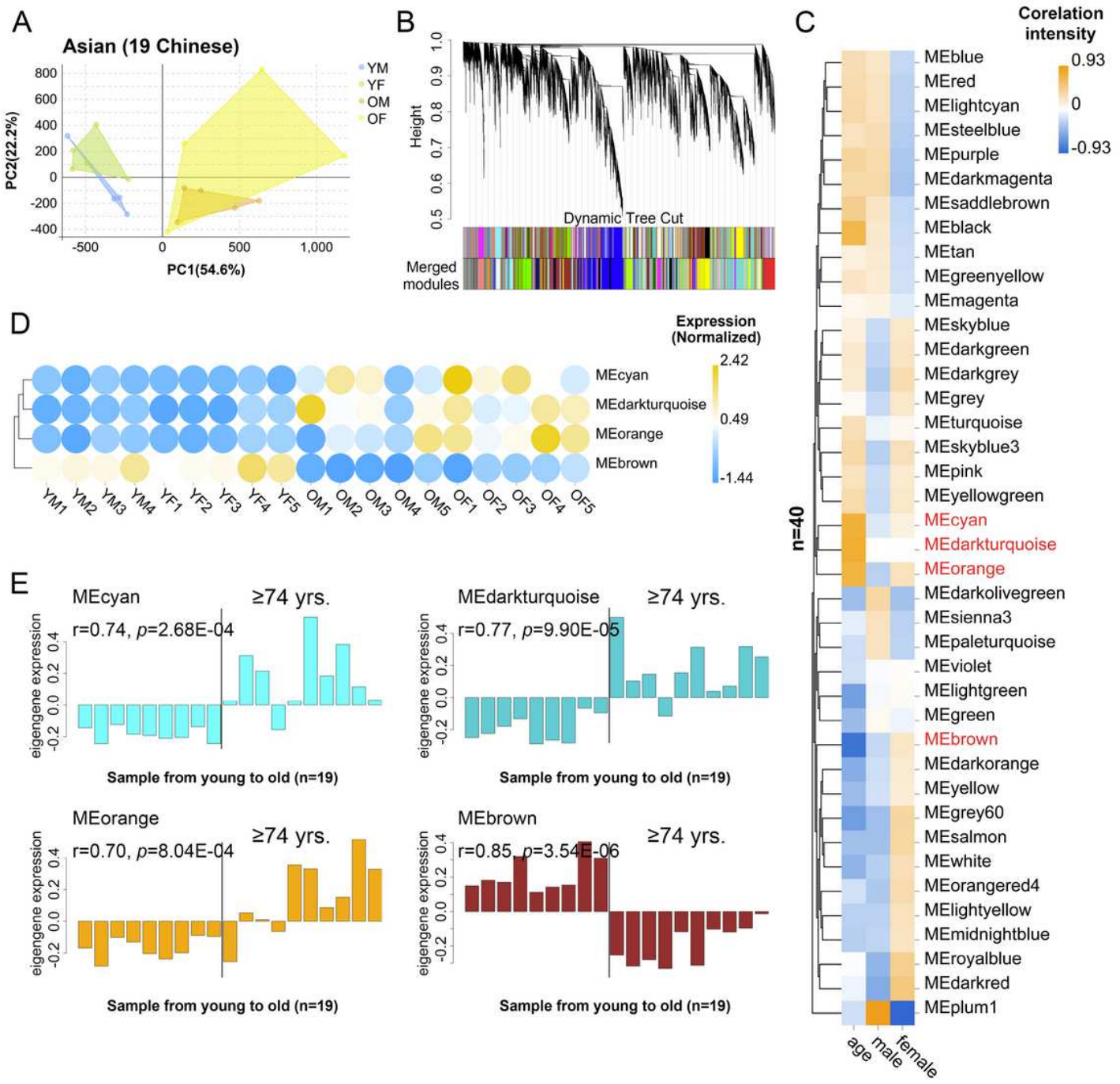



**Figure 1**

Age and race influenced the transcriptomic variations over human adult lifespan. WGCNA approach was applied for gene module consturction for the transcriptome data of 153 healthy human subjects in 10KIP. Principal component analyses (PCA) were calculated individually. a Cluster dendrogram. Each branch

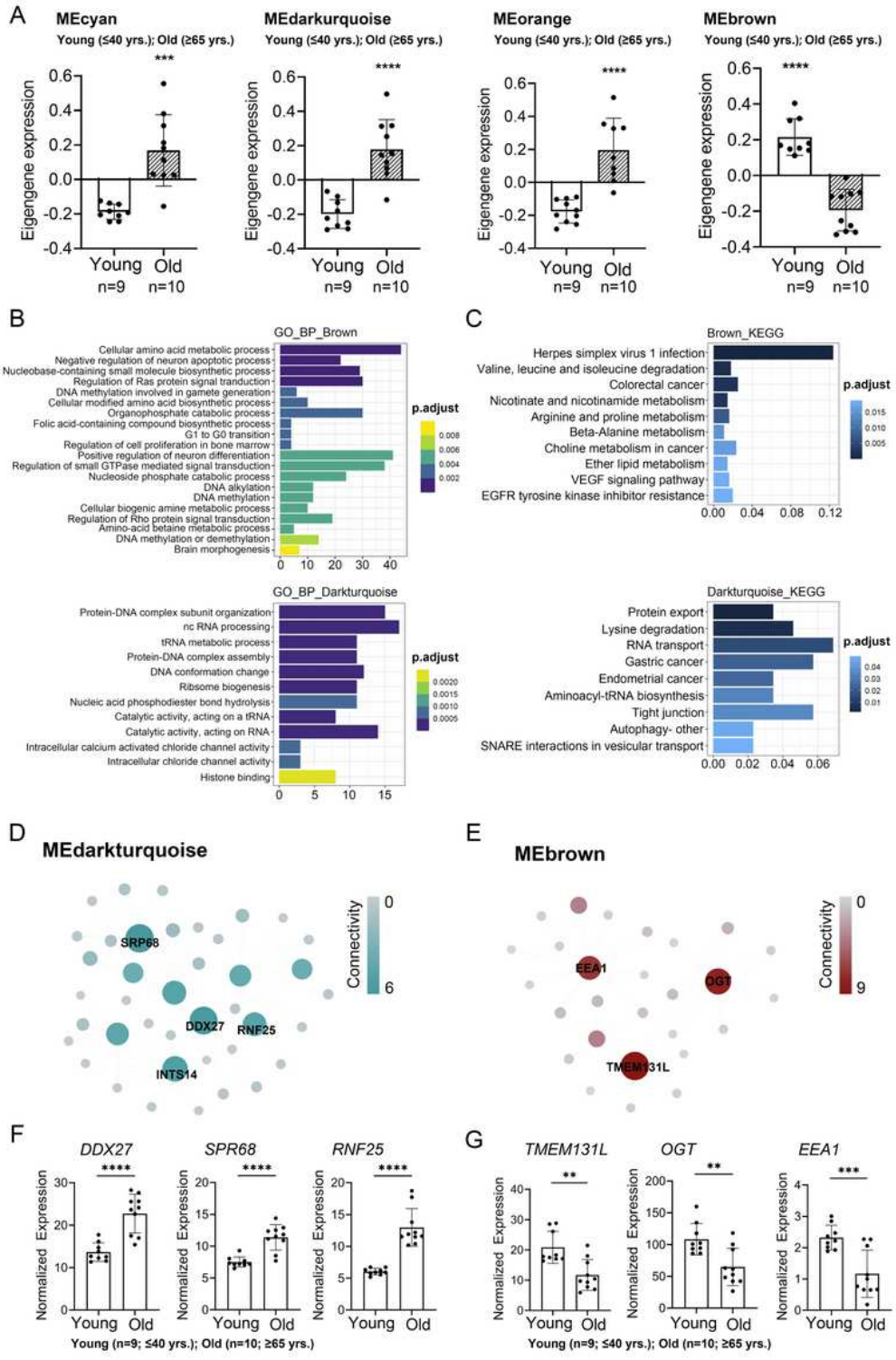
represented one gene and each color below denoted one co-expression gene module. The two colored rows below the dendrogram represented the original and merged modules, respectively. b Eigengene adjacency heatmap of different modules. Each module showed independent validation to each other, and higher correlation indicated higher co-expression interconnectedness. c Heatmap of the Pearson's correlation coefficient between trait (age, race and gender) and module eigengenes (ME, n=22). The column and row corresponded to ME and trait, respectively. Each cell contained the value of Pearson's correlation coefficient. The table was color-coded by correlation according to the color legend. The p-value < 0.05 represented statistical significance. d-e The characteristic gene expression changed during PBMC aging. The left and right-hand Y-axis represented the eigengene expression of each module, and trend line for each individuals, respectively. f-g Principal component1 scores (PC1) were calculated for each individual from principal component analyses (PCA). PC1 scores from transcriptomic data were differentially expressed among different races. Wilcoxon rank-sum test was used to compare data from Asian (n = 19) and Caucasian (n = 113) or African American subjects. Dot plot represented median and IQR values; ****p < 0.0001, ***p < 0.001, **p < 0.01, *p < 0.05, n.s.: non-significant.




## Figure 2

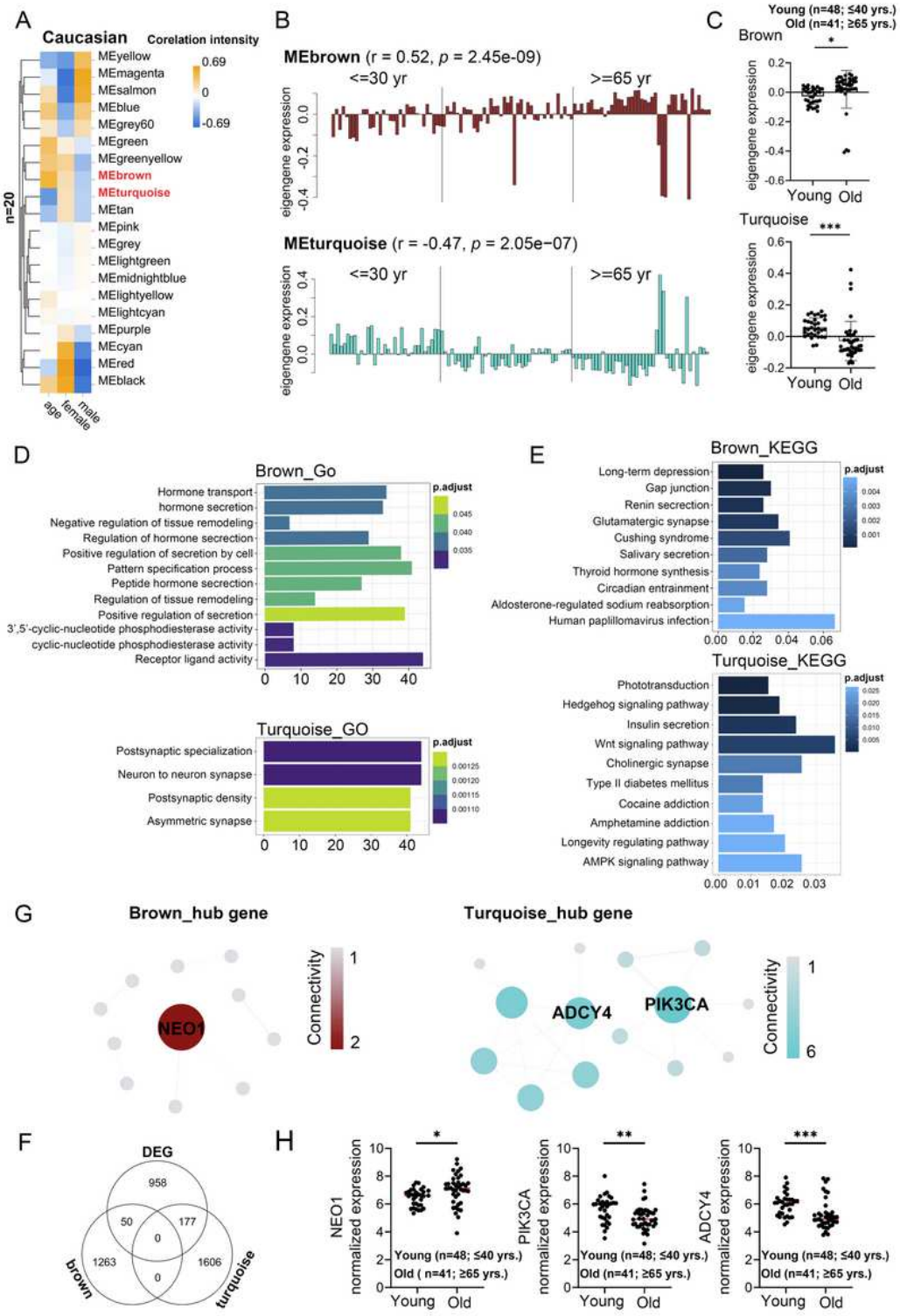
The characteristic gene expression of PBMC aging in Asian (Chinese). Transcriptome data of 19 healthy human subjects in Guangdong China were analyzed, and gene modules were constructed by WGCNA. a Principal component analyses. Young and old individuals were largely separated according to the principal component1 scores (PC1). b Cluster dendrogram. Each branch represents one gene and each color below denotes one co-expression gene module. The two colored rows below the dendrogram, represented the original and merged modules, respectively. c Heatmap of the Pearson's correlation

between trait (age and gender) and module eigengenes (ME, n=40). The column and row corresponded to trait and ME, respectively. The color in each cell represented corresponding correlation, and scaled in the color legend. d Hierarchical cluster analyses of four interested modules. Based on the module-trait's Pearson's r and p value (absolute 732 r > 0.5, p< 0.05), three modules (cyan, darkturquoise, and orange) showed relatively lower expression in young PBMC groups and high expression in the aged, while the brown modules showed the opposite result. Each dot represented an individual. e The histograms of the eigengene expression in the four age-related modules from young to old.



## Figure 3

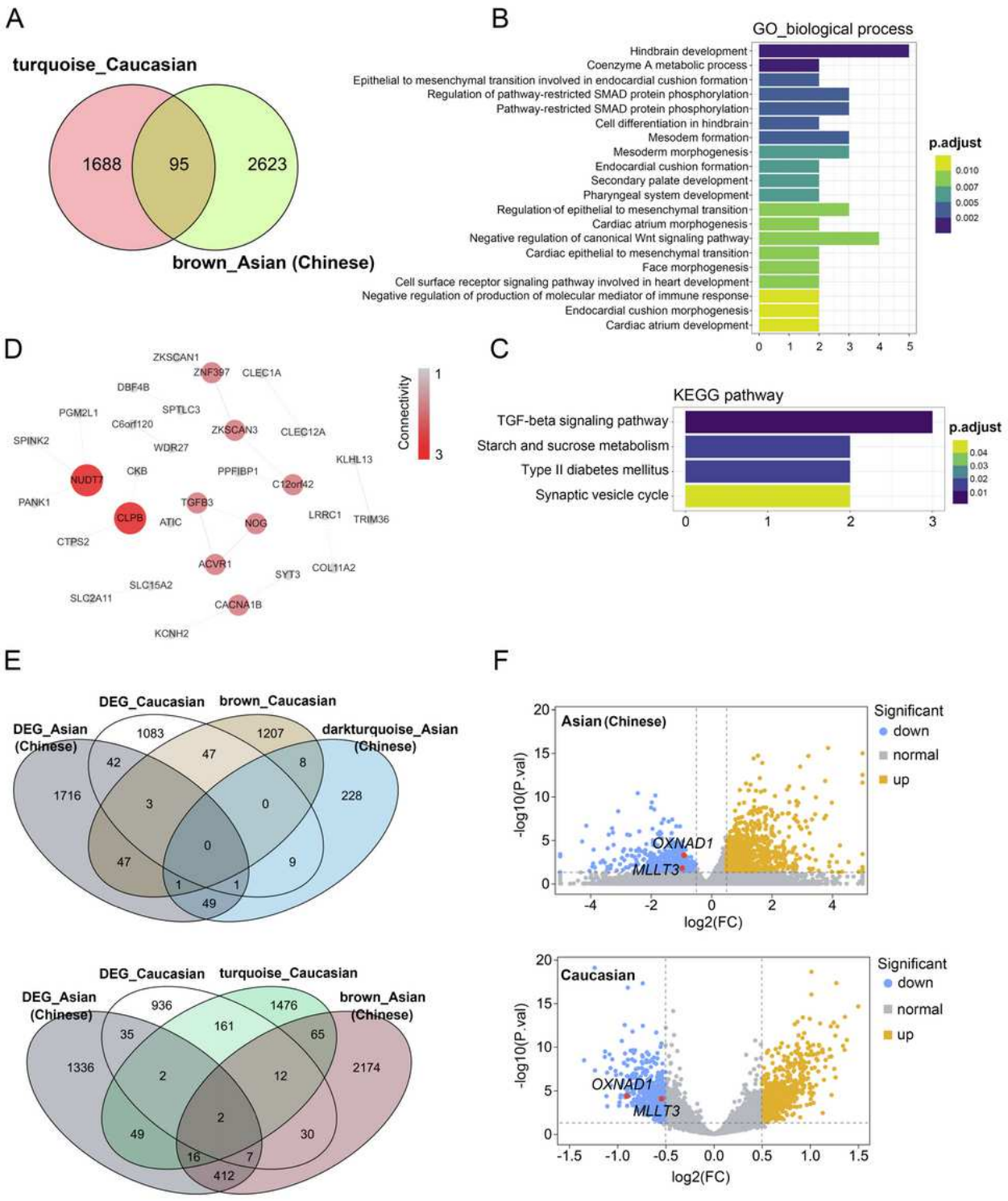
Novel and known age-associated genes and pathways associated with PBMC aging in Asian (Chinese). Gene Ontology (GO) and KEGG pathway enrichment analyses were conducted to analyze the biological functions of modules. Module eigengene (ME) was defined as the first principal component of the expression matrix of the corresponding module and was considered as a representation of the gene expression profiles in a module. a The transcriptomic expression of age related modules changed significantly among young and old individuals. Based on ME expression profile of the four interesting modules, the expression of cyan, darkturquoise and orange modules were downregulated, while brown module showed the opposite results. Box plots represent mean ± standard deviation (SD). The p-value was calculated by the student's t-test, n=19, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. b-c Top 10 GO biological process functional annotation (B) and KEGG pathway enrichment (C) analyses for the brown and darkturquoise modules. The Benjamini-Hochberg FDR multiple test correction was also applied to assess significance of hypergeometric p-values at a false discovery rate (FDR) of 5%. The color represented the adjusted p-values. d-e Hub gene detection for the darkturquoise (D) and brown (E) modules. PPI network of the brown and darkturquoise modules were based on the STRING database. And each node represented a protein-coding gene and the size of each node was mapped to its connectivity (also known as degree). f-g The verification of the hub genes. The top of three genes in brown and darkturquoise modules were selected and its mRNA abundance of these hub genes were detected in young and old individuals. Box plots represent mea±standard deviation (SD). The p-value was calculated by the student's t-test, young=9, old=10, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.




## Figure 4

Aging-related changes in PBMC transcriptomes in Caucasian and its hub gene detection. Transcriptome data of 113 healthy Caucasian subjects in 10KIP were analyzed and gene modules were constructed by WGCNA. a Module-trait relationships. Row and column corresponded to module eigengenes and clinical trait (age and gender), respectively. Each cell contained the corresponding Pearson's correlation coefficient and p-value. b The histograms presentation 766 of the eigengene expression in the brown and

turquoise modules from young to old. c Comparison of eigengene expression of the brown and turquoise modules between young and older. The transcriptomic expression of age related modules changed significantly among young and old individuals. Box plots represent mean±standard deviation (SD). The p-value was calculated by the student's t-test, n=113, *p < 0.05, **p < 0.01, ***p < 0.001. d-e Gene Ontology (GO) and KEGG enrichment analyses for the genes in the brown and turquoise modules. The top 10 of the GO enriched biological process and enriched KEGG pathway were shown. The Benjamini-Hochberg FDR multiple test correction was also applied to assess significance of hypergeometric p-values at a false discovery rate (FDR) of 5% in GO and KEGG enrichment analyses. The color represented the adjusted p-values, and the size of the bars represented the gene number. f The Venn diagram showed the overlapping genes among differential expression genes (DEG) and co-expression modules. In total, 50 and 177 overlapping genes were listed in the intersection of DEG lists and two co-expression modules, respectively. g Hub genes detection for the brown and turquoise modules by using the STRING database. h The verification of the hub genes. The top of genes in brown and turquoise modules were selected and mRNA abundance of these hub genes were detected in young and old individuals. Box plots represent mean±standard deviation (SD). The p-value was calculated by the student's t-test, young=33, old=41, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.




## Figure 5

Shared transcriptomic signatures of PBMC aging between Caucasian and Asian (Chinese). The brown module from Asian (Chinese), the turquoise module from Caucasian and its differentially expressed genes (DEGs) were compared and listed. a The Venn diagram of genes among turquoise module from Caucasian and brown module from Asian (Chinese). Despite thousands of race-specific gene associated with aging corresponding to 2623 and 1688 genes in Asian (Chinese) and Caucasian, 95 genes in Asian

(Chinese) and Caucasian significantly overlapped. b-c GO and KEGG enrichment analyses for the 95 common shared genes. d Hub genes detection for 95 genes by using the STRING database, and visualized by the cytoscape. e The Venn diagram of differentially expressed genes (DEGs) with the age-related modules in White and Caucasian revealed two aging-specific gene markers. f The two overlapping genes (OXNAD1 and MLLT3) were both downregulated, as shown in the volcano diagram for the DEG genes in Asian and Caucasian dataset.

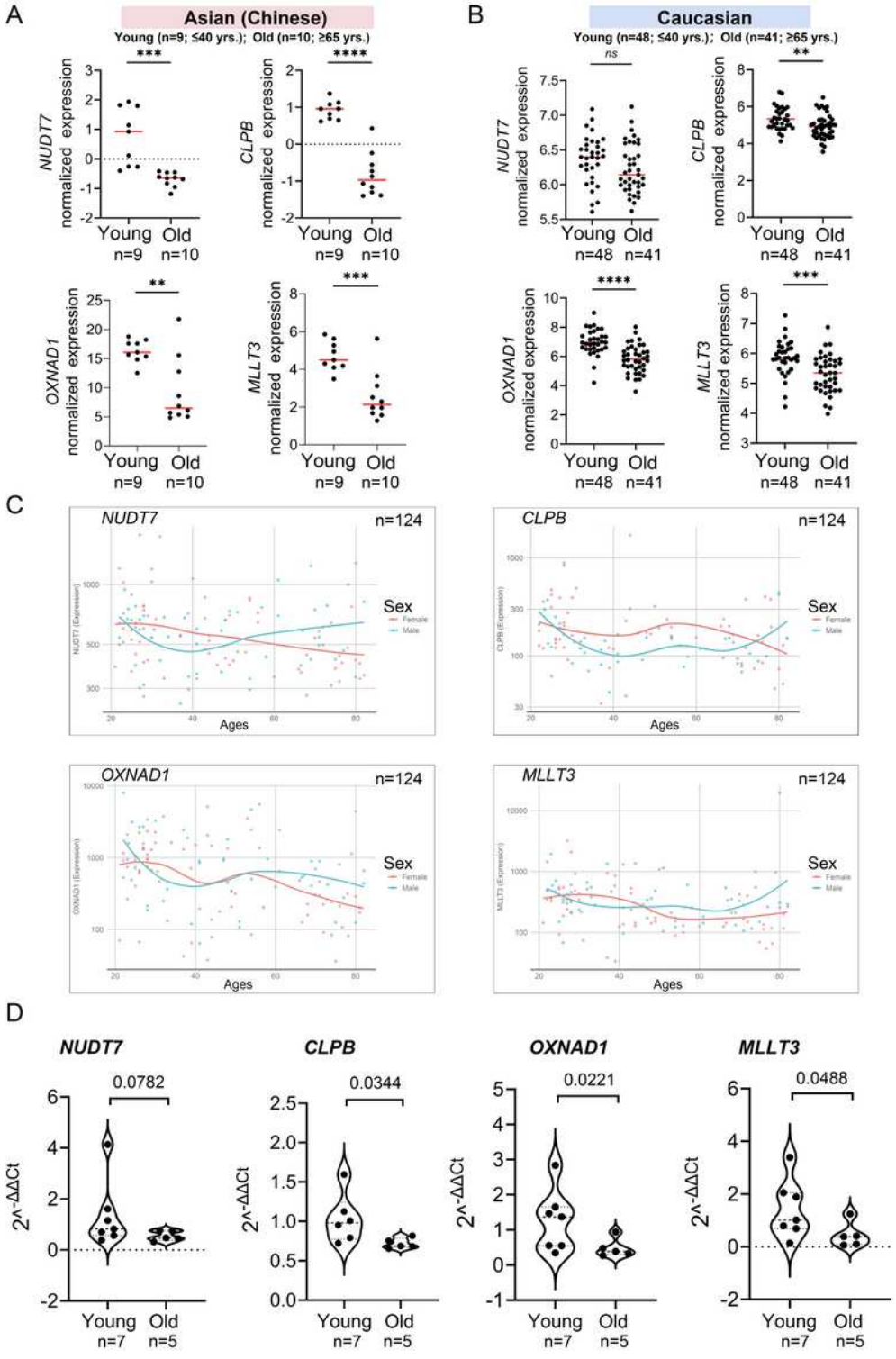



Figure 6

Validation of expression levels of the common hub 800 genes involved in PBMC aging. The validation in Asian (Chinese) and Caucasian were performed using additional samples and 10KIP data, respectively. a-b Gene expression value of the hub genes among young and old samples in 19 Asian (Chinese) (A) and 153 Caucasian (B). The data were expressed as the mean ±standard deviation (SD). Student's t-test was used for statistical analyses. In Asian (Chinese), young n=9, old n=10; in Caucasian, young n=33, old n=40, *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001. c Gene expression of hub genes among samples of man and woman during their lifespan, n=124, including 19 Asian and 105 caucasian. d Quantification of the four hub genes was confirmed and presented by the qPCR assay. The data were expressed as the mean±standard deviation (SD). The p-value was calculated by the student's t-test. Young n=7; old n=5, *p < 0.05, **p < 0.01.

## Supplementary Files

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

- SupplementaryFile.docx
- SupplementaryTables.xlsx