ORIGINAL PAPER

Genetic associations for two biological age measures point to distinct aging phenotypes

Chia-Ling Kuo[1,2] | Luke C. Pilling[2,3] | Zuyun Liu[4] | Janice L. Atkins[3] |
Morgan E. Levine[5]

[1]Connecticut Convergence Institute for Translation in Regenerative Engineering, University of Connecticut Health, Farmington, Connecticut, USA

[2]Center on Aging, School of Medicine, University of Connecticut, Farmington, Connecticut, USA

[3]College of Medicine and Health, University of Exeter, Exeter, UK

[4]Department of Big Data in Health Science, School of Public Health and the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

[5]Department of Pathology, Yale School of Medicine, New Haven, Connecticut, USA

Correspondence
Morgan E. Levine, 310 Cedar St. Suite 315A, New Haven, CT 06519, USA.
Email: morgan.levine@yale.edu

Chia-Ling Kuo, 195 Farmington Ave, Suite 2080, Farmington, CT 06032, USA.
Email: kuo@uchc.edu

Abstract

Biological age measures outperform chronological age in predicting various aging outcomes, yet little is known regarding genetic predisposition. We performed genome-wide association scans of two age-adjusted biological age measures (PhenoAgeAcceleration and BioAgeAcceleration), estimated from clinical biochemistry markers (Levine et al., 2018; Levine, 2013) in European-descent participants from UK Biobank. The strongest signals were found in the *APOE* gene, tagged by the two major protein-coding SNPs, PhenoAgeAccel—rs429358 (*APOE* e4 determinant) ($p = 1.50 \times 10^{-72}$); BioAgeAccel—rs7412 (*APOE* e2 determinant) ($p = 3.16 \times 10^{-60}$). Interestingly, we observed inverse *APOE* e2 and e4 associations and unique pathway enrichments when comparing the two biological age measures. Genes associated with BioAgeAccel were enriched in lipid related pathways, while genes associated with PhenoAgeAccel showed enrichment for immune system, cell function, and carbohydrate homeostasis pathways, suggesting the two measures capture different aging domains. Our study reaffirms that aging patterns are heterogeneous across individuals, and the manner in which a person ages may be partly attributed to genetic predisposition.

KEYWORDS
APOE, biomarkers, cardiac aging, inflammaging, polygenic risk score

1 | INTRODUCTION

While chronological age is arguably the strongest determinant of risk for major chronic diseases, it alone is not sufficient to reflect the state of biological aging. Individuals with similar chronological ages are heterogeneous in their physiological states, and subsequent health risks, due to differences in both the rate and manner of biological aging. As a result, efforts have been launched to develop measures that can capture the concept of biological age (BA) (Ferrucci et al., 2020). Typically, these measures encompass single or composite biomarkers found to be associated with a surrogate of biological age, usually chronological age or mortality. In principle, a valid BA measure needs to outperform chronological age in predicting lifespan and a wide range of age-sensitive tests in multiple

physiological and behavioral domains (Butler et al., 2004). BA predictors may help discover drivers of the aging process and can be used for secondary prevention by identifying at-risk individuals prior to disease onset. Additionally, they have been proposed as tools to monitor intervention or treatment effects aimed at targeting the aging process.

A variety of data modalities, covering different biological levels of organization, have been used to develop BA predictors, for example, DNA methylation data, gene expression data, proteomic data, metabolomic data, and clinical chemistry measures (Jylhävä et al., 2017). A study comparing 11 BA predictors found low agreement between those based on DNA methylation, clinical biomarkers, and telomere length, in terms of both their correlations with each other and their relative associations with healthspan-related characteristics: balance, grip strength, motor coordination, physical limitations, cognitive decline, self-rated health, and facial aging (Belsky et al., 2018). These findings suggest that different BA predictors may be capturing distinct aspects of the aging process. Furthermore, a recent paper (Ahadi et al., 2020) combining various omics data identified distinct "ageotypes", which represent diverse aging patterns across individuals.

We hypothesize that individual susceptibility to one biological aging domain versus another may be due in part to underlying genetic mechanisms. To test the hypothesis, we used data from the UK Biobank study to understand genetic predisposition to accelerated aging, measured by two validated biological age predictors, PhenoAge (Levine et al., 2018) and BioAge (Levine, 2013). PhenoAge and BioAge were previously trained for mortality and chronological age, respectively, as surrogates of biological age, using the biomarker data from the National Health and Nutrition Examination Survey (NHANES) III in the United States (US). Biomarkers were selected based on associations with the biological age surrogates for both measures and knowledge regarding their role or dependency in the aging process also for BioAge. Details on variable selection and methods to construct PhenoAge and BioAge are provided in the Methods. PhenoAge is a function of chronological age, albumin, creatinine, C-reactive protein (CRP), alkaline phosphatase, glucose, lymphocyte percentage, mean corpuscular volume, red blood cell distribution width (RDW), and white blood cell count. BioAge is a function of chronological age, albumin, creatinine, CRP, and alkaline phosphatase (also in PhenoAge), plus glycated hemoglobin (HbA1c), systolic blood pressure, and total cholesterol. Both aging measures have been shown to be robust predictors of aging outcomes (Levine, 2013; Levine et al., 2018; Liu et al., 2018) yet are clearly distinct. Biological age acceleration measured by either biological age measure adjusted for chronological age (PhenoAgeAccel or BioAgeAccel) was similarly associated with morbidity and mortality in the full sample and subgroups of National Health and Nutrition Examination Survey (NHANES) IV, but PhenoAgeAccel outperformed BioAgeAccel in those disease-free and with normal body mass index (Liu et al., 2018). We applied PhenoAge and BioAge that were previously trained using the US NHANES III data to the UK Biobank for genome-wide association studies (GWASs). Overall, this study is fully supported by UK Biobank, featured by a large sample and extensive genetic and phenotypic data.

2 | RESULTS

451,367 genetically-determined Europeans were identified in UK Biobank. Of whom, 379,703 unrelated participants were included in analyses. Among the included samples (Table S1), 204,736 (54%) participants were female. After a mean follow-up time of 11.49 years (standard deviation (SD) = 1.55) to April 26, 2020 (last death in the data), 23,060 participants died with the mean age at death 69.06 years (SD = 7.21, range: 40.84 to 82.50). A summary of PhenoAge or BioAge biomarkers is provided in Table S1.

Participants were biologically younger than their chronological ages, with the mean PhenoAge and BioAge, 54.43 years (SD = 9.56) and 56.16 years (SD = 8.17) versus the mean chronological age 56.74 years (SD = 8.02). PhenoAge acceleration (PhenoAgeAccel) estimated by residualizing PhenoAge based on chronological age via a linear regression model was weakly correlated ($r = 0.23$) with BioAge acceleration (BioAgeAccel) that was similarly defined. Both BioAgeAccel and PhenoAgeAccel were significantly associated with all-cause mortality in this young cohort ($p < 2 \times 10^{-6}$), with the hazard ratio (HR) 1.100 (95% CI: 1.097 to 1.102) per year increase in PhenoAgeAccel and 1.054 (95% CI: 1.046 to 1.062) per year increase in BioAgeAccel. In the above models, sex was included, additional to chronological age and PhenoAge or BioAge. When both PhenoAge and BioAge were included, the hazard ratio with PhenoAgeAccel was little changed (HR = 1.099, 95% CI: 1.097 to 1.102) but that with BioAgeAccel (HR = 1.000, 95% CI: 0.992 to 1.008) was reduced toward the null.

The data was split with a 1 to 2 ratio, where PhenoAge was available for 107,460 participants in the training set and for 214,192 participants in the testing set. Similarly, BioAge was available for 98,446 participants in the training set and for 195,847 participants in the testing set. The training set was used to perform genome-wide association analysis and the GWAS summary statistics were used to construct polygenic risk scores (PRSs) in the testing set, evaluated for the use of risk stratification for a variety of age-related outcomes. Demographics and PhenoAge or BioAge biomarker levels were quite balanced between the training and testing sets (Table S2).

2.1 | PhenoAgeAccel GWAS

In the Genome-Wide Association Study (GWAS) of PhenoAgeAccel, 7,561 genetic variants were identified ($p < 5 \times 10^{-8}$) (Manhattan plot in Figure 1, including the top 10 mapped genes of lead SNPs detailed in Table 1). The SNP p-value distribution showed sizable deviation from the null distribution of no association. However, there was a lack of evidence of population stratification or cryptic relatedness (LD score regression intercept 1.02 with the standard error (SE) 0.01, compared to the null value 1; genomic inflation factor 1.12),
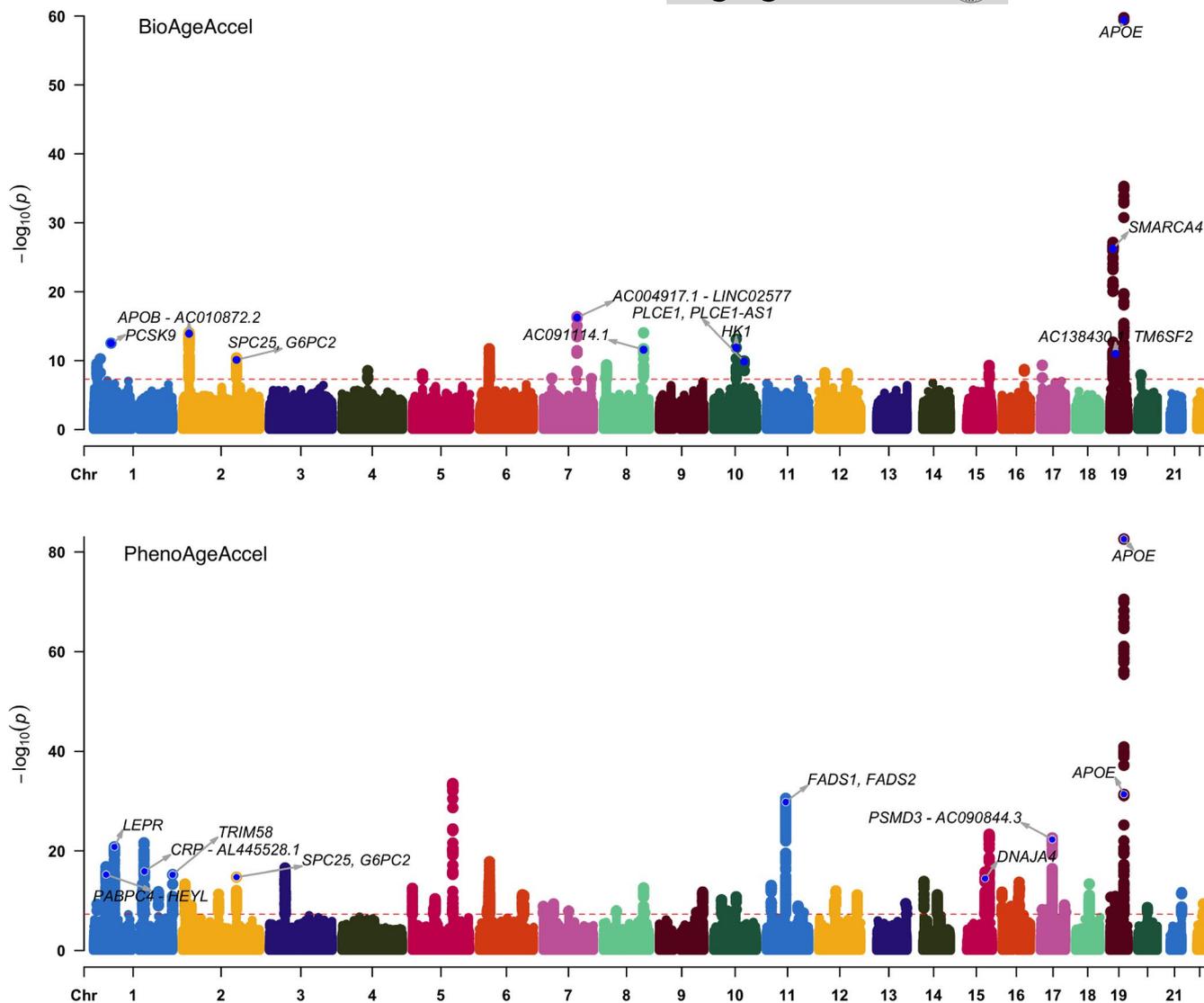
FIGURE 1 PhenoAgeAccel (bottom) and BioAgeAccel (top) Manhattan plots (colors to separate adjacent chromosomes without other indications), including the top 10 mapped genes of lead SNPs (see Tables 2 and 3)

and the proportion of inflation not explained by polygenic heritability was only 6.33% (SE = 3.06%).

The SNP-heritability for PhenoAgeAccel was estimated to be 14.45% (SE = 0.95%). We identified 55 independent signals ($p < 5 \times 10^{-8}$) tagged by 55 lead SNPs. Of which, 29 were near genes (Table 1). Both *APOE* isoform coding SNPs (rs429358 and rs7412 on chromosome 19) were identified. Multiple lead SNPs were associated with CRP, glucose or HbA1c, and hematology traits, based on previous GWAS catalog (Buniello et al., 2019) results (Table S3). Lead SNPs nearby *GCKR*, *FTO*, *ZPR1*, and *APOE* were associated with various traits including cardiovascular diseases and/or PhenoAge biomarkers (Table S3).

The Multi-marker Analysis of GenoMic Annotation (MAGMA) gene set analysis identified 11 gene sets at the Bonferroni-corrected level of 5%, including regulation of signaling and transcription, homeostasis (carbohydrate homeostasis, homeostasis of number of

cells, and myeloid cell homeostasis), and immune system process (Figure 2). In the MAGMA tissue expression analysis, we found that genes expressed in whole blood and liver were more likely to be associated with PhenoAgeAccel than genes expressed in other tissues (Figure 3).

2.2 | BioAgeAccel GWAS

In the GWAS for BioAgeAccel, 996 genetic variants were identified ($p < 5 \times 10^{-8}$) (Manhattan plot in Figure 1 including the top 10 mapped genes of lead SNPs detailed in Table 2). The observed *p*-value distribution was significantly deviant from the expected under the null (Figure 2). However, there was no evidence to suggest population stratification or cryptic relatedness (LD score regression intercept = 1.02, SE = 0.01; genomic control factor 1.11), and the

TABLE 1 Genetic loci associated with PhenoAgeAccel ($p < 5 \times 10^{-8}$) that can be mapped to genes

SNP	Chr	bp	refA	freq	bJ	bJ_se	pJ	Genes
rs1801133	1	11856378	G	0.66	−0.13	0.022	1.28E−09	MTHFR
rs12037222	1	40064961	G	0.77	−0.2	0.025	1.98E−15	PABPC4 - HEYL
rs1805096	1	66102257	G	0.63	0.2	0.021	1.01E−20	LEPR
rs4129267	1	154426264	C	0.59	0.16	0.021	6.73E−15	IL6R
rs7553007	1	159698549	G	0.67	0.19	0.022	2.44E−17	CRP - AL445528.1
rs12239046	1	247601595	T	0.37	−0.12	0.021	1.40E−08	NLRP3
rs3811444	1	248039451	C	0.66	0.18	0.022	2.61E−16	TRIM58
rs1260326	2	27730940	T	0.39	−0.13	0.021	2.29E−09	GCKR
rs6734238	2	113841030	A	0.6	−0.13	0.021	1.51E−10	IL1F10 - RNU6-1180P
rs560887	2	169763148	T	0.3	−0.18	0.023	1.72E−15	SPC25, G6PC2
rs35188965	5	1104938	C	0.42	0.15	0.021	5.45E−13	SLC12A7
rs7775698	6	135418635	C	0.74	0.16	0.024	9.60E−12	HBS1L
rs592423	6	139840693	A	0.45	0.14	0.021	3.38E−11	AL592429.2
rs17321515	8	126486409	A	0.53	−0.14	0.021	4.08E−12	AC091114.1
rs8176746	9	136131322	G	0.94	0.28	0.043	4.54E−11	ABO
rs7908745	10	45953767	A	0.68	−0.14	0.022	1.31E−09	MARCH8
rs16926246	10	71093392	C	0.87	−0.2	0.031	1.46E−10	HK1
rs174548	11	61571348	C	0.69	0.26	0.022	8.18E−31	FADS1, FADS2
rs964184	11	116648917	G	0.13	−0.17	0.03	3.28E−08	ZPR1
rs2393791	12	121423956	C	0.38	−0.15	0.021	7.45E−12	HNF1A
rs8013143	14	23494277	A	0.72	−0.17	0.023	4.66E−14	PSMB5
rs3169166	15	78563103	A	0.58	0.16	0.021	1.34E−14	DNAJA4
rs9939609	16	53820527	T	0.61	−0.16	0.021	1.62E−13	FTO
rs9914988	17	27183104	G	0.2	−0.14	0.026	3.20E−08	ERAL1
rs8078723	17	38166879	T	0.61	−0.21	0.021	6.25E−23	PSMD3 - AC090844.3
rs9944715	18	43831259	A	0.25	−0.16	0.024	6.09E−11	C18orf25
rs1985157	19	18513594	T	0.59	−0.14	0.021	1.10E−10	LRRC25 - SSBP4
rs429358	19	45411941	T	0.84	0.52	0.029	1.50E−72	APOE
rs7412	19	45412079	C	0.92	−0.36	0.038	3.07E−21	APOE

Abbreviations: Chr: chromosome, bp: base pairs (Genome Reference Consortium Human Build 37), refA: reference/effect allele, freq: reference allele frequency, bJ, bJ_se, pJ: regression coefficient and the associated standard error and *p*-value, adjusted for other lead SNPs, SNPs overlapped between PhenoAgeAccel and BioAgeAccel in gray.

proportion of inflation not explained by polygenic heritability was small, 6.58% (SE = 3.78%).

The SNP-heritability for BioAgeAccel was estimated to be 12.39% (SE = 0.95%). Twenty lead SNPs were identified ($p < 5 \times 10^{-8}$) and 16 were nearby genes (Table 2). The strongest signal appeared in the *APOE* gene, tagged by the *APOE* isoform coding SNP rs7412. Several lead SNPs were associated with blood pressures. Other lead SNPs were associated with HbA1c, cardiovascular disease, and/or lipid biomarkers (Table S4).

The MAGMA gene set analysis identified 10 lipid-related gene sets at the Bonferroni-corrected level of 5%, including lipid homeostasis, lipid protein particle clearance, and triglyceride-rich plasma lipoprotein particle (Figure 2). None of the 53 tissues showed significant specificity in gene expression for the genes associated with BioAgeAccel (Figure 3).

2.3 | PhenoAgeAccel GWAS vs BioAgeAccel GWAS

PhenoAgeAccel and BioAgeAccel shared the lead SNPs, rs560887 (near *SPC25*, *G6PC2*), rs17321515 (near *AC091114.1*), rs16926246 (near *HK1*), and rs7412 (near *APOE*). Interestingly, three out of four common lead SNPs were oppositely associated with PhenoAgeAccel and BioAgeAccel.

- The rs560887T allele, associated with decreased HbA1c (An et al., 2014; Wheeler et al., 2017), is associated with decreased BioAgeAccel and PhenoAgeAccel.
- The rs16926246 C allele, associated with increased HbA1c (Soranzo et al., 2010), is associated with increased BioAgeAccel, but decreased PhenoAgeAccel.
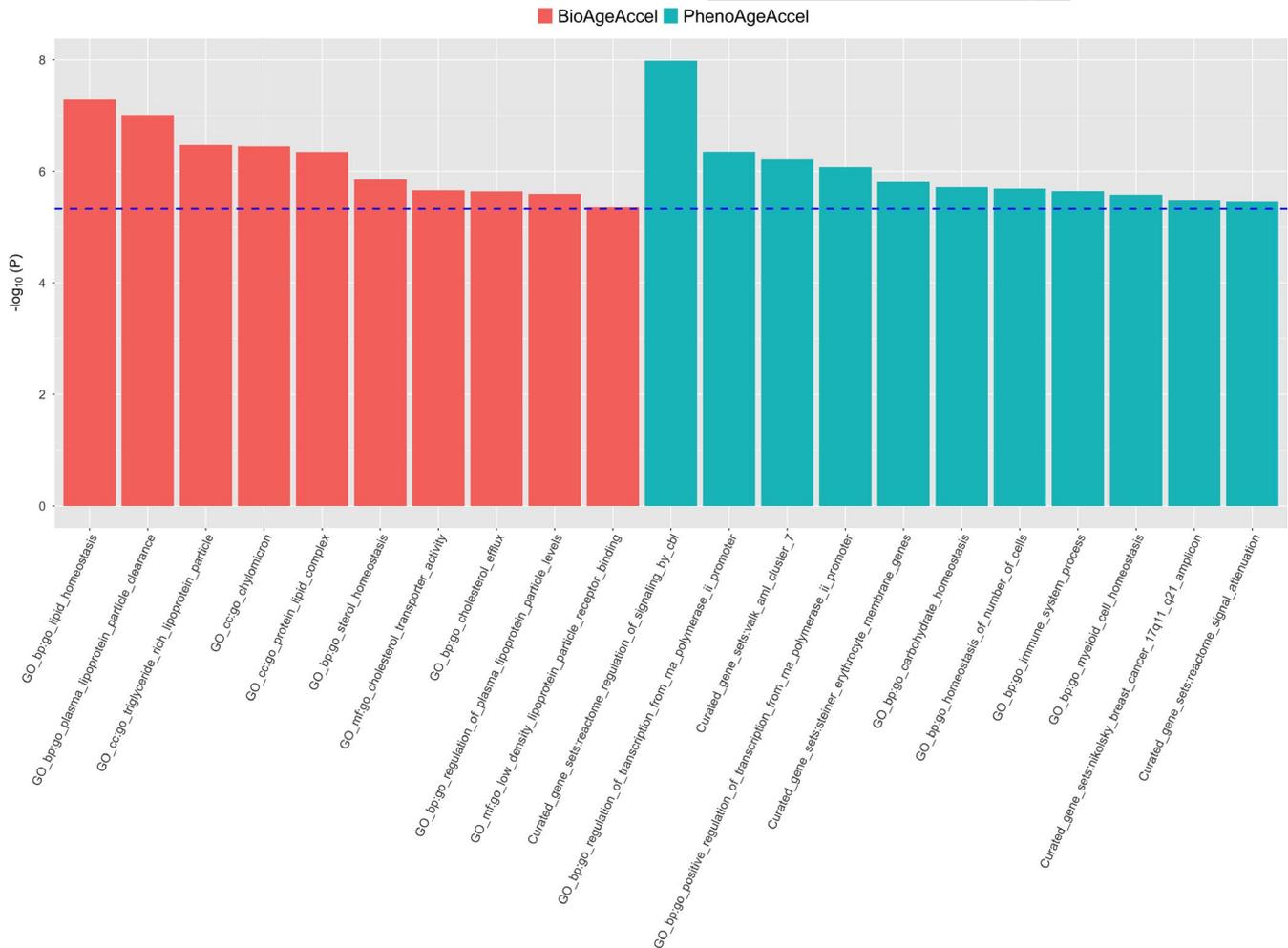
FIGURE 2 Significant gene sets identified by MAGMA for PhenoAgeAccel (in red) and BioAgeAccel (in green) at the Bonferroni-corrected level, 0.05/10,678 for 10,678 gene sets)

- The rs17321515 A allele, associated with increased triglycerides (Kathiresan et al., 2008), is associated with increased BioAgeAccel, but decreased PhenoAgeAccel.
- The rs7412T allele, or *APOE* e2 determined allele, associated with increased longevity (Deelen et al., 2019), is associated with decreased BioAgeAccel, but increased PhenoAgeAccel.

2.4 | Genetic associations

Among the biomarkers, PhenoAge acceleration was genetically highly correlated with RDW (r_g = 0.65), followed by CRP (r_g = 0.48), and then white blood cell count (r_g = 0.46) (Figure S2). BioAge acceleration was genetically highly correlated with systolic blood pressure (r_g = 0.84), followed by alkaline phosphatase (r_g = 0.43), and then CRP (r_g = 0.36) (Figure S2).

The genetic correlation between PhenoAgeAccel and BioAgeAccel was 0.42 (SE = 0.047). Both PhenoAgeAccel and BioAgeAccel had low genetic correlations with gastrointestinal diseases (GWAS summary statistics from (Liu et al., 2015)), prostate, and breast cancers (Michailidou et al., 2017; Schumacher et al.,

2018), and Alzheimer's disease (Jansen et al., 2019) (Figure S3). Both were genetically correlated with coronary artery disease (CAD) (CARDIoGRAMplusC2013 Consortium et al., 2013) (r_g = 0.27 with PhenoAgeAccel, r_g = 0.38 with BioAgeAccel), osteoarthritis (Zengini et al., 2018) (r_g =0.30 with PhenoAgeAccel, r_g = 0.25 with BioAgeAccel), stroke (Malik et al., 2018) (r_g = 0.30 with PhenoAgeAccel, r_g = 0.34 with BioAgeAccel), chronic kidney disease (Pattaro et al., 2016) (r_g = 0.35 with PhenoAgeAccel, r_g = 0.26 with BioAgeAccel), type II diabetes (Mahajan et al., 2018) (r_g = 0.36 with PhenoAgeAccel, r_g = 0.33 with BioAgeAccel), a 49-item frailty including pains and diseases (Atkins et al., 2019) (r_g = 0.34 with PhenoAgeAccel, r_g = 0.27 with BioAgeAccel), and parental mortality (Timmers et al., 2019) (r_g = 0.42 with PhenoAgeAccel, r_g = 0.45 with BioAgeAccel) (Figure S3).

PhenoAgeAccel (r_g = 0.44) was genetically more positively correlated with body mass index (BMI) than BioAgeAccel (r_g = 0.24). Waist circumstance and waist-hip ratio, adjusted for BMI and physical activity (Pulit et al., 2019), were not correlated with PhenoAgeAccel genetically, but a modest genetic correlation was found between BioAgeAccel and waist-hip ratio (r_g = 0.16). BioAgeAccel (systolic: r_g = 0.84, diastolic: r_g = 0.57) also was
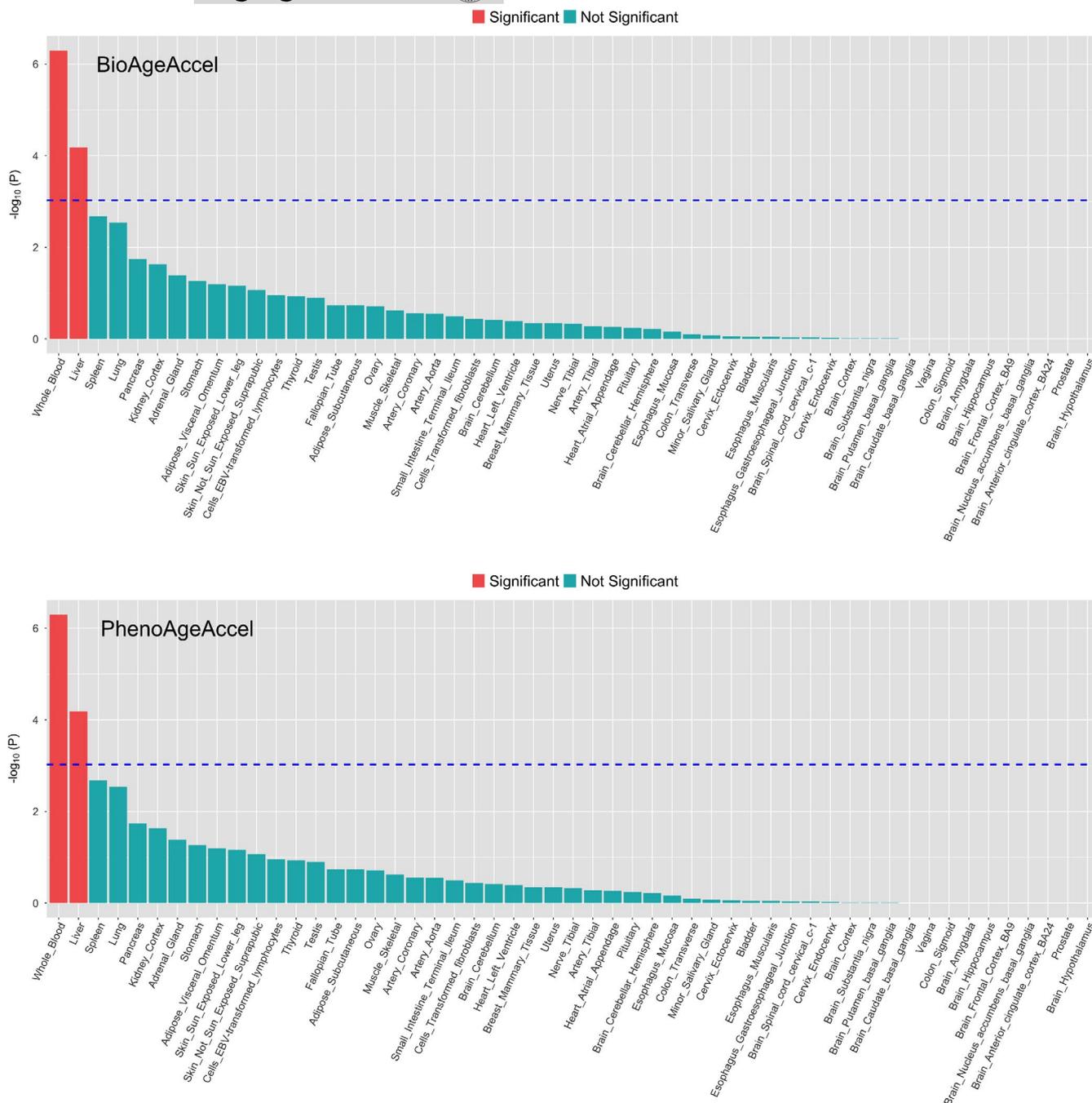
FIGURE 3 Association between tissue-specific gene expression and PhenoAgeAccel-gene or BioAgeAccel-gene association (*p*-values significant at the Bonferroni-corrected level 0.05/53 for 53 tissue types in red bars and others in green bars)

genetically more positively correlated with systolic and diastolic blood pressures than PhenoAgeAccel (systolic: r_g = 0.23, diastolic: r_g = 0.17). Genetically increased PhenoAgeAccel and BioAgeAccel were correlated with lower forced vital capacity (FVC) and forced expiratory volume in one second (FEV1) to a moderate degree but not with the FEV1/FVC ratio. The genetic correlations between PhenoAgeAccel or BioAgeAccel with heel bone mineral density and heart rate variability (the root mean square of the successive differences of inter beat intervals, RMSSD) (Nolte et al., 2017) were minimal (Figure S4).

PhenoAgeAccel was genetically more correlated than BioAgeAccel with hematology traits with no surprise as four hematological measures are included in PhenoAge versus none in BioAge (Figure S5). PhenoAgeAccel and BioAgeAccel were genetically associated with different cholesterol biomarkers: total cholesterol, LDL cholesterol, and apolipoprotein B with BioAgeAccel, and HDL cholesterol and apolipoprotein A-1 with PhenoAgeAccel. Similarly, BioAgeAccel was genetically more correlated than PhenoAgeAccel with the liver biomarkers of alanine aminotransferase, aspartate aminotransferase, and gamma glutamyltransferase, whereas albumin,

TABLE 2 Genetic loci associated with BioAgeAccel ($p < 5 \times 10^{-8}$) that can be mapped to genes

SNP	Chr	bp	refA	freq	bJ	bJ_se	pJ	Genes
rs17367504	1	11862778	A	0.84	0.07	0.012	9.03E−10	MTHFR
rs11591147	1	55505647	G	0.98	0.23	0.033	7.91E−13	PCSK9
rs541041	2	21294975	G	0.18	−0.09	0.011	2.33E−14	APOB - AC010872.2
rs560887	2	169763148	T	0.3	−0.06	0.009	9.83E−11	SPC25, G6PC2
rs16998073	4	81184341	A	0.71	−0.05	0.009	2.46E−08	PRDM8 - FGF5
rs1173771	5	32815028	A	0.4	−0.05	0.009	6.19E−09	NPR3 - AC025459.1
rs17477177	7	106411858	T	0.8	−0.09	0.011	4.62E−17	AC004917.1 - LINC02577
rs17321515	8	126486409	A	0.53	0.06	0.009	2.20E−12	AC091114.1
rs16926246	10	71093392	C	0.87	0.09	0.013	7.77E−13	HK1
rs2274224	10	96039597	G	0.57	0.05	0.009	2.41E−10	PLCE1, PLCE1-AS1
rs17249754	12	90060586	G	0.83	0.07	0.011	9.41E−09	ATP2B1
rs7497304	15	91429176	G	0.67	−0.05	0.009	1.89E−08	FES
rs55791371	19	11188153	A	0.88	0.14	0.013	4.95E−26	SMARCA4
rs58542926	19	19379549	C	0.92	0.11	0.016	1.78E−11	AC138430.1, TM6SF2
rs7412	19	45412079	C	0.92	0.26	0.016	3.16E−60	APOE
rs1327235	20	10969030	A	0.52	−0.05	0.009	1.02E−08	AL050403.2

Abbreviations: Chr: chromosome, bp: base pairs (Genome Reference Consortium Human Build 37), refA: reference/effect allele, freq: reference allele frequency, bJ, bJ_se, pJ: regression coefficient and the associated standard error and *p*-value, adjusted for other lead SNPs, SNPs overlapped between PhenoAgeAccel and BioAgeAccel in gray.

another liver biomarker, was more correlated with PhenoAgeAccel genetically. PhenoAgeAccel also was genetically more correlated than BioAgeAccel with creatinine, cystatin C, HbA1c, and CRP – biomarkers linked to kidney function, diabetes, and inflammation (Figure S6).

2.5 | Polygenic risk scores

5,198 SNPs (*p* < 0.0064) were selected to calculate PRSs for PhenoAgeAccel, which explained 0.50% of the variance in PhenoAge, in addition to 74.23% by other covariates, primarily baseline chronological age, plus sex, baseline assessment center, genotyping array type, and the first five genetic principal components. Similarly, 146,223 SNPs (*p* < 0.45) were selected for BioAgeAccel, accounting for 0.068% of the variance in BioAge, independent of 94.49% by baseline chronological age and other covariates. SNPs were selected for PRS to best explain the variance of PhenoAge or BioAge given other covariates were in the model. More SNPs were included in the PRS of BioAgeAccel than in the PRS of PhenoAgeAccel likely due to the small residual variance after accounting for other covariates and also small SNP effects in general for BioAgeAccel. While the variance independently explained by the PRS was minimal, the top 20% (high-risk class) and bottom 20% (low-risk class) of PRS showed distinct aging phenotypes. We here focus on the top 20% PRS results but include those comparing the top 40–60% to the bottom 20% in the forest plots (Figures 4, 5, S7, and S8). Overall, the disease risk or mean trait value increased or decreased with PRS quintiles, but the trend was

not always linear. We chose to present the results for the third and fifth quintile groups versus the first quintile group to preserve the trends, while also minimizing noise in the figure. A summary of the aging traits in the testing set is provided in Table S5.

The top 20% was compared to the bottom 20% of PhenoAgeAccel or BioAgeAccel PRS for a variety of aging traits adjusting for baseline chronological age, sex, baseline assessment center, genotyping array type, and the first five genetic principal components. The mean difference in PhenoAge between the top and bottom 20% of PhenoAgeAccel PRS (0.20 SD, 95% CI: 0.19 to 0.21 SD) was larger than the mean difference in BioAge between the top and bottom 20% of BioAge PRS (0.073 SD, 95% CI: 0.070 to 0.076 SD) in terms of either SD, PhenoAge SD = 9.56 and BioAge SD = 8.17. PhenoAge and BioAge share CRP, creatinine, and alkaline phosphatase in composition. Higher levels of the three biomarkers, particularly CRP, were observed in the top 20% than in the bottom 20% of PhenoAgeAccel or BioAgeAccel PRS (top left, Figure 4). The top-and-bottom mean difference of PhenoAgeAccel PRS was larger than that of BioAgeAccel PRS in biomarkers that appear in PhenoAge but not in BioAge, and vice versa (top left, Figure 4). Interestingly, the top 20% of PhenoAgeAccel PRS had lower mean cholesterol (−0.09 SD, 95% CI: −0.10 to −0.07 SD) than the bottom 20%, which was opposite for BioAgeAccel that the top 20% had 0.09 SD (95% CI: 0.08 to 0.11 SD) higher mean cholesterol than the bottom 20% (top left, Figure 4). The opposite trend was also found in mean corpuscular volume, with smaller top-and-bottom mean differences (top left, Figure 4). Additional biomarker and blood count PRS association results are provided in Figures S7 and S8.
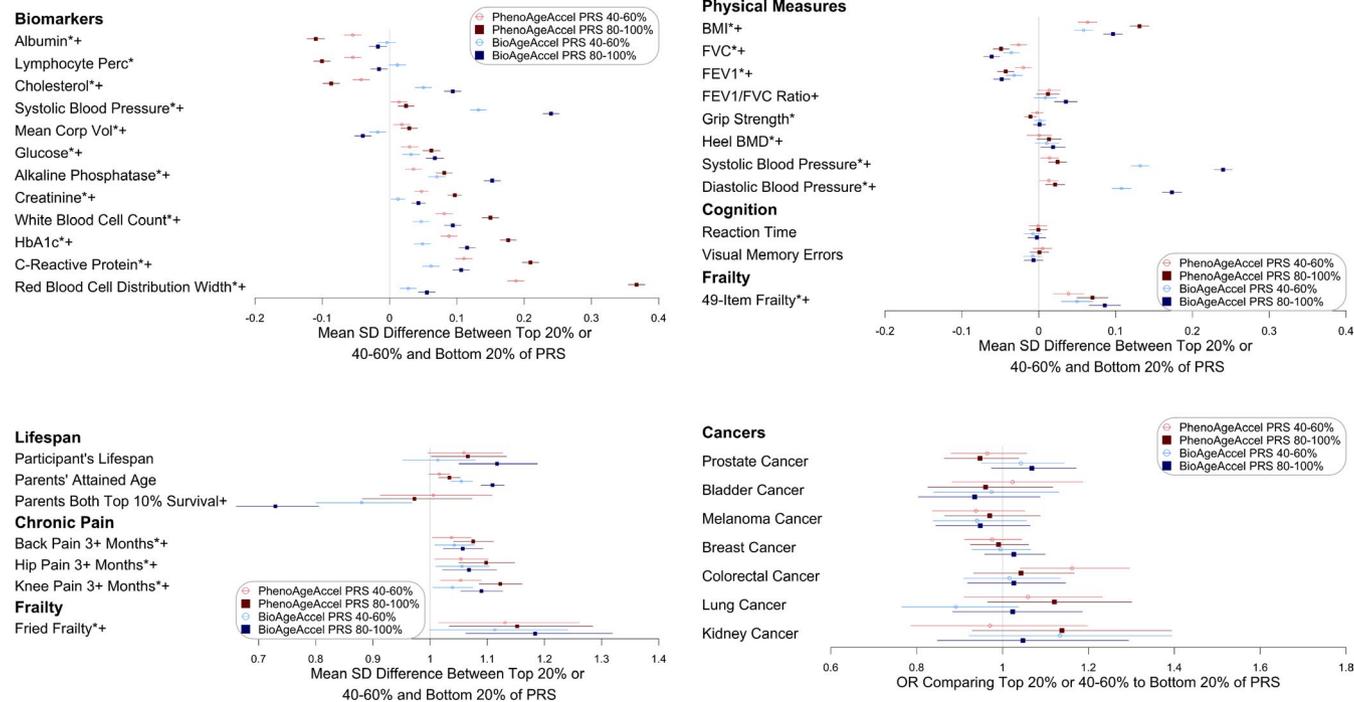
FIGURE 4 Comparisons between the top 20% or 40%-60% and the bottom 20% of PhenoAgeAccel (in red) or BioAgeAccel (in blue) Polygenic Risk Score (PRS) for biomarkers included in PhenoAge or BioAge and a variety of aging phenotypes (*significantly associated with the top 20% of PhenoAgeAccel PRS at the 5% false-discovery-rate adjusted level; +significantly associated with the top 20% of BioAgeAccel PRS at the 5% false-discovery adjusted level)

The top 20% of BioAgeAccel PRS were more likely to die early (HR = 1.12, 95% CI: 1.05 to 1.19) and have higher parental mortality risk (HR = 1.11, 95% CI: 1.09 to 1.13), and were less likely to have both parents survive to the top 10% of sex-specific lifespans (OR = 0.73, 95% CI: 0.66 to 0.81) than the bottom 20%. Similar results were observed for PhenoAgeAccel PRS, but with smaller risk ratios, participant mortality HR = 1.07 (95% CI: 1.00 to 1.13), parental mortality HR = 1.03 (95% CI: 1.01 to 1.05), and parental longevity OR = 0.97 (95% CI: 0.88 to 1.07) (bottom left, Figure 4). We also found higher likelihoods of chronic pain and Fried frailty (Fried et al., 2001) (frail, if 3 or more items checked) for the top 20% versus the bottom 20% when considering either PhenoAgeAccel or BioAgeAccel PRS. The top 20% of PhenoAgeAccel or BioAgeAccel PRS was associated with higher mean BMI and more deficits in a 49-item frailty (Williams et al., 2019) (a modified Rockwood frailty index, essentially accumulation of deficits), plus lower FVC and FEV1 (top right, Figure 4). The mean differences in systolic and diastolic blood pressures were much larger between the top 20% and the bottom 20% of BioAgeAccel PRS than that of PhenoAgeAccel PRS (top right, Figure 4).

Both PhenoAgeAccel and BioAgeAccel PRSs were not associated with prevalent cancers including prostate cancer, breast cancer, and colorectal cancer (bottom right, Figure 4). The associations of BioAgeAccel PRS were stronger than those of PhenoAgeAccel PRS with prevalent cardiovascular diseases, particularly CAD and hypertension (Figure 5). The odds ratio of CAD was 1.27 (95% CI: 1.22 to 1.33) and that of hypertension was 1.58 (95% CI: 1.53 to 1.62) comparing the top 20% to the bottom 20% of BioAge PRS. At

the biomarker levels, total cholesterol, low LDL cholesterol, apolipoprotein B, and triglycerides, risk factors of CAD, were elevated in the top 20% of BioAgeAccel PRS but reduced in the top 20% of PhenoAgeAccel PRS compared to the bottom 20% of each (Figure S7). These biomarker results suggested the association between PhenoAgeAccel PRS and CAD was likely driven by non-lipid mechanisms as indicated by the gene set analysis results.

PhenoAgeAccel PRS was more strongly associated than BioAgeAccel PRS with liver and kidney diseases (Figure 5) and the associated biomarkers, for example, albumin, total bilirubin, creatinine, and cystatin C (Figure S7), plus COPD, hypothyroidism, type I and type II diabetes, and rheumatoid arthritis (Figure 5). The odds ratio of type I diabetes was 1.76 (95% CI: 1.52 to 2.03) and that of type II diabetes was 1.38 (95% CI: 1.30 to 1.45) comparing the top 20% to the bottom 20% of PhenoAgeAccel PRS, and those comparing the top 20% to the bottom 20% of BioAgeAccel PRS were 1.12 (95% CI: 0.98 to 1.30) for type I diabetes and 1.27 (95% CI: 1.20 to 1.34) for type II diabetes. The associations of PhenoAgeAccel or BioAgeAccel PRS were minimal with bone diseases such as osteoporosis and osteoarthritis, age-related macular degeneration (AMD), anxiety and depression, and two neurological disorders, Parkinson's disease and delirium (Figure 5). A negative association was observed between PhenoAgeAccel PRS and dementia (OR = 0.80, 95% CI: 0.67 to 0.96). This was mainly driven by *APOE*, which when adjusted for completely accounted for the association (OR = 0.97, 95% CI: 0.80 to 1.16). Consistently, the genetic association between PhenoAgeAccel and Alzheimer's disease was not statistically significant (Figure S3).

FIGURE 5 Odds ratios (ORs) for diseases comparing the top 20% or 40%-60% to the bottom 20% of PhenoAgeAccel (in red) or BioAgeAccel (in blue) polygenic risk score (PRS) (*significantly associated with the top 20% of PhenoAgeAccel PRS at the 5% false-discovery-rate adjusted level; +significantly associated with the top 20% of BioAgeAccel PRS at the 5% false-discovery-rate adjusted level)

2.6 | Biological age measures and APOE genotypes

Some of the strongest associations for both PhenoAgeAccel and BioAgeAccel were with APOE isoform coding SNPs, but the effect directions were opposite. The APOE e2 determined T allele of rs7412 was associated with increased PhenoAgeAccel but decreased BioAgeAccel. Similarly, the rs429358 C allele (APOE e4), a risk factor for Alzheimer's disease, was associated with decreased PhenoAgeAccel but increased BioAgeAccel although the association with BioAgeAccel didn't reach genome-wide significance ($p = 1.3 \times 10^{-7}$).

Taking a step further, we associated PhenoAge and BioAge with APOE isoforms determined based on the genotypes of rs429358 and rs7412. For BioAge, results suggested that e2e3 and e2e2 were both associated with younger BioAge relative to the reference genotype (e3e3), while e3e4 and e4e4 exhibited higher BioAges, where the

results were adjusted for baseline chronological age, sex, genotyping array type, baseline assessment center, and the first five genetic principal components (Figure 6). When comparing APOE genotypes as a function of PhenoAge, we find the reverse—e2e3 and e2e2 appeared older than e3e3, whereas e3e4, and e4e4 appeared younger.

To further disentangle the associations between APOE genotypes and accelerated aging by the two biological age measures, we examined the associations between APOE genotypes and the individual biomarkers that make up the composites. We found that the trend of mean BioAge (e4e4 > e4e3 > e3e3 > e2e4 > e2e3 > e2e2) also held for total cholesterol, which was most strongly associated with APOE genotypes among the biomarkers of BioAge (Figure 6). When adjusting for total cholesterol, the trend of mean BioAge was reversed, that is, e3e4 and e4e4 younger than e2e3 and e2e2, suggesting that decelerated BioAge associated with e2 was driven by differences in plasma total cholesterol levels. The biomarkers that
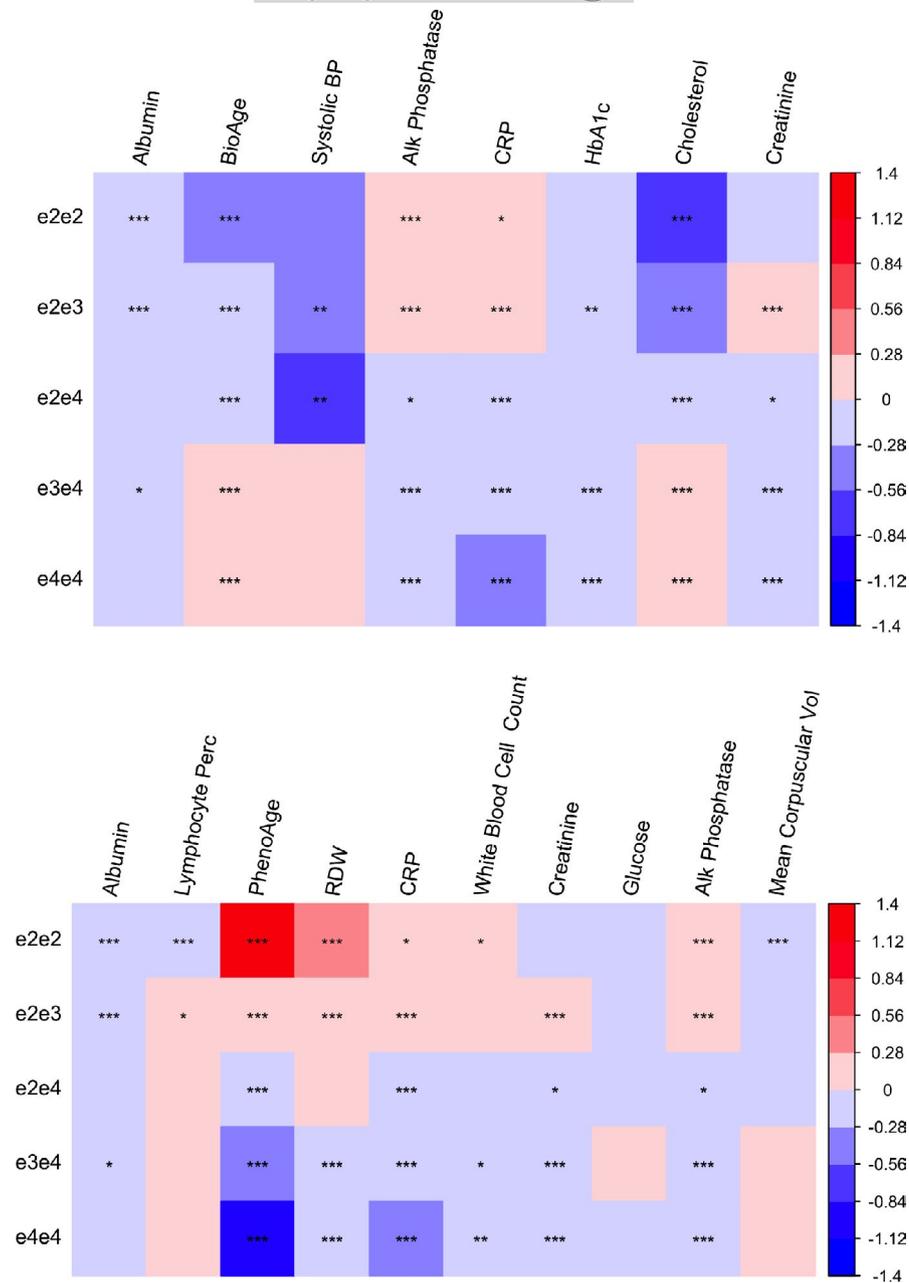
appeared to show inverse associations (similar to PhenoAge) were RDW, CRP, alkaline phosphatase, creatinine, and white blood cell count. For all of these biomarkers there was a trend towards higher levels among participants with e2 alleles and lower levels among those with e4 alleles (Figure 6).

2.7 | Replication in the US heath retirement study

Complete genetic and 2016 biomarker data to estimate PhenoAge and BioAge were available for 5,572 and 1,782 participants, respectively. A summary for the PhenoAge and BioAge samples is provided in Table S6. We are underpowered to detect genome-wide significance based on the HRS sample size and the UKB effect sizes for most lead SNPs. However, we do find that in HRS both the *APOE*

SNPs show similar effect sizes to what was observed in UKB for the associations with PhenoAgeAccel. For instance, rs429358 in HRS has an effect size of $b = 0.58$ ($p = 0.01$), which is similar to the effect size in UKB ($b = 0.55$, $p = 7.8 \times 10^{-83}$). Similarly, rs7412 has an effect size of $b = -0.38$ ($p = 0.20$) in HRS and $b = -0.44$ ($p = 1.6 \times 10^{-31}$) in UKB (Tables S7 and S8).

We also compared *APOE* genotypes, using e3e3 as the reference (Table S9). Overall, the three e4 genotypes showed a trend toward decreased PhenoAge, similar to what was observed in UKB, although only e2e4 significantly differed from e3e3 ($p = 0.039$), while e3e4 was marginally decreased ($p = 0.054$), and e4e4 was not significant ($p = 0.912$). Similarly, e2e2 and e2e3 showed a trend toward increased PhenoAge, with e2e2 being marginally significant ($p = 0.058$) and e2e3 not significant ($p = 0.291$). Overall, while we were underpowered to replicate the findings for all genotypes, the trends remained

unchanged. Further, given that this was a replication, a one-sided p-value would be justified, and in that case, three of the genotypes significantly differed from e3e3 in the expected direction.

Finally, the GRS using the PhenoAgeAccel lead SNPs from UKB was significantly associated with PhenoAgeAccel in HRS. The mean PhenoAge was increased by 0.80 years (95% CI: 0.58 to 1.02, p = 1.42×10^{-12}) per SD increase in the GRS of PhenoAgeAccel, with adjustment for age in 2016, sex, and PC1-PC5. rs7412 was strongly associated with BioAgeAccel in UKB but the association cannot be replicated in HRS (Table S8), which also led to negative results with *APOE* genotypes (Table S9) and the GRS, 0.07 year increase in mean BioAge (95% CI: −0.03 to 0.17, p = 0.150) per SD increase in the GRS of BioAgeAccel, adjusted for age in 2016, sex, and PC1-PC5.

We hypothesized that part of the issue of replication for BioAge may stem from the lack of data for two of the measures – HbA1c and systolic blood pressure, which were only available for half the sample due to the study design of HRS. For sensitivity analysis, we imputed HbA1c in 2016 using HbA1c in 2014 for the other subsample and similarly for systolic blood pressure for BioAge. In the model, we included a dummy indicator to signify which samples used the 2014 systolic blood pressure and HbA1c to calculate BioAge, which was not statistically significant (p > 0.05) across models. The sample size increased from 1,782 to 4,909, where the mean age and BioAge decreased but the sex distribution and biomarker statistics were similar (Table S6). The associations between BioAge and BioAge GRS or *APOE* genotypes remained statistically insignificant (p > 0.05). One SD increase in BioAgeAccel GRS increased the mean BioAge by 0.05 years (95% CI: 0.01 to 0.11, p = 0.095), adjusted for age in 2016, sex, PC1-PC5, and an indicator of using 2014 HbA1c and systolic blood pressure data. Additionally, the trend of associations between BioAge and *APOE* genotypes were inconclusive still (Table S9).

3 | DISCUSSION

Overall, our analysis using the UK Biobank biomarker data identified both overlapping and distinct genetic underpinnings of two widely applied biological age measures. Our results suggested that although the estimated heritability is similar for PhenoAgeAccel (14.45%) and BioAgeAccel (12.39%) with the genetic correlation being 0.42, these two measures capture distinct aging domains with different genetic determinants, as a result of their differential biomarker compositions. SNPs associated with BioAgeAccel (p < 5×10^{-8}) tended to relate to systolic blood pressure and lipid biomarkers, with enrichment analysis pointing to an increased proportion of genes involved in lipid homeostasis, plasma lipoprotein particle clearance, chylomicron, sterol homeostasis, and cholesterol transport activity. Conversely, SNPs associated with PhenoAgeAccel were shown to relate to CRP, white blood cell count, and RDW, and were enriched in biological processes involved in regulation of cell signaling by *CBL*, transcription, immune system process, and myeloid cell homeostasis.

The immune/inflammation versus lipid findings for PhenoAgeAccel and BioAgeAccel, respectively, were also recapitulated when comparing the associations between PRS and age-related outcomes. Results suggested that the top 20% of PhenoAgeAccel and BioAgeAccel PRS were differentially linked to a variety of diseases. For instance, BioAgeAccel PRS outperformed PhenoAgeAccel PRS in prioritizing cardiovascular and all-cause mortality risk in this young cohort, while PhenoAgeAccel PRS showed more robust associations than BioAgeAccel PRS for liver/kidney diseases, and chronic inflammatory and autoimmune diseases. The stronger link between BioAgeAccel PRS and all-cause mortality (compared to PhenoAgeAccel PRS) may be driven in part by its association with cardiovascular disease, which is the leading cause of death in the UK. By comparison, the diseases associated with PhenoAgeAccel PRS tend to contribute to major morbidity, while being less common causes of death. This may suggest that individuals genetically predisposed to accelerated BioAge may be more likely to experience shortened lifespan, while those genetically predisposed to accelerated PhenoAge, may not experience major reductions in lifespan, but may experience decreased healthspan (disease-free life expectancy). The hypothesis needs to be tested in older adults, however. Of note, accelerated aging is not only determined by genetics but also by environment. Interestingly, when considering the actual values rather than the PRS, accelerated PhenoAge is more strongly associated with all-cause mortality than accelerated BioAge in UK Biobank, which implies that the association between accelerated PhenoAge and all-cause mortality may be explained to a larger degree by the environmental components.

The PhenoAgeAccel PRS was also related to dementia, but in the opposite than the expected direction, such that individuals with increased PhenoAgeAccel had reduced odds of dementia. This result was almost entirely driven by the association between PhenoAgeAccel and *APOE*, which is the most well-known genetic risk factor for late-onset Alzheimer's disease (LOAD). Our results suggested that while PhenoAgeAccel and BioAgeAccel were both associated with the two *APOE* isoform coding SNPs (rs429358 and rs7412), the relationships were inverse. For instance, the *APOE* e4 allele is traditionally associated with adverse health outcomes, including an increased risk of Alzheimer's disease, cardiovascular disease, and reduced life expectancy, while the e2 allele confers protection. However, in our results, we observed increased PhenoAgeAccel associated with e2 genotypes and decreased PhenoAgeAccel associated with e4 genotypes, relative to the common e3e3 genotype. This paradoxical result was also found for a number of the biomarkers that make up PhenoAge, which likely explains this finding. For instance, *APOE* e2 allele was associated with higher CRP, RDW, alkaline phosphatase, creatinine, and white blood cell count, while *APOE* e4 allele was generally associated with lower levels of these biomarkers (Kuo et al., 2020). These biomarkers are not specific to physiological functions or diseases but are more or less associated with inflammation, which suggests that e2 genotypes may be prone to higher inflammation levels on average, compared to e3e3s and e4 genotypes. Our biomarker specific finding is supported by a previous study via CRP (one of the most popular inflammation markers) (Hubacek et al., 2010). Although the underlying mechanism is

not fully understood, it has been reported that high inflammation (high CRP) is associated with increased risk of Alzheimer's disease among e4 genotypes (e3e4 or e4e4) but less so among non-e4 genotypes (Tao et al., 2018). This may suggest that increased inflammation levels in e2 genotypes does not increase their Alzheimer's disease risk, but that higher PhenoAge and inflammation among e4 genotypes could increase Alzheimer's disease risk. Contrary to PhenoAgeAccel, BioAgeAccel showed an expected association with *APOE* that consisted of decelerated aging among participants with e2 alleles and accelerated aging among participants with e4 alleles. This association was accounted for by higher levels of total cholesterol among those with increased BioAgeAccel. This is in-line with *APOE* known function as a transporter of extracellular cholesterol and the existing evidence suggesting those with the e2 allele exhibit reduced circulating cholesterol, particularly low-density lipoproteins (LDL) (Kuo et al., 2020).

We were able to replicate the association in HRS between PhenoAge and PhenoAgeAccel GRS using the lead SNPs from UKB and showed similar trends for *APOE*, in which e2 is associated with higher PhenoAgeAccel and e4 is associated with lower PhenoAgeAccel. The unsuccessful replication for BioAgeAccel may be explained by a small sample size or the instruction from the interviewer to the respondent with a high systolic blood pressure to consult their physician as soon as possible (stated in the Documentation of Physical Measures, Anthropometrics and Blood Pressure in the Health Retirement Study), which drove systolic blood pressures (SBPs) towards the mean, reduced the BioAge-SBP correlation (Pearson correlation 0.34 in HRS versus 0.51 in UKB) and associations between genetic variants and BioAgeAccel, particularly those (e.g., rs7412) associated with systolic blood pressure. Systolic blood pressure, as we have shown in Figure S2, had the highest genetic correlation with BioAgeAccel (r_g = 0.84, Figure S2), compared to other BioAge biomarkers. At the end, more replication studies are needed to validate the UKB findings although it is challenging to find a large cohort with both genetic and biomarker data.

Inevitably, our study has limitations. The UK Biobank participants are healthier than the general population (Fry et al., 2017); therefore, are less susceptible to accelerated aging. The disease status was determined based on self-reported doctor diagnoses at baseline and inpatient electronic health records to 2017. Given that some participants were still relatively young and will likely go on to develop late-onset morbidity, this will contribute to misclassification, which could lead to biased associations, towards the null if the misclassification is non-differential. Nevertheless, when disease prognostic biomarkers were analyzed, we observed consistent results. Last but not least, our findings are based on European-descent participants and may not be generalizable to other ancestry populations.

Overall, the mapped genes and enriched genes sets highlight that these two biological age measures may capture different aspects of the aging process—cardiometabolic by BioAge and inflammaging/immunoscenece by PhenoAge. Nevertheless, PhenoAgeAccel and BioAgeAccel PRSs are not disease-specific and can be used to prioritize genetic risk for multiple morbidity or mortality outcomes—particularly cardiovascular diseases and all-cause mortality via BioAge, and liver

or kidney diseases, COPD, rheumatoid arthritis, hypothyroidism, and type I and type II diabetes via PhenoAge. Our findings confirm the hypothesis that individuals may age in different ways, due in part to different underlying genetic susceptibility. In moving forward, understanding personalized aging susceptibility phenotypes has important implications for primary and secondary disease interventions.

4 | METHODS

4.1 | UK Biobank

Over 500,000 participants between the ages of 40 and 70 were recruited by UK Biobank from 2006 to 2010 (Bycroft et al., 2018; Sudlow et al., 2015), of which, over 90% of the cohort were European-descent. Phenotypes considered in this study include participant mortality, parental lifespan, cognitive function, physical measures, and diseases. The death status was determined based on death certificate data, updated to March 2020 for all participants. Some deaths were recorded in April 2020 but the mortality data for that month is incomplete. The disease diagnosis was confirmed based on self-reported doctor diagnoses at baseline, cancer registry data to 2016, and hospital admission records from 1996 to 2017. A list of disease ICD-10 codes used to identify diseases is provided in Table S10. At recruitment, participants completed online questionnaire and physical measurements and their biological samples were collected for biomarker assays. Physical measurements were described elsewhere (Kuo et al., 2020). A full list and technical details are available via the UK Biobank Biomarker Panel and UK Biobank Haematology Data Companion Document.

4.2 | Genetic data

DNA was extracted from blood samples and was genotyped using Affymetrix UK BiLEVE Axiom array for the first ~50,000 participants and Affymetrix UK Biobank Axiom array for the remaining cohort – the two arrays have over 95% content overlap (Bycroft et al., 2018). Imputation was performed by the UK Biobank team using the reference panels of 1000 Genomes and the Haplotype Reference Consortium (HRC), yielding ~93 million variants in 487,442 participants. Of whom, participants (*n* = 968) with unusually high heterozygosity or missing genotype calls were further removed (Bycroft et al., 2018).

4.3 | Biological age measures

PhenoAge and BioAge were previously trained using data from a US cohort (NHANES III) in separate papers (Levine, 2013; Levine et al., 2018) that are available in the literature. PhenoAge was trained for mortality as a surrogate of biological age using all available biomarkers (*n* = 42) in the NHANES III, where 9 biomarkers selected by Cox penalized regression model and chronological age were used

to construct PhenoAge in a parametric proportional hazards model based on the Gompertz distribution. The 10-year mortality risk was converted to the unit of years to derive PhenoAge (Levine et al., 2018). The formula of PhenoAge is given by

$$\text{PhenoAge} = 141.50 + \frac{\ln\left\{-0.00553 \times \ln(1 - \text{mortality risk})\right\}}{0.09165}$$
$$= 141.50 + \frac{\ln\left\{-0.00553 \times \frac{(-1.51714) \times \exp(xb)}{0.0076927}\right\}}{0.09165}$$

where mortality risk = $1 - \exp\{-\exp(xb) [\exp(120\gamma) - 1]/\gamma\}$, $\gamma = 0.0076927$ and $xb = -19.907 - 0.0336 \times$ albumin $+ 0.0095 \times$ creatinine $+ 0.1953 \times$ glucose $+ 0.0954 \times \ln(\text{CRP}) - 0.0120 \times$ lymphocyte percentage $+ 0.0268 \times$ mean corpuscular volume $+ 0.3306 \times$ RDW $+ 0.00188 \times$ alkaline phosphatase $+ 0.0554 \times$ white blood cell count $+ 0.0804 \times$ age, estimated using the NHANES III data and age denotes the chronological age.

BioAge (Levine, 2013) was trained for the biological age surrogate of chronological age, using the NHANES III data. 21 biomarkers were preselected for BioAge based on prior knowledge regarding their role or dependency in the aging process and a significant correlation (r) with chronological age ($p < 0.05$) in the NHANES III. Of which, 10 were correlated with chronological age with Pearson correlation coefficients greater than 0.1 or less than −0.1. 3 biomarkers were further removed for not significantly loaded on the first principal component in men or in women based on the principal components analysis results using the 10 biomarkers. Seven biomarkers and chronological age were used to calculate BioAge by applying an algorithm previously proposed by Klenmera and Doubal (Klemera & Doubal, 2006),

$$\text{BioAge} = \frac{\sum_{j=1}^{7} (x_j - q_j)\left(\frac{k_j}{s_j^2}\right) + \frac{\text{age}}{31.63}}{\sum_{j}^{7} \left(\frac{k_j}{s_j}\right)^2 + \frac{1}{31.63}}$$

where x_j denotes the level of j-th biomarker, with the corresponding parameters q_j, k_j, and s_j provided in Table S11. age here, again, denotes the chronological age.

The PhenoAge or Bioage biomarkers in UK Biobank are summarized in Table S1, after setting the bottom 1% of values to the first percentile and the top 1% to the 99th percentile to correct the skewness of distributions. PhenoAge and BioAge were calculated per individual, applying the equations above to the UK Biobank biomarker data. Biological age acceleration was estimated by the residual of PhenoAge or BioAge after subtracting the effect of chronological age using a linear regression model, termed PhenoAgeAccel and BioAgeAccel, respectively.

4.4 | Included samples

Participants of European descent were included, identified using genetic principal components analysis in detail in Thompson and colleagues (Thompson et al., 2019). Additionally, one in third-degree or closer pairs were removed, identified via pairwise kinship coefficients. The sample was randomly split into a training and a testing set, with a 1 to 2 ratio. The training set was used to perform genome-wide association analysis with the results being used to create PRSs in the testing set to evaluate the use for risk stratification for age-related outcomes.

4.5 | SNP quality control

Of 93,095,623 genotyped or imputed SNPs, 16,446,666 SNPs passed the quality control, where SNPs were excluded if meeting any of the criteria: (1) imputation information score <0.3, (2) minor allele frequency <0.1%, (3) Hardy–Weinberg equilibrium test p-value significant at the Bonferroni-corrected level, (4) missing imputation information score, minor allele frequency, or Hardy–Weinberg equilibrium test result. The SNP summary statistics were calculated using the QCTOOL software version 2.

4.6 | Genome-wide association analysis

The association between accelerated PhenoAge or BioAge with each SNP was examined using an efficient Bayesian linear mixed effects model (BOLT-LMM software version 2.2) (Loh et al., 2015) for the outcome of PhenoAge or BioAge with additive allelic effect of the candidate SNP, and other fixed effects: chronological age (to make the case of accelerated biological age), sex, genotyping array type, and assessment center, plus random polygenic and environment effects. By default, the LD scores included in the BOLT-LMM for European-ancestry samples were used to calibrate the BOLT-LMM statistic. SNP p-values smaller than 5×10^{-8} were deemed to be statistically significant. Manhattan and tile–quantile (Q–Q) plots were created for visualization using the CMplot R package. The genomic inflation due to population stratification or cryptic relatedness was evaluated by linkage disequilibrium (LD) score regression (Bulik-Sullivan et al., 2015), where SNPs were filtered to the HapMap3 SNPs, well imputed in most studies to avoid bias from poor imputation quality. The LD scores were downloaded from the url (https://data.broadinstitute.org/alkesgroup/LDSCORE/), precomputed using the European data from the 1000 genome project phase 3.

We performed a stepwise model selection procedure on the genome-wide SNP summary statistics to identify independent signals ($p < 5 \times 10^{-8}$) using the COJO (Conditional & Joint association analysis) function in the GCTA (Genome-wide Complex Trait Analysis) software version 1.92.1 beta6 Linux (Yang et al., 2012). SNPs more than 10,000 kb away from each other were assumed to be in complete linkage equilibrium. As SNPs were selected, those with multiple regression R^2 greater than 0.9 with already pre-selected SNPs were excluded, so as not to include redundant signals from high LD. The loci marked by the selected SNPs were mapped to genes based on

GRCh37/hg19 coordinates, and were used in searches for published GWAS associations based on GWAS catalog (Buniello et al., 2019).

4.7 | Gene enrichment analysis

The GWAS p-values were analyzed by Multi-marker Analysis of GenoMic Annotation (MAGMA) in FUMA (Functional Mapping and Annotation) (Watanabe et al., 2017) to perform a comparative gene-set analysis to test if genes in the gene set were more strongly associated with PhenoAgeAccel or BioAgeAccel than others, for 10,678 gene sets (curated gene sets: 4,761, GO terms: 5,917) from the MsigDB v6.2. Additionally, a gene-property analysis was performed to test for positive relationships (one-sided test) between tissue-specific gene expression profiles and gene associations with PhenoAgeAccel or BioAgeAccel, using 53 tissue types from the GTEx repository version. Both test results were adjusted for multiple testing using the Bonferroni correction method.

4.8 | Genetic correlations

Genetic correlations of PhneoAgeAccel or BioAgeAccel were calculated by LD score regression (Bulik-Sullivan et al., 2015) using GWAS summary statistics, filtered to HapMap3 SNPs. GWAS summary statistics were downloaded from previous published GWAS. Those of biomarkers, not limited to PhenoAge or BioAge biomarkers, were downloaded from the Ben Neale Lab round 2, where biomarkers were transformed by the rank-based inverse normal transformation, and the SNP-biomarker associations were adjusted for age, age^2, sex, age × sex, age^2 × sex and the top 20 genetic principal components in over 361,000 UK Biobank participants.

4.9 | Polygenic risk scores

The PRSice-2 software version 2.2.2 (Choi & O'Reilly, 2019) was used to perform polygenic risk score (PRS) analysis. SNPs were clumped to obtain SNPs in low LD ($r^2 < 0.1$) in a 250 base-pair window. SNPs with p-values smaller than a threshold were used to calculate the PRS, sum of the effect alleles associated with accelerated aging, weighted by the effect size. The optimal threshold was chosen by a grid search from $1 × 10^{-5}$ to 0.5 with the increment of $1 × 10^{-5}$ plus 1, such that the variance of PhenoAge or BioAge in the testing set was best explained by PRS, in addition to that by baseline chronological age, sex, genotyping array type, baseline assessment center, and the first five genetic principal components. Subjects were equally divided into five groups by the PRS, where the top 20% (high-risk class) was compared to the bottom 20% (low-risk class) for a variety of aging traits (n = 111). The association analysis was conducted using a regression model, with adjustment for baseline chronological age, sex, genotyping array type, baseline assessment center, and the first five genetic principal components: (1) Cox regression models for lifespan outcomes with hazard ratios reported; (2) logistic regression models for binary outcomes such as disease and pain outcomes with odds ratios reported; (3) linear regression models for continuous outcomes such as physical measures and 49-item frailty that were z-transformed so the regression coefficients associated with PRS quintiles were unit-free and represented the mean differences between quintiles in standard deviations (SDs). Log (e.g., cognitive function measures and 49-item frailty) or the inverse normal transformation (e.g., blood counts and biomarkers) may be applied before the z-transformation to correct the distributional skewness. Traits that show significant differences between the top 20% and the bottom 20% of PRS were highlighted at the 5% false-discovery-rate adjusted level. Results comparing the top 40–60% to the bottom 20% of PRS were also included in forest plots to examine the dosage effects.

4.10 | Replication in the US heath retirement study

For replication, we used European-descent participants in the US Health Retirement Study (HRS) (Fisher & Ryan, 2018; Sonnega et al., 2014) to test the associations between PhenoAgeAccel/BioAgeAccel and the UKB lead SNPs individually, their genetic risk score (GRS), or *APOE* genotypes. Linear regression models were used, with adjustment for age, sex, and the first five genetic principal components (PCs) (Pilling et al., 2017).

CONFLICT OF INTEREST

MEL is named on patent applications for epigenetic clocks and holds licenses for the clocks she has developed. MEL also serves as the Bioinformatics Advisor for Elysium Health.

AUTHOR CONTRIBUTIONS

Study design: CLK, MEL; data analysis: CLK; manuscript preparation: all.

ORCID

Chia-Ling Kuo https://orcid.org/0000-0003-4452-2380

Janice L. Atkins https://orcid.org/0000-0003-4919-9068

Morgan E. Levine https://orcid.org/0000-0001-9890-9324

REFERENCES

Ahadi, S., Zhou, W., Schüssler-Fiorenza Rose, S. M., Sailani, M. R., Contrepois, K., Avina, M., Ashland, M., Brunet, A., & Snyder, M. (2020). Personal aging markers and ageotypes revealed by deep longitudinal profiling. *Nature Medicine*, 26, 83–90.

An, P., Miljkovic, I., Thyagarajan, B., Kraja, A. T., Daw, E. W., Pankow, J. S., Selvin, E., Kao, W. H. L., Maruthur, N. M., Nalls, M. A., Liu, Y., Harris, T. B., Lee, J. H., Borecki, I. B., Christensen, K., Eckfeldt, J. H., Mayeux, R., Perls, T. T., Newman, A. B., & Province, M. A. (2014). Genome-wide association study identifies common loci influencing circulating glycated hemoglobin (HbA1c) levels in non-diabetic subjects: The Long Life Family Study (LLFS). *Metabolism*, 63(4), 461–468.

Atkins, J. L., Jylhava, J., Pedersen, N., Magnusson, P., Lu, Y., Wang, Y., Hagg, S., Melzer, D., Williams, D., & Pilling, L. C. (2019). A Genome-Wide Association Study of the Frailty Index Highlights Synaptic Pathways in Aging. *medRxiv*, 19007559. Available at: http://medrxiv.org/content/early/2019/09/25/19007559.abstract

Belsky, D. W., Moffitt, T. E., Cohen, A. A., Corcoran, D. L., Levine, M. E., Prinz, J. A., Schaefer, J., Sugden, K., Williams, B., Poulton, R., & Caspi, A. (2018). Eleven telomere, epigenetic clock, and biomarker-composite quantifications of biological aging: do they measure the same thing? *American Journal of Epidemiology*, 187(6), 1220–1230.

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Consortium SWG of the PG, Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47(3), 291–295.

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P. L., Amode, R., Guillen, J. A., Riat, H. S., Trevanion, S. J., Hall, P., Junkins, H., ... Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, 47(D1), D1005–D1012.

Butler, R. N., Sprott, R., Warner, H., Bland, J., Feuers, R., Forster, M., Fillit, H., Harman, S. M., Hewitt, M., Hyman, M., Johnson, K., Kligman, E., McClearn, G., Nelson, J., Richardson, A., Sonntag, W., Weindruch, R., & Wolf, N. (2004). Aging: The reality: biomarkers of aging: from primitive organisms to humans. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59, B560–B567.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562, 203–209.

CARDIoGRAMplusC4D Consortium, Deloukas, P., Kanoni, S., Willenborg, C., Farrall, M., Assimes, T. L., Thompson, J. R., Ingelsson, E., Saleheen, D., Erdmann, J., Goldstein, B. A., Stirrups, K., König, I. R., Cazier, J.-B., Johansson, A., Hall, A. S., Lee, J.-Y., Willer, C. J.,

Chambers, J. C., ... Samani, N. J. (2013). Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature Genetics*, 45(1), 25–33.

Choi, S. W., & O'Reilly, P. F. (2019). PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*, 8(7), giz082. https://doi.org/10.1093/gigascience/giz082

Deelen, J., Evans, D. S., Arking, D. E., Tesi, N., Nygaard, M., Liu, X., Wojczynski, M. K., Biggs, M. L., van der Spek, A., Atzmon, G., Ware, E. B., Sarnowski, C., Smith, A. V., Seppälä, I., Cordell, H. J., Dose, J., Amin, N., Arnold, A. M., Ayers, K. L., ... Murabito, J. M. (2019). A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nature Communications*, 10(1), 3669.

Ferrucci, L., Gonzalez-Freire, M., Fabbri, E., Simonsick, E., Tanaka, T., Moore, Z., Salimi, S., Sierra, F., & de Cabo, R. (2020). Measuring biological aging in humans: A quest. *Aging Cell*, 19(2), e13080. https://doi.org/10.1111/acel.13080

Fisher, G. G., & Ryan, L. H. (2018). Overview of the health and retirement study and introduction to the special issue. *Work Aging Retire*, 4(1), 1–9.

Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., Seeman, T., Tracy, R., Kop, W. J., Burke, G., & McBurnie, M. A. (2001). Frailty in older adults: evidence for a phenotype. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 56(3), M146–M157.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., & Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of UK biobank participants with the general population. *American Journal of Epidemiology*, 186(9), 1026–1034.

Hubacek, J. A., Peasey, A., Pikhart, H., Stavek, P., Kubinova, R., Marmot, M., & Bobak, M. (2010). APOE polymorphism and its effect on plasma C-reactive protein levels in a large general population sample. *Human Immunology*, 71, 304–308.

Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., Sealock, J., Karlsson, I. K., Hägg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I. S., Andersen, F., Bergh, S., Bettella, F., ... Posthuma, D. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics*, 51(3), 404–413.

Jylhävä, J., Pedersen, N. L., & Hägg, S. (2017). Biological age predictors. *EBioMedicine*, 21, 29–36. https://doi.org/10.1016/j.ebiom.2017.03.046.

Kathiresan, S., Melander, O., Guiducci, C., Surti, A., Burtt, N. P., Rieder, M. J., Cooper, G. M., Roos, C., Voight, B. F., Havulinna, A. S., Wahlstrand, B., Hedner, T., Corella, D., Tai, E. S., Ordovas, J. M., Berglund, G., Vartiainen, E., Jousilahti, P., Hedblad, B., ... Orho-Melander, M. (2008). Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature Genetics*, 40(2), 189–197.

Klemera, P., & Doubal, S. (2006). A new approach to the concept and computation of biological age. *Mechanisms of Ageing and Development*, 127, 240–248.

Kuo, C. L., Pilling, L. C., Atkins, J. L., Kuchel, G. A., & Melzer, D. (2020). ApoE e2 and aging-related outcomes in 379,000 UK Biobank participants. *Aging (Albany., NY)*. 12(12), 12222–12233. https://doi.org/10.18632/aging.103405

Levine, M. E. (2013). Modeling the rate of senescence: Can estimated biological age predict mortality more accurately than chronological age? *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 68(6), 667–674.

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., Hou, L., Baccarelli, A. A., Stewart, J. D., Li, Y., Whitsel, E. A., Wilson, J. G., Reiner, A. P., Aviv, A., Lohman, K., Liu, Y., Ferrucci, L., & Horvath, S. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany., NY)*. 10(4), 573–591.

Liu, J. Z., Van Sommeren, S., Huang, H., Ng, S. C., Alberts, R., Takahashi, A., Ripke, S., Lee, J. C., Jostins, L., Shah, T., Abedian, S., Cheon, J. H., Cho, J., Daryani, N. E., Franke, L., Fuyuno, Y., Hart, A., Juyal, R. C., Juyal, G., ... Weersma, R. K. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and shared genetic risk across populations. *Nature Genetics*, 47(9), 979–986.

Liu, Z., Kuo, P. L., Horvath, S., Crimmins, E., Ferrucci, L., & Levine, M. (2018). A new aging measure captures morbidity and mortality risk across diverse subpopulations from NHANES IV: A cohort study. *PLoS Medicine*, 15(12), e1002718. https://doi.org/10.1371/journal.pmed.1002718

Loh, P.-R., Bhatia, G., Gusev, A., Finucane, H. K., Bulik-Sullivan, B. K., Pollack, S. J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, de Candia, T. R., Lee, S. H., Wray, N. R., Kendler, K. S., O'Donovan, M. C., Neale, B. M., Patterson, N., & Price, A. L. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature Genetics*, 47(12), 1385–1392.

Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., Payne, A. J., Steinthorsdottir, V., Scott, R. A., Grarup, N., Cook, J. P., Schmidt, E. M., Wuttke, M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., ... McCarthy, M. I. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics*, 50, 1505–1513.

Malik, R., Rannikmäe, K., Traylor, M., Georgakis, M. K., Sargurupremraj, M., Markus, H. S., Hopewell, J. C., Debette, S., Sudlow, C. L. M., & Dichgans, M. (2018). Genome-wide meta-analysis identifies 3 novel loci associated with stroke. *Annals of Neurology*, 84, 934–939.

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., Bolla, M. K., Wang, Q., Tyrer, J., Dicks, E., Lee, A., Wang, Z., Allen, J., Keeman, R., Eilber, U., ... Easton, D. F. (2017). Association analysis identifies 65 new breast cancer risk loci. *Nature*, 551(7678), 92–94.

Nolte, I. M., Munoz, M. L., Tragante, V., Amare, A. T., Jansen, R., Vaez, A., Von Der Heyde, B., Avery, C. L., Bis, J. C., Dierckx, B., Van Dongen, J., Gogarten, S. M., Goyette, P., Hernesniemi, J., Huikari, V., Hwang, S. J., Jaju, D., Kerr, K. F., Kluttig, A., ... De Geus, E. J. C. (2017). Genetic loci associated with heart rate variability and their effects on cardiac disease risk. *Nature Communications*, 8, 15805. https://doi.org/10.1038/ncomms15805

Pattaro, C., Teumer, A., Gorski, M., Chu, A. Y., Li, M., Mijatovic, V., Garnaas, M., Tin, A., Sorice, R., Li, Y., Taliun, D., Olden, M., Foster, M., Yang, Q., Chen, M. H., Pers, T. H., Johnson, A. D., Ko, Y. A., Fuchsberger, C., ... Fox, C. S. (2016). Genetic associations at 53 loci cell types and biological pathways relevant for kidney function. *Nature Communications*, 7, 10023.

Pilling, L. C., Kuo, C.-L., Sicinski, K., Tamosauskaite, J., Kuchel, G. A., Harries, L. W., Herd, P., Wallace, R., Ferrucci, L., & Melzer, D. (2017). Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany. NY)*, 9(12), 2504–2520.

Pulit, S. L., Stoneman, C., Morris, A. P., Wood, A. R., Glastonbury, C. A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., Yang, J., Jones, S., Beaumont, R., Croteau-Chonka, D. C., Winkler, T. W., Hattersley, A. T., Loos, R. J. F., Hirschhorn, J. N., Visscher, P. M., ... Lindgren, C. M. (2019). Meta-Analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Human Molecular Genetics*, 28(1), 166–174.

Schumacher, F. R., Al Olama, A. A., Berndt, S. I., Benlloch, S., Ahmed, M., Saunders, E. J., Dadaev, T., Leongamornlert, D., Anokian, E., Cieza-Borrella, C., Goh, C., Brook, M. N., Sheng, X., Fachal, L., Dennis, J., Tyrer, J., Muir, K., Lophatananon, A., Stevens, V. L., ... Eeles, R. A. (2018). Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nature Genetics*, 50(7), 928–936.

Sonnega, A., Faul, J. D., Ofstedal, M. B., Langa, K. M., Phillips, J. W. R., & Weir, D. R. (2014). Cohort profile: The Health and Retirement Study (HRS). *International Journal of Epidemiology*, 43(2), 576–585.

Soranzo, N., Sanna, S., Wheeler, E., Gieger, C., Radke, D., Dupuis, J., Bouatia-Naji, N., Langenberg, C., Prokopenko, I., Stolerman, E., Sandhu, M. S., Heeney, M. M., Devaney, J. M., Reilly, M. P., Ricketts, S. L., Stewart, A. F. R., Voight, B. F., Willenborg, C., Wright, B., ... Meigs, J. B. (2010). Common variants at 10 genomic loci influence hemoglobin A1C levels via glycemic and nonglycemic pathways. *Diabetes*, 59(12), 3229–3239.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12, e1001779.

Tao, Q., Ang, T. F. A., DeCarli, C., Auerbach, S. H., Devine, S., Stein, T. D., Zhang, X., Massaro, J., Au, R., & Qiu, W. Q. (2018). Association of chronic low-grade inflammation with risk of Alzheimer disease in ApoE4 carriers. *JAMA Network Open*, 1, e183597.

Thompson, W. D., Tyrrell, J., Borges, M.-C., Beaumont, R. N., Knight, B. A., Wood, A. R., Ring, S. M., Hattersley, A. T., Freathy, R. M., & Lawlor, D. A. (2019). Association of maternal circulating 25(OH)D and calcium with birth weight: A mendelian randomisation analysis. *PLoS Medicine*, 16(6), e1002828. https://doi.org/10.1371/journal.pmed.1002828

Timmers, P. R., Mounier, N., Lall, K., Fischer, K., Ning, Z., Feng, X., Bretherick, A. D., Clark, D. W., Agbessi, M., Ahsan, H., Alves, I., Andiappan, A., Awadalla, P., Battle, A., Bonder, M. J., Boomsma, D., Christiansen, M., Claringbould, A., Deelen, P., ... Joshi, P. K. (2019) Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *Elife*, 8, e39856. https://doi.org/10.7554/eLife.39856.001

Watanabe, K., Taskesen, E., van Bochoven, A., Posthuma, D., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature Communications*, 8(1), 1826.

Wheeler, E., Leong, A., Liu, C. T., Hivert, M. F., Strawbridge, R. J., Podmore, C., Li, M., Yao, J., Sim, X., Hong, J., Chu, A. Y., Zhang, W., Wang, X., Chen, P., Maruthur, N. M., Porneala, B. C., Sharp, S. J., Jia, Y., Kabagambe, E. K., ... Meigs, J. B. (2017). Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Medicine*, 14(9), e1002383. https://doi.org/10.1371/journal.pmed.1002383

Williams, D. M., Jylhävä, J., Pedersen, N. L., & Hägg, S. (2019). A frailty index for UK biobank participants. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences.*, 74(4), 582–587.

Yang, J., Ferreira, T., Morris, A. P., & Medland, S. E. Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Madden, P. A. F., Heath, A. C., Martin, N. G., Montgomery, G. W., Weedon, M. N., Loos, R. J., Frayling, T. M., McCarthy, M. I., Hirschhorn, J. N., Goddard, M. E., & Visscher, P. M. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics*, 44(4), 369–375 S1-3.

Zengini, E., Hatzikotoulas, K., Tachmazidou, I., Steinberg, J., Hartwig, F. P., Southam, L., Hackinger, S., Boer, C. G., Styrkarsdottir, U., Gilly, A., Suveges, D., Killian, B., Ingvarsson, T., Jonsson, H., Babis, G. C., McCaskie, A., Uitterlinden, A. G., Van Meurs, J. B. J., Thorsteinsdottir, U., ... Zeggini, E. (2018). Genome-wide analyses

using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nature Genetics*, *50*(4), 549–558.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.