

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** McCartney, Daniel L.; Min, Josine L.; Richmond, Rebecca C.; Lu, Ake T.; Sobczyk, Maria K.; Davies, Gail; Broer, Linda; Guo, Xiuqing; Jeong, Ayoung; Jung, Jeesun; Kasela, Silva; Katrinli, Seyma; Kuo, Pei-Lun; Matias-Garcia, Pamela R.; Mishra, Pashupati P.; Nygaard, Marianne; Palviainen, Teemu; Patki, Amit; Raffield, Laura M.; Ratliff, Scott M.; Richardson, Tom G.; Robinson, Oliver; Soerensen, Mette;

**Title:** Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging

**Year:** 2021

**Version:** Published version

**Copyright:** © The Author(s). 2021

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

McCartney, D. L., Min, J. L., Richmond, R. C., Lu, A. T., Sobczyk, M. K., Davies, G., Broer, L., Guo, X., Jeong, A., Jung, J., Kasela, S., Katrinli, S., Kuo, P.-L., Matias-Garcia, P. R., Mishra, P. P., Nygaard, M., Palviainen, T., Patki, A., Raffield, L. M., . . . Marioni, R. E. (2021). Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. *Genome Biology*, 22, Article 194. <https://doi.org/10.1186/s13059-021-02398-9>

RESEARCH

Open Access



# Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging

Daniel L. McCartney<sup>1†</sup>, Josine L. Min<sup>2,3†</sup>, Rebecca C. Richmond<sup>2,3†</sup>, Ake T. Lu<sup>4†</sup>, Maria K. Sobczyk<sup>2,3†</sup>, Gail Davies<sup>5†</sup>, Linda Broer<sup>6†</sup>, Xiuqing Guo<sup>7†</sup>, Ayoung Jeong<sup>8,9†</sup>, Jeusun Jung<sup>10†</sup>, Silva Kasela<sup>11†</sup>, Seyma Katrinli<sup>12†</sup>, Pei-Lun Kuo<sup>13†</sup>, Pamela R. Matias-Garcia<sup>14,15,16†</sup>, Pashupati P. Mishra<sup>17†</sup>, Marianne Nygaard<sup>18,19†</sup>, Teemu Palviainen<sup>20†</sup>, Amit Patki<sup>21†</sup>, Laura M. Raffield<sup>22†</sup>, Scott M. Ratliff<sup>23†</sup>, Tom G. Richardson<sup>2,3†</sup>, Oliver Robinson<sup>24†</sup>, Mette Soerensen<sup>18,19,25†</sup>, Dianjianyi Sun<sup>26†</sup>, Pei-Chien Tsai<sup>27,28,29†</sup>, Matthijs D. van der Zee<sup>30,31†</sup>, Rosie M. Walker<sup>1†</sup>, Xiaochuan Wang<sup>32†</sup>, Yunzhang Wang<sup>33†</sup>, Rui Xia<sup>34†</sup>, Zongli Xu<sup>35†</sup>, Jie Yao<sup>36†</sup>, Wei Zhao<sup>23†</sup>, Adolfo Correa<sup>37</sup>, Eric Boerwinkle<sup>38</sup>, Pierre-Antoine Dugue<sup>32,39,40</sup>, Peter Durda<sup>41</sup>, Hannah R. Elliott<sup>2,3</sup>, Christian Gieger<sup>14,15</sup>, The Genetics of DNA Methylation Consortium<sup>42</sup>, Eco J. C. de Geus<sup>30,31</sup>, Sarah E. Harris<sup>5</sup>, Gibran Hemani<sup>2,3</sup>, Medea Imboden<sup>8,9</sup>, Mika Kähönen<sup>43</sup>, Sharon L. R. Kardia<sup>23</sup>, Jacob K. Kresovich<sup>35</sup>, Shengxu Li<sup>44</sup>, Kathryn L. Lunetta<sup>45</sup>, Massimo Mangino<sup>27,46</sup>, Dan Mason<sup>47</sup>, Andrew M. McIntosh<sup>48</sup>, Jonas Mengel-From<sup>18,19</sup>, Ann Zenobia Moore<sup>13</sup>, Joanne M. Murabito<sup>49</sup>, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium<sup>50</sup>, Miina Ollikainen<sup>20</sup>, James S. Pankow<sup>51</sup>, Nancy L. Pedersen<sup>33</sup>, Annette Peters<sup>15,52</sup>, Silvia Polidoro<sup>24</sup>, David J. Porteous<sup>1</sup>, Olli Raitakari<sup>53,54,55</sup>, Stephen S. Rich<sup>56</sup>, Dale P. Sandler<sup>35</sup>, Elina Sillanpää<sup>20,57</sup>, Alicia K. Smith<sup>12,58</sup>, Melissa C. Southey<sup>32,39,40</sup>, Konstantin Strauch<sup>59,60,61</sup>, Hemant Tiwari<sup>21</sup>, Toshiko Tanaka<sup>13</sup>, Therese Tillin<sup>62</sup>, Andre G. Uitterlinden<sup>6,63</sup>, David J. Van Den Berg<sup>64</sup>, Jenny van Dongen<sup>30,31</sup>, James G. Wilson<sup>65,66</sup>, John Wright<sup>47</sup>, Idil Yet<sup>27,67</sup>, Donna Arnett<sup>68</sup>, Stefania Bandinelli<sup>69</sup>, Jordana T. Bell<sup>27</sup>, Alexandra M. Binder<sup>70,71</sup>, Dorret I. Boomsma<sup>30,31</sup>, Wei Chen<sup>72</sup>, Kaare Christensen<sup>18,19,25</sup>, Karen N. Conneely<sup>73</sup>, Paul Elliott<sup>24</sup>, Luigi Ferrucci<sup>13</sup>, Myriam Fornage<sup>34</sup>, Sara Hägg<sup>33</sup>, Caroline Hayward<sup>74</sup>, Marguerite Irvin<sup>75</sup>, Jaakko Kaprio<sup>20,76</sup>, Deborah A. Lawlor<sup>2,3,77</sup>, Terho Lehtimäki<sup>17</sup>, Falk W. Lohoff<sup>10</sup>, Lili Milani<sup>11</sup>, Roger L. Milne<sup>32,39,40</sup>, Nicole Probst-Hensch<sup>8,9</sup>, Alex P. Reiner<sup>78</sup>, Beate Ritz<sup>70</sup>, Jerome I. Rotter<sup>7</sup>, Jennifer A. Smith<sup>23</sup>, Jack A. Taylor<sup>35</sup>, Joyce B. J. van Meurs<sup>6,63</sup>, Paolo Vineis<sup>24</sup>, Melanie Waldenberger<sup>14,15,52</sup>, Ian J. Deary<sup>5</sup>, Caroline L. Relton<sup>2,3</sup>, Steve Horvath<sup>4,79\*†</sup> and Riccardo E. Marioni<sup>1\*†</sup> 

\* Correspondence: [shorvath@mednet.ucla.edu](mailto:shorvath@mednet.ucla.edu); [riccardo.marioni@ed.ac.uk](mailto:riccardo.marioni@ed.ac.uk)

<sup>†</sup>Daniel L. McCartney, Josine L. Min, Rebecca C. Richmond, Ake T. Lu, Maria K. Sobczyk, Gail Davies, Linda Broer, Xiuqing Guo, Ayoung Jeong, Jeusun Jung, Silva Kasela, Seyma Katrinli, Pei-Lun Kuo, Pamela R. Matias-Garcia, Pashupati P. Mishra, Marianne Nygaard, Teemu Palviainen, Amit Patki, Laura M. Raffield, Scott M. Ratliff, Tom G. Richardson, Oliver Robinson, Mette

Soerensen, Dianjianyi Sun, Pei-Chien Tsai, Matthijs D. van der Zee, Rosie M. Walker, Xiaochuan Wang, Yunzhang Wang, Rui Xia, Zongli Xu, Jie Yao, Wei Zhao, Steve Horvath and Riccardo E. Marioni contributed equally to this work.

<sup>4</sup>Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA

<sup>1</sup>Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Crewe Road South, Edinburgh EH4 2XU, UK

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Abstract

**Background:** Biological aging estimators derived from DNA methylation data are heritable and correlate with morbidity and mortality. Consequently, identification of genetic and environmental contributors to the variation in these measures in populations has become a major goal in the field.

**Results:** Leveraging DNA methylation and SNP data from more than 40,000 individuals, we identify 137 genome-wide significant loci, of which 113 are novel, from genome-wide association study (GWAS) meta-analyses of four epigenetic clocks and epigenetic surrogate markers for granulocyte proportions and plasminogen activator inhibitor 1 levels, respectively. We find evidence for shared genetic loci associated with the Horvath clock and expression of transcripts encoding genes linked to lipid metabolism and immune function. Notably, these loci are independent of those reported to regulate DNA methylation levels at constituent clock CpGs. A polygenic score for GrimAge acceleration showed strong associations with adiposity-related traits, educational attainment, parental longevity, and C-reactive protein levels.

**Conclusion:** This study illuminates the genetic architecture underlying epigenetic aging and its shared genetic contributions with lifestyle factors and longevity.

**Keywords:** DNA methylation, GWAS, Epigenetic clock

## Background

Aging is associated with an increased risk of physical, cognitive, and degenerative disorders [1]. While the rate of chronological aging is constant between individuals, there are inter-individual differences in the risk of age-associated morbidities. Biological aging is influenced by both environmental and genetic factors [2]. Multiple measures of biological age exist, several of which have drawn information from DNA methylation (DNAm) across the genome. DNAm is a common epigenetic modification typically characterized by the addition of a methyl group to a cytosine-guanine dinucleotide (CpG). DNAm levels can be influenced by both genetic and environmental factors, and in recent years, DNAm signatures have become established correlates of multiple health-related outcomes [3–5]. Such signatures include “epigenetic clocks”, accurate markers of aging which associate with several health outcomes [6, 7]. Epigenetic clocks use weighted linear combinations of CpGs to predict an individual’s chronological age and have common single-nucleotide polymorphism (SNP)-based heritability estimates ranging from 0.15 to 0.19 [8, 9]. Individuals with epigenetic clock estimates greater than their chronological age display “age acceleration” and have been shown to be at a greater risk of all-cause mortality and multiple adverse health outcomes [10]. Consequently, identification of genetic and environmental contributors to the variation in these measures in populations has become a major goal in the field [11].

The first generation of epigenetic aging clocks used penalized regression models to predict chronological age on the basis of DNA methylation data, e.g., the widely used clocks from Hannum (2013) and Horvath (2013) apply to blood and 51 human tissues/cell types, respectively [12–14]. A derivative of the Horvath clock, intrinsic epigenetic age acceleration (IEAA) has since been developed, conditioning out (i.e., removing) estimates of blood cell composition. An increasing literature supports the view that IEAA relates to properties of hematopoietic stem cells [2, 8, 15]. The second generation of epigenetic clocks move beyond estimating chronological age by incorporating

information on morbidity and mortality risk (e.g., smoking, plasma protein levels, white blood cell counts), and chronological age. Two such predictors, termed PhenoAge (a DNAm predictor trained on a measure that itself was trained on mortality, using 42 clinical measures and age as input features) and GrimAge (trained on mortality, including a DNAm measure of smoking as a constituent part), outperform both Hannum and Horvath clocks in predicting mortality and are associated with various measures of morbidity and lifestyle factors [16, 17]. DNAm GrimAge outperforms PhenoAge and the first generation of epigenetic clocks when it comes to predicting time to death [8, 18, 19].

While nothing is known about the genetics of the second generation of epigenetic clocks, 13 genetic loci have been associated with the first generation of epigenetic clocks. A study of nearly ten thousand individuals revealed a regulatory relationship between human telomerase (hTERT) and epigenetic age acceleration [8]. More recently, a larger genome-wide association study (GWAS;  $n = 13,493$ ) revealed that metabolic and immune pathways share genetic underpinnings with epigenetic clocks [9].

Here, we greatly expand on these studies across several dimensions. First, we analyze a large, multi-ethnic dataset comprised of over 41,000 individuals from 29 European ancestry studies, seven African American studies, and one Hispanic ancestry study. Second, we characterize for the first time the genetic architecture of the second-generation epigenetic clocks, GrimAge and PhenoAge. All of these clocks have been trained on European ancestry populations. Third, we also conduct GWAS of two important DNAm-based surrogate markers: DNAm plasminogen activator inhibitor-1 (PAI1) levels and granulocyte proportion, respectively. Although not considered in risk prediction scores such as the Framingham Heart score, DNAm PAI1 was chosen because it exhibited stronger associations with cardiometabolic disease than the epigenetic clocks [16]. The DNAm-based estimate of granulocyte proportions was chosen because it exhibited significant associations with several epigenetic clocks (including GrimAge and PhenoAge) and with health outcomes such as Parkinson's disease [16, 17, 20]. The unprecedented sample size of the current study allowed us to develop polygenic risk scores for these six epigenetic biomarkers.

We report 137 independent loci, including 113 novel loci (i.e., not previously identified in previous GWAS meta-analyses of epigenetic age estimators [8, 9]), and examine the genetic and causal relationships between epigenetic aging, lifestyle behaviors, health outcomes, and longevity.

## Results

To identify genetic variants associated with six methylation-based biomarkers, genome-wide association studies of 34,710 European ancestry and 6195 African American individuals were performed (Additional file 1 and Additional file 2: Tables S1-S2). A fixed effects meta-analysis was performed to combine the summary statistics within each ancestry group (summary statistics available at <https://datashare.is.ed.ac.uk/handle/10283/3645>). Genomic inflation factors ranged between 1.01 and 1.06 (Additional file 3: Figure S1-S6) for the European-only meta-analyses, indicating appropriate adjustment for population stratification, and from 1.11 and 1.21 for the meta-analyses comprising African American participants (Additional file 3: Figures S7-S12; Table 1). Inflation was present and consistent across all allele frequencies in the African American analyses; there was much greater variability in the effect sizes in the African American analyses

**Table 1** Summary of key findings of post GWA analyses. African American meta-analyses based on 6195 participants from 7 cohorts; European ancestry meta-analyses based on 34,710 participants from 28 cohorts

|  | IEAA             | Hannum           | PhenoAge         | GrimAge         | Granulocyte      | PAI-1           |
|--|------------------|------------------|------------------|-----------------|------------------|-----------------|
| GWAS loci <sup>a</sup>                                     | 24 EUR<br>9 AFR  | 9 EUR<br>11 AFR  | 12 EUR<br>12 AFR | 4 EUR<br>11 AFR | 2 EUR<br>27 AFR  | 5 EUR<br>11 AFR |
| Gene-based GWAS <sup>b</sup>                               | 40 EUR<br>13 AFR | 29 EUR<br>13 AFR | 46 EUR<br>18 AFR | 5 EUR<br>14 AFR | 7 EUR<br>209 AFR | 2 EUR<br>9 AFR  |
| Functional enrichment (EUR) <sup>c</sup>                   | 193              | 3                | 1                | 0               | 0                | 0               |
| Pathway analyses <sup>d</sup>                              | 0                | 0                | 0                | 0               | 35 AFR<br>0 EUR  | 0               |
| GWAS Catalog Enrichment (EUR) <sup>e</sup>                 | 5                | 0                | 0                | 6               | 21               | 2               |
| Colocalisation (EUR)                                       | 7                | 0                | 0                | 1               | 0                | 0               |
| Heritability (EUR)   | 0.17             | 0.11             | 0.10             | 0.10            | 0.08             | 0.02            |
| Genetic correlations (EUR)                                 | 0                | 0                | 36               | 107             | 0                | 0               |
| Polygenic Risk Score Max R <sup>2</sup> (EUR) <sup>f</sup> | 2.37%            | 1.13%            | 1.38%            | 1.07%           | 0.76%            | 1.62%           |
| Mendelian Randomisation (EUR) <sup>g</sup>                 | 0                | 3                | 6                | 11              | 13               | 1               |

Analyses labelled EUR were performed only on European ancestry GWAS results due to availability of reference data  
a: Independent loci identified by cojo at  $p < 5 \times 10^{-8}$   
b: Genes identified as significant using a MAGMA p-value threshold of  $p < 2.69 \times 10^{-6}$   
c: Enrichments identified using GARFIELD and MendelVar  
d: Enrichments identified among KEGG pathways/GO terms  
e: Number of GWAS traits enriched for genes identified in the current meta-analysis  
f: Maximum average percentage of variance explained by polygenic scores at a range of p-value thresholds  
g: Mendelian randomisation analysis of epigenetic predictors (outcomes) and published GWAS traits (exposures) significant using the IVW estimate

(Additional file 4). Phenotypic correlations were examined in Generation Scotland, the largest participating cohort in the study. Correlations ranged from 0 (IEAA and granulocyte proportions) to 0.48 (PhenoAge acceleration and Hannum age acceleration (Additional file 2: Table S3). We examined the relationship between means and standard deviations of predicted age versus means and standard deviations of chronological age for each cohort, separated by ancestry group, observing weaker mean correlations in the African American cohorts. There was little difference in the relationship between the standard deviations of age acceleration and chronological age by ancestry group (Additional file 3: Figure S13). Heterogeneity between studies may decrease power to detect genetic associations. We found little evidence of systematic between-study heterogeneity in both the European ancestry and African American meta-analyses, as determined by M-statistic outlier analysis, meta-regressions against cohort characteristics, and analysis of heterogeneity  $I^2$  statistics [21] (Additional file 4).

The key findings of the post GWA analyses are summarized in Table 1 along with a summary of the input features of each clock in Table 2. The latter highlights the value in discriminating novel associations from those that are likely driven by the construction of the clocks.

#### European ancestry GWAS meta-analysis: 56 independently associated loci

We identified 56 conditionally independent associations ( $P < 5 \times 10^{-8}$ ) across the six epigenetic biomarkers in European ancestry populations using a stepwise model (Additional file 3: Figures S14–S19; Table 1; Additional file 2: Table S4) [22, 23]. We replicated 10/10 loci associated with IEAA and 1/1 locus associated with a cell-adjusted

**Table 2** Summary of input features of first-generation (Hannum and Horvath) and second-generation (DNAm PhenoAge, GrimAge) epigenetic clocks

| Clock          | Training Outcome   | Input  |
|----------------|--|--|
| Hannum         | Chronological Age  | 71 CpGs from the Illumina 450k array, filtered from all sites via elastic net regression   |
| Horvath (IEAA) | Chronological Age  | 353 CpGs used to generate Horvath Age, then regressed on age and cell count information imputed from the DNAm data (naïve CD8+ T cells, exhausted CD8+ T cells, plasmablasts, CD4+ T cells, natural killer cells, monocytes, and granulocytes) |
| DNAm PhenoAge  | “Phenotypic” Age (PhenoAge) – selection of 9 biomarkers and chronological age used to predict ageing-related mortality via penalised Cox regression. | 513 CpG sites used to predict PhenoAge via elastic net regression  |
| GrimAge        | All-cause mortality  | Age, sex, and 1,030 CpGs from DNAm proxies for smoking, leptin, adrenomedullin, beta-2-microglobulin, cystatin C, growth differentiation factor 15, plasminogen activation inhibitor 1, tissue inhibitor metalloproteinase-1.                  |

Hannum-based measure of epigenetic age acceleration identified in an earlier GWAS ( $P < 0.05/11 = 0.0045$ ) [9]. All but three loci (associated with IEAA) were replicated at the genome-wide significant level. To validate the associations with DNAm-derived granulocyte counts, we compared our results to a previous GWAS of FACS granulocyte counts which identified 155 independent loci [24]. In the current meta-analysis, we replicated 13/129 present loci ( $P < 0.05/129 = 3.88 \times 10^{-4}$ ; Additional file 2: Table S5; Additional file 3: Figures S20-S21), two of which replicated at the genome-wide significant level. Effect sizes at the 129 loci were strongly correlated between studies ( $r = 0.85$ ). Conversely, we failed to replicate four genome-wide significant lead SNPs from a previous GWAS of measured PAI1 levels ( $P \geq 0.327$ ) [25]. There was no clear concordance of effect sizes at these loci in the current study ( $r = -0.59$ ;  $P = 0.41$ )

To examine whether genetic variation across the six epigenetic biomarkers was shared, we performed genetic colocalization analyses of the 56 loci [26]. There was evidence for colocalization at 30 loci between epigenetic biomarkers (posterior probability (PP)  $> 0.8$ ; Additional file 2: Table S6). IEAA was associated with the greatest number of independent loci ( $n = 24$ ), whereas granulocyte proportion was associated with the fewest ( $n = 2$ ).

#### African American GWAS meta-analysis: 81 independently associated loci

We identified 81 conditionally independent associations for the six epigenetic biomarkers in the African American analyses (Additional file 3: Figures S22-S27; Table 1; Additional file 2: Table S4). The number of associated loci per epigenetic biomarker ranged from 9 (IEAA) to 27 (granulocyte proportion).

#### Trans-ethnic meta-analyses identify 69 loci

To determine if any loci were shared across the European and African American populations, a trans-ethnic meta-analysis was carried out for each of the six epigenetic biomarkers using MR-MEGA [27]. Sixty-nine risk loci were identified across the six predictors that were common to all ancestries, ranging from 6 (GrimAge acceleration)



to 23 (IEAA). Ten loci were significant in the African American analyses, 33 were significant in the European analyses, and five were significant in both. This left 21 novel loci from the trans-ethnic meta-analyses (Additional file 2: Table S7). Among the allele frequencies of the lead SNPs for these loci, 11/21 differed by > 10% between European and African American populations (Additional file 2: Table S8).

We compared effect sizes of the lead SNPs from these loci in a Hispanic-American ancestry subset of the MESA cohort ( $n = 287$ ). Correlations between the respective effect sizes ranged from very weak ( $r = 0.16$  for 10 granulocyte proportion SNPs) to near unit ( $r = 0.92$  for 10 Hannum age acceleration SNPs; Additional file 2: Table S9 and Additional file 3: Figure S28).

### Gene-based GWAS identifies 364 significant genes

Gene-based GWASs carried out using MAGMA identified between two and 46 genes (111 unique genes in total) associated with the six epigenetic biomarkers in the European ancestry data (Additional file 2: Table S10) [28]. In the African American data, between nine and 209 genes (264 unique genes in total) were associated with the epigenetic biomarkers (Additional file 2: Table S10). Across all epigenetic biomarkers and ancestries, there were 364 unique gene-based associations.

### Independently associated loci are associated with DNA methylation levels

One obvious genetic effect that may influence our GWA findings is the overlap with *cis* methylation quantitative trait loci (mQTLs) for epigenetic clock DNAm sites. To explore whether any of the 56 loci from the European GWAS shared genetic variation influencing epigenetic clock DNAm sites, colocalization analyses were conducted using GoDMC summary statistics [29] (“Methods”). We found strong evidence ( $PP > 0.8$ ) that 1/4 loci (25%) for GrimAge acceleration, 3/12 loci (25%) for PhenoAge acceleration, 11/24 loci (46%) for IEAA, and 5/9 loci (56%) for Hannum age acceleration had shared genetic variation influencing epigenetic clock DNAm sites (Additional file 2: Tables S11–S12). Next, we used genetic colocalization to assess whether GWAS loci for aging biomarkers were associated with methylation levels at established DNAm sites for BMI [30] and smoking [31]; 29/56 loci (52%) colocalized with genetic variation influencing smoking-associated DNAm sites and 1/56 loci (1.8%) was colocalized with genetic variation influencing a BMI-associated DNAm site. Specifically, GrimAge acceleration (75%) and Hannum age acceleration (78%) loci showed a large overlap of genetic variation influencing smoking DNAm sites.

Utilizing results from a published GWAS of IEAA in brain tissue [32], we tested whether genetic variation influencing IEAA in blood and brain was shared for the 24 blood-related IEAA loci (cross-tissue plot for lead SNPs shown in Additional file 3: Figure S29). Colocalization analysis showed that there was no strong evidence ( $PP > 0.80$ ) for a single SNP being associated with both traits. However, the true extent of sharing is difficult to estimate because the sample size of the brain study ( $n = 1796$ ) is much smaller than our blood-based study, limiting power to detect shared loci. Previous simulations using a sample size of 2000 individuals have indicated that the shared variant must explain close to 2% of the variance of a biomarker to attain a posterior probability > 0.8 for shared genetic effects [26]. Nevertheless, we observed suggestive evidence for

colocalization (PP = 0.53; LocusZoom plot in Additional file 3: Figure S30) for a locus mapping to *DSCR6* on chromosome 21. This locus also shares genetic variation with an IEAA clock CpG, cg13450409 (PP = 0.99; Additional file 2: Table S12).

### SNP- and gene-based enrichment within published GWAS

To determine whether any of the 56 lead SNPs in the European ancestry meta-analyses for the six epigenetic biomarkers showed evidence for pleiotropic associations, a lookup of published GWAS significant associations ( $P < 5 \times 10^{-8}$ ) was carried out (Additional file 2: Table S13). Four of the IEAA-associated SNPs (rs2736100 in *TERT*, rs2275558 in *PBX1*, rs144317085 in *TET2*, and rs2492286 in *RPNI*) were associated with 16 unique traits including multiple cancers (e.g., lung cancer, glioma) [33, 34] and blood cell counts (e.g., platelet count, eosinophil count, red blood cell count) [24, 35]. Whereas there was considerable overlap with cell-related traits (e.g., white/red blood cell, platelet, and eosinophil counts, and mean corpuscular volume) [24, 35, 36], there were no associations with non-cancer-related disease or lifestyle measures (Additional file 2: Table S13). A gene-based test of enrichment among traits within the GWAS catalog output (Additional file 2: Table S14) showed genes associated with IEAA, GrimAge, and granulocyte proportion were enriched among those associated with white blood cell counts [24, 35]. Several genes associated with granulocyte proportions were also enriched among those associated with inflammatory traits (e.g., inflammatory bowel disease, rheumatoid arthritis, asthma) [37–42]. IEAA-associated genes were also significantly enriched among those identified in a previous GWAS of IEAA [8, 9].

### Colocalization to identify GWAS loci that might regulate expression levels

Colocalization analyses were conducted to investigate whether any of the 56 loci from the European ancestry GWASs showed evidence of regulating gene expression levels (“Methods”). There was strong evidence (PP > 0.8) that eight loci had shared genetic effects with expression quantitative trait loci (eQTLs; Table 1; Additional file 2: Table S15). Of these, one was associated with GrimAge acceleration and seven were associated with IEAA. The locus associated with GrimAge acceleration was linked to the expression of *C6orf183* whereas IEAA-associated loci were linked to the expression of 11 transcripts including genes related to lipid transport and immune function (e.g., *ATP8B4*, *CD46*, *TRIM59*). Colocalization plots are presented in Additional file 5. Notably, four loci (all associated with IEAA) were independent of variants associated with DNAm levels at constituent clock CpGs (Additional file 2: Table S12; Table S15).

### Functional enrichment analysis

To gain an understanding of the regulatory properties of the variants that underlie the six epigenetic biomarkers, we performed functional enrichment analyses across various gene annotations and regulatory and cell-type specific elements on the summary statistics for each of the European ancestry GWAS results (see “Methods,” Additional file 2: Table S16) [43]. At an epigenetic biomarker-specific adjusted  $P$  value calculated from the effective number of annotations, significant enrichments were present for IEAA ( $n = 191$ ) and Hannum age acceleration ( $n = 3$ ). Associations with IEAA were enriched in DNaseI hypersensitive site (DHS) hotspots in several tissues, which might reflect that



Horvath's pan-tissue clock applies to all tissues. The strongest enrichment of associations with IEAA could be observed for mobilized CD34 primary cells ( $OR = 6.06$ ,  $P = 6.1 \times 10^{-12}$ ), which supports the view that IEAA reflects properties of hematopoietic stem cells [2, 8, 15].

### Pathway enrichment analysis

In the African American analysis, genes associated with granulocyte proportions were enriched among 35 Gene Ontology (GO) terms (Bonferroni  $P < 0.05$ ), the majority of which were immune-related (e.g., adaptive immune response, lymphocyte activation, regulation of immune system process, or skin development; Additional file 2: Table S17). By contrast, there was no significant enrichment among Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways or GO terms for the significantly associated genes in the European ancestry analysis.

### Overlap with Mendelian disease genes

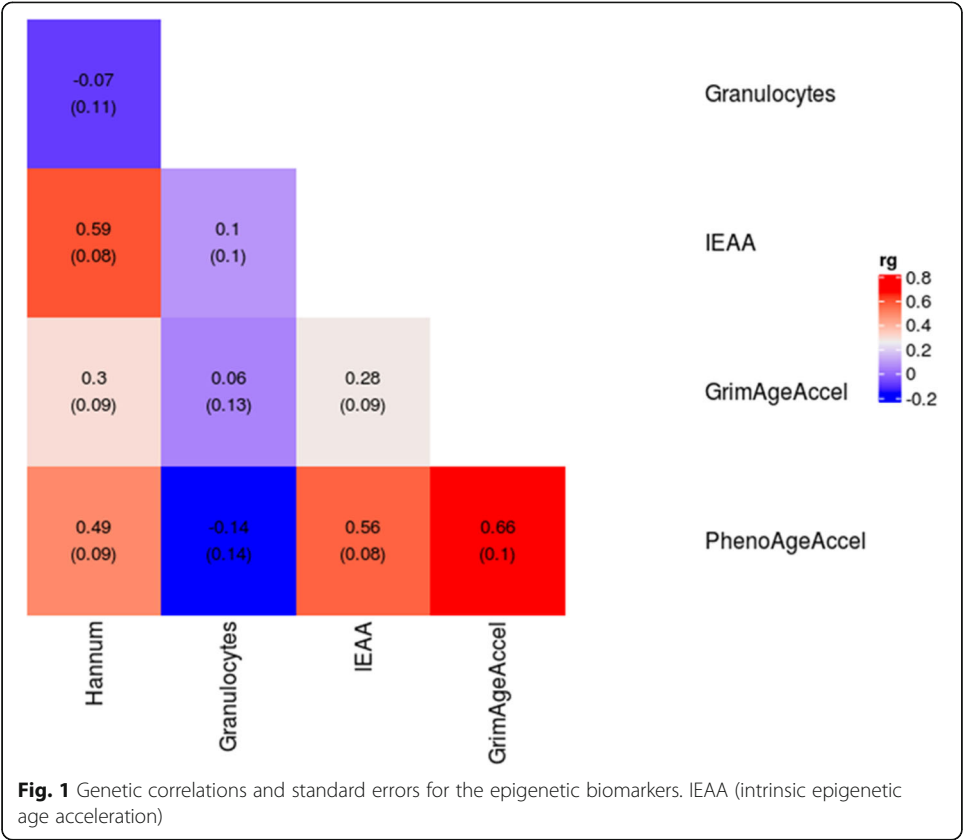
In order to characterize the potential overlap of Mendelian disease genes and associated pathways with our findings, a series of enrichment analyses were conducted ("Methods") [44]. In the European ancestry analysis, enrichment of Mendelian disease genes was observed for IEAA and two gene sets ("disorders of platelet function" and "vascular skin abnormality": bootstrapped  $P = 0.027$  and  $0.049$ , respectively). Enrichment analysis of PhenoAge loci revealed overrepresentation of methylation-related Mendelian disease genes (Additional file 2: Table S18; bootstrapped  $P = 0.04$ ) with *MTR* (methyltetrahydrofolate-homocysteine S-methyltransferase) present in addition to *TPMT*. No significant over-representations of Mendelian disease gene sets were observed for any of the genes identified in the African American analysis.

### Heritability and LD score regression

We quantified the proportion of variance in the six epigenetic biomarkers from the European ancestry meta-analyses that can be explained by our SNP sets using Linkage Disequilibrium (LD) Score regression [45]. The GWAS summary statistic SNP-based heritability ranged from 0.02 (SE = 0.02) for PAI1 levels, to 0.17 (SE = 0.02) for IEAA (Table 1; Additional file 3: Figure S31; Additional file 2: Table S19). We omitted DNAm PAI1 from the genetic correlation analysis due to its low heritability estimate. Several of the remaining five epigenetic biomarkers exhibited significant ( $P < 0.05$ ) pairwise genetic correlation coefficients ranging from  $r = 0.28$  (GrimAge acceleration and IEAA) to  $r = 0.66$  (GrimAge acceleration and PhenoAge acceleration; Fig. 1; Additional file 2: Table S20).

For each epigenetic biomarker, large-scale genetic correlation analyses were conducted with 693 different traits. A selection of significant associations after Bonferroni correction for multiple testing ( $P < 0.05/693 = 7.22 \times 10^{-5}$ ) is presented in Fig. 2.

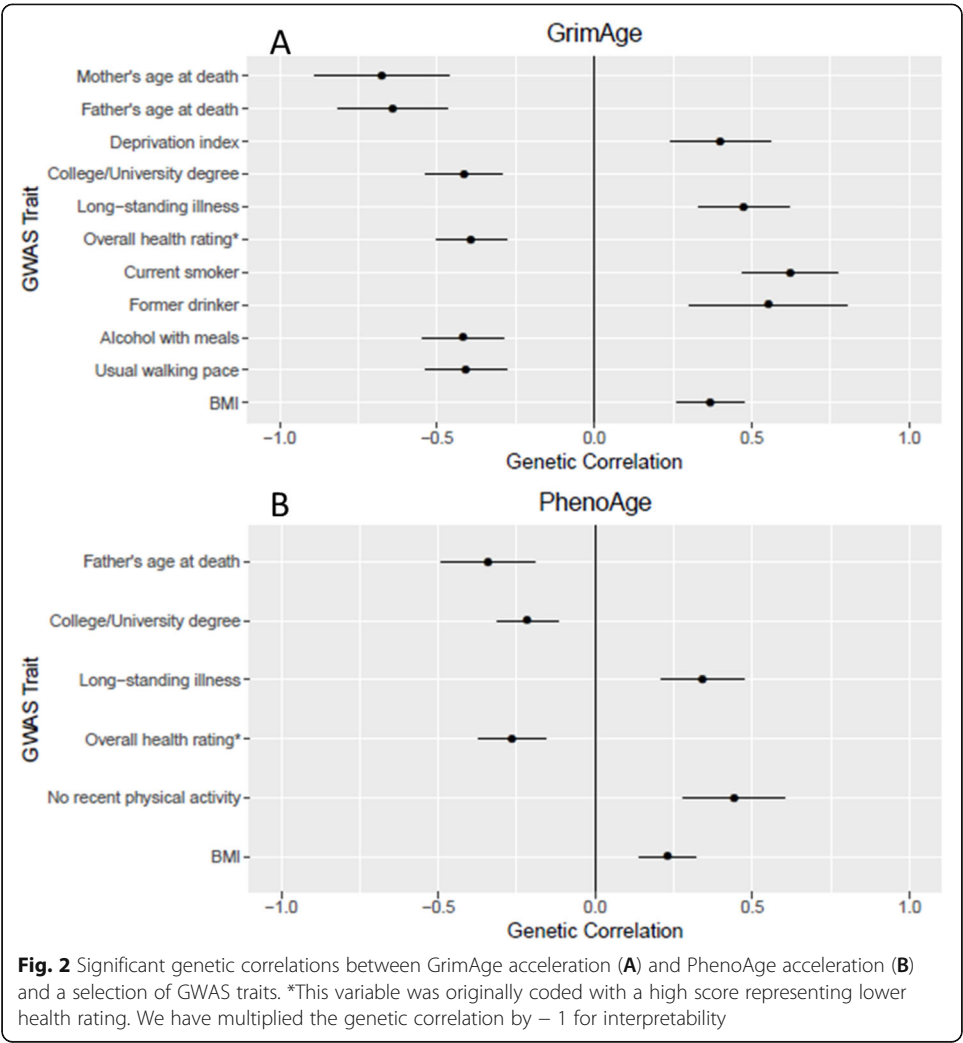
PhenoAge acceleration had significant genetic correlations with 36 health-related traits including educational and cognitive traits (e.g., years of schooling, intelligence;  $r_g = -0.26$  and  $-0.30$ ;  $P \leq 3.29 \times 10^{-5}$ ) [46–48], anthropometric traits (e.g., waist circumference, obesity, extreme BMI, hip circumference;  $r_g = 0.22$ – $0.31$ ;  $P \leq 3.61 \times 10^{-5}$ ) [49, 50], (<http://www.nealelab.is/uk-biobank/>), adiposity (e.g., leg, arm, and trunk fat mass;



$r_g = 0.22-0.23$ ;  $P \leq 3.46 \times 10^{-6}$ ) (<http://www.nealelab.is/uk-biobank/>), and longevity (e.g., father's age at death:  $r_g = -0.34$ ;  $P = 9.66 \times 10^{-6}$ ) (<http://www.nealelab.is/uk-biobank/>). GrimAge acceleration was genetically correlated with similar traits to PhenoAge (e.g., father's age at death:  $r_g = -0.64$ ;  $P = 6.2 \times 10^{-13}$ ), along with smoking-related traits (e.g., current tobacco smoking:  $r_g = 0.62$ ;  $P = 1.5 \times 10^{-15}$ ) (<http://www.nealelab.is/uk-biobank/>) and cancer-related traits (e.g., lung cancer:  $r_g = 0.48$ ;  $P = 8.3 \times 10^{-6}$ ) [51]. The shared genetic contributions to GrimAge and smoking/mortality are expected given that GrimAge uses a DNAm-based estimator of smoking pack-years in its definition. There were no significant genetic correlations between Hannum age acceleration, IEAA, or granulocyte proportions and any of the traits tested after correction for multiple testing (Additional file 2: Table S21).

**Polygenic risk score (PRS) profiling**

To determine how well SNP-based genetic scores can approximate the six epigenetic biomarkers and investigate whether these genetic scores associate with health outcomes, a polygenic risk score analysis was conducted on the European ancestry data. Re-running the meta-analysis with an iterative leave-one-cohort-out process (and on the full summary statistics in a completely independent cohort—the Young Finns Study), the mean polygenic predictions explained between 0.21 and 2.37% of the epigenetic biomarkers (Table 1; Fig. 3; Additional file 2: Table S22). The maximum prediction for a single cohort was 4.21% for PAI1 levels in ARIES. Parsimonious predictors



(built using SNPs with  $P < 5 \times 10^{-8}$ ) performed well for IEAA, PAI1 levels, and PhenoAge acceleration, whereas predictors including more SNPs ( $P < 0.01$ – $P < 1$ ) tended to explain the most variance in GrimAge acceleration and granulocyte proportions.

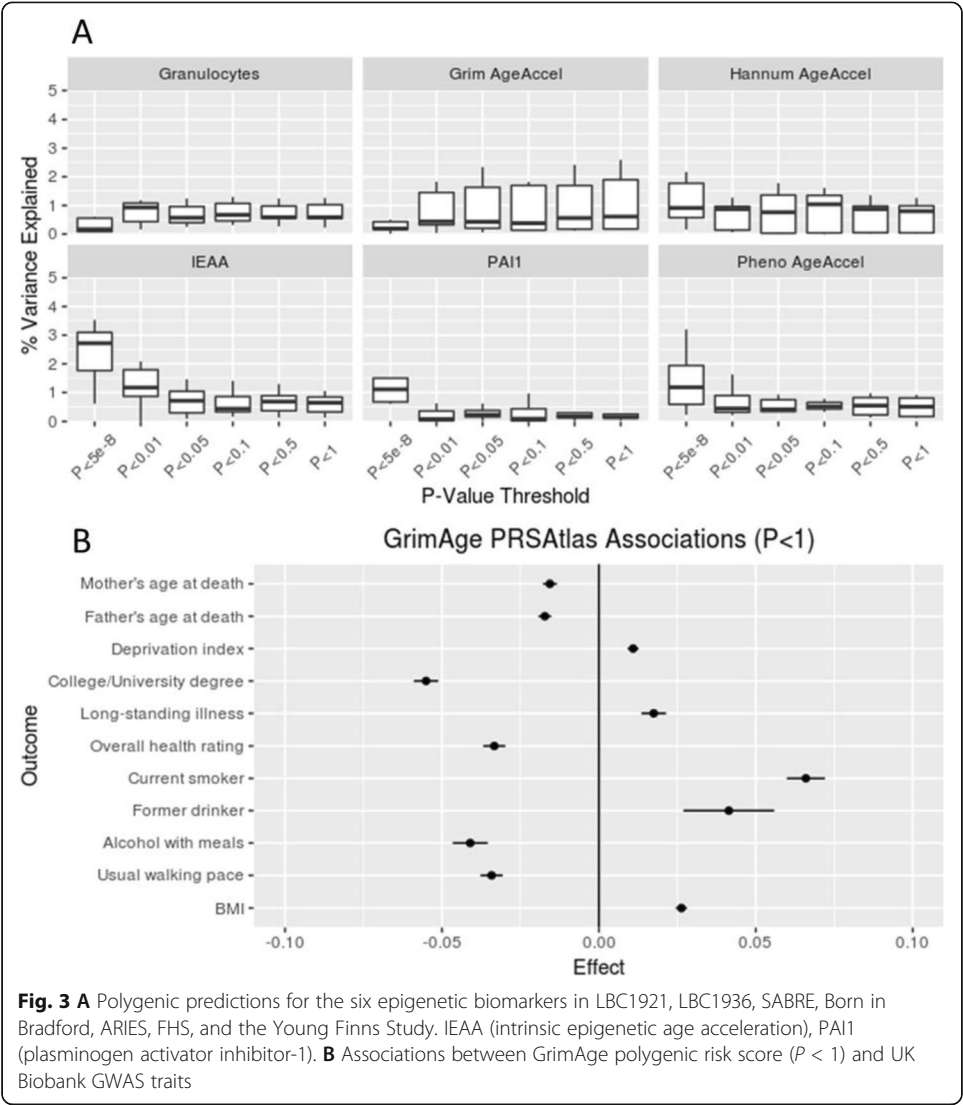
In order to investigate the association between the polygenic risk scores and health outcomes, we utilized a PRS Atlas to model associations with 581 heritable traits ( $n_{\text{range}} = 10,299$  to 334,915) from the UK Biobank study [52]. The PRS inputs included the independent SNPs with  $P < 1$  for GrimAge acceleration and granulocyte proportion and  $P < 5 \times 10^{-8}$  for the other four epigenetic biomarkers, with thresholds based on the results from the leave-one-out predictions (Fig. 3). Using a false discovery rate (FDR)-corrected  $P$  value for each of the six epigenetic biomarkers, we found between 7 and 250 significant associations for GrimAge acceleration, granulocyte proportions, and Hannum age acceleration ( $P_{\text{FDR}} < 0.05$ ; Additional file 2: Table S23). The strongest associations were between the GrimAge acceleration PRS and the following traits: adiposity-related traits (e.g., body fat percentage:  $\beta = 0.02$ ;  $P_{\text{FDR}} = 7.3 \times 10^{-39}$ ); education (e.g., college or university degree:  $\beta = -0.06$ ;  $P_{\text{FDR}} = 2.6 \times 10^{-43}$ ); and parental longevity (e.g., father's age at death:  $\beta = -0.02$ ;  $P_{\text{FDR}} = 5.7 \times 10^{-16}$ ; mother's age at death:  $\beta = -0.02$ ;  $P_{\text{FDR}} = 1.6 \times 10^{-11}$ ). Higher C-reactive protein was associated with a higher

PRS for both granulocyte proportions and GrimAge acceleration (granulocyte proportions:  $\beta = 0.01$ ;  $P_{\text{FDR}} = 8.2 \times 10^{-4}$ ; GrimAge acceleration:  $\beta = 0.02$ ;  $P_{\text{FDR}} = 2.1 \times 10^{-29}$ ), and a lower score for Hannum age acceleration ( $\beta = -0.006$ ;  $P_{\text{FDR}} = 0.02$ ). A higher Hannum age acceleration PRS was also associated with an increased likelihood of taking insulin medication and lower total protein levels.

### Mendelian randomization between age acceleration phenotypes and health and lifestyle outcomes

To investigate if the epigenetic measures were causally influenced by lifestyle factors and had a causal effect on aging and disease outcomes, we performed Mendelian randomization (MR) analyses on 150 traits for the European ancestry data (Additional file 2: Table S24). We found 12 inverse-variance weighted MR effects between the main exposures and epigenetic outcomes (GrimAge acceleration, PhenoAge acceleration, and PAI1 levels), after adjustment for multiple testing (Table 1; Additional file 2: Table S25). Of these, three remained significant ( $P < 0.05$ ) across the other three MR methods. All of these consistent effects were with GrimAge as the outcome. Greater adiposity was associated with greater GrimAge acceleration: body mass index (BMI;  $\text{Beta}_{\text{IVW}} = 0.76$  years per standard deviation (SD) increase in BMI,  $P = 3.7 \times 10^{-16}$ ); hip circumference ( $\text{Beta}_{\text{IVW}} = 0.42$  years per SD increase in hip circumference,  $P = 2.5 \times 10^{-5}$ ); waist circumference ( $\text{Beta}_{\text{IVW}} = 0.59$  years per SD increase in waist circumference,  $P = 5.9 \times 10^{-6}$ ; Fig. 4). Current tobacco smoking showed evidence for a causal effect on increased GrimAge in two of the MR methods ( $\text{Beta}_{\text{IVW}} = 3.42$  years for smokers,  $P = 9.0 \times 10^{-6}$ ; Fig. 4), as anticipated given that it incorporates a DNAm-based estimator of smoking pack-years [8]. Past tobacco smoking showed evidence for an inverse causal effect ( $\text{Beta}_{\text{IVW}} = -1.09$  years,  $P = 6.6 \times 10^{-9}$ ), indicating that GrimAge acceleration is reduced upon smoking cessation. As a DNAm-proxy for leptin was included in the derivation of GrimAge, the smoking and adiposity findings may act as positive controls. There was evidence from three of the four MR methods to support a link between higher educational attainment (both years of schooling and college/university degree) and lower GrimAge acceleration (Fig. 4). For the secondary exposures, there was evidence across all methods for a causal effect of a greater body size at age 10 on higher GrimAge acceleration ( $\text{Beta}_{\text{IVW}} = 0.70$ ,  $P = 1.6 \times 10^{-4}$ ; Additional file 2: Tables S24; S26). Consistent findings across all four MR methods provided evidence to support a causal effect of 13 cell count traits and DNA methylation-estimated granulocyte proportions, and between lower lymphocyte proportions and higher Hannum and GrimAge acceleration (Additional file 2: Table S27). There was evidence for heterogeneity in the causal effects for most of the cell types on epigenetic age measures, as well as years of schooling on GrimAge acceleration, although weaker evidence for directional pleiotropy was detected, based on the Egger intercept (Additional file 2: Table S28).

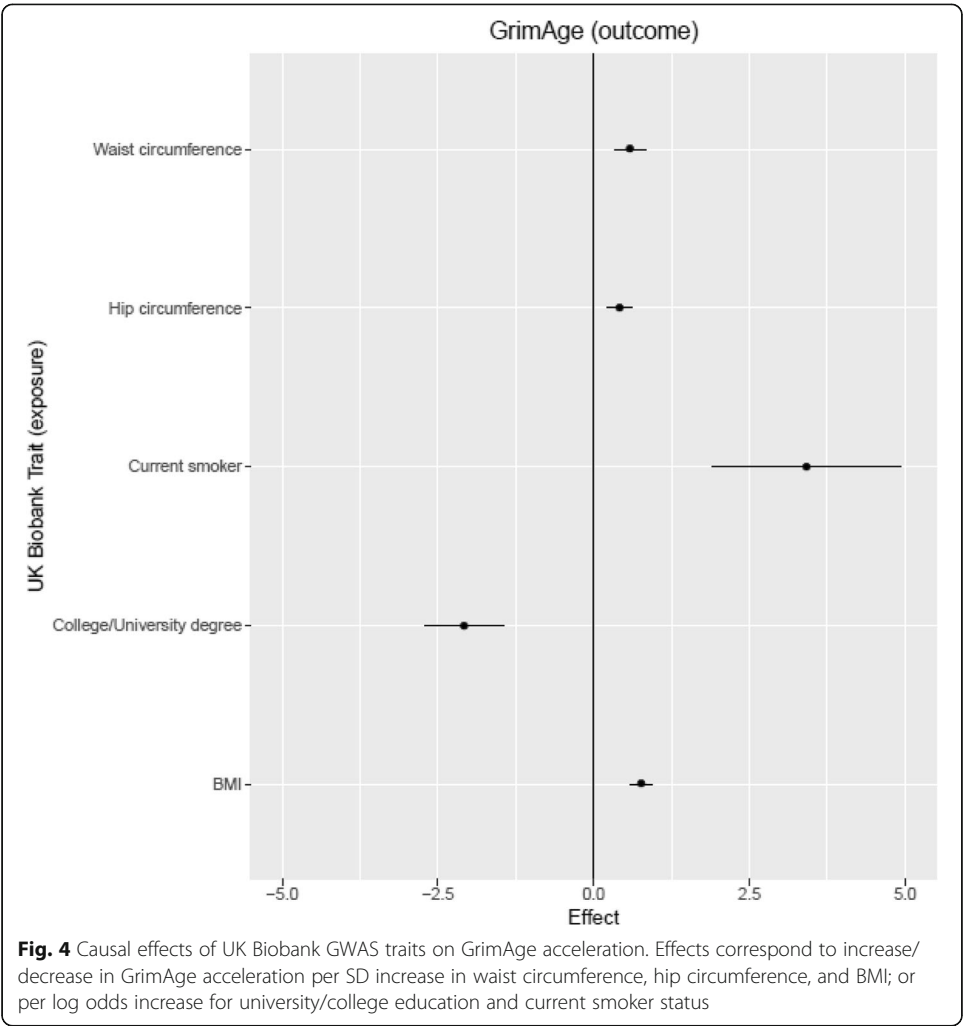
The biomarker analyses with epigenetic measures as outcomes identified no consistent effects across all four MR methods (Additional file 2: Table S29). We found limited evidence to support any causal effects of the epigenetic measures (as exposures) on key disease and health outcomes, including longevity (Additional file 2: Tables S30–S31).



**Fig. 3** **A** Polygenic predictions for the six epigenetic biomarkers in LBC1921, LBC1936, SABRE, Born in Bradford, ARIES, FHS, and the Young Finns Study. IEAA (intrinsic epigenetic age acceleration), PAI1 (plasminogen activator inhibitor-1). **B** Associations between GrimAge polygenic risk score ( $P < 1$ ) and UK Biobank GWAS traits

**Discussion**

Epigenetic biomarkers of aging and mortality have been extensively studied in relation to a plethora of health and disease outcomes. Here, we conducted a comprehensive suite of analyses in a meta-analysis sample of over 40,000 individuals, including the first GWA studies of two DNAm-based estimators of mortality risk (PhenoAge and GrimAge), as well as for DNAm-based proxies for granulocyte proportion and plasminogen activation inhibitor 1 protein levels. We identified 137 loci, of which 113 were novel, related to the six epigenetic biomarkers. Whereas previous studies have shown a general decline in longitudinal Hannum and Horvath age acceleration [53], there was no evidence of heterogeneity by cohort age for the lead independent loci. Although our comparison of genetic architectures across different ancestries was limited by sample size, particularly in the Hispanic ancestry lookup cohort of just 287 individuals, our European and African American trans-ethnic meta-analyses implicated many shared genetic loci. However, heterogeneity of effect sizes between European and African American ancestries was found for genome-wide significant loci which may reflect



differential tagging of underlying causal variants. Alternatively, gene-environment interactions or poor prediction of epigenetic biomarkers in African Americans, possibly due to the construction of the clocks relying on European ancestry individuals, may explain heterogeneity in these effect sizes.

The IEAA clock is a derivative of Horvath’s pan-tissue predictor that regresses out DNA methylation-based estimates for naive CD8+ T cells, exhausted CD8+ T cells, plasmablasts, CD4+ T cells, natural killer cells, monocytes, and granulocytes. Although other cell sub-types may influence the findings, the GWAS results that are shared between IEAA and the other epigenetic clocks (Hannum clock, PhenoAge, GrimAge) are less likely to be influenced by differences in cell composition. However, there was still an enrichment in the GWAS results for IEAA and GrimAge with findings from previous white blood cell GWASs, such as basophils. Mendelian randomization analyses indicated putative causal effects of lymphocyte count on PhenoAge, GrimAge, and Hannum age acceleration but not on IEAA. Furthermore, whereas we observed enrichment of Mendelian disease genes for IEAA relating to platelet disorders, the Mendelian randomization analyses did not support evidence for a causal link between platelets and IEAA. IEAA and PhenoAge acceleration share the following genome-wide



significant gene-based associations: *TPMT*, *TERT*, *NHLRC1*, *KDM1B*, *EDARADD*. IEAA and Hannum age acceleration share the following associations: *TERT*, *TRIM59*, *KPNA4*, *RP11-432B6.3*, *IFT80*, and *TET2*. *TET2* is particularly interesting in light of its mechanistic role (catalyzing the conversion of methylcytosine to 5-hydroxymethylcytosine) and its established role in several aging/regenerative phenotypes [54, 55].

Several of the GWAS overlapping genes in the European ancestry analysis (e.g., *TRIM59*, *KPNA4*) are also implicated by our eQTL colocalization analysis. While it is important to note that these analyses do not test for causality, we identified regulatory relationships between SNPs associated with IEAA and expression of nearby genes. Moreover, four of these loci were not merely mQTLs for CpG sites that were used to construct the clocks. The IEAA-associated locus on chromosome 3 (lead SNP rs1047210) colocalized with an eQTL for *TRIM59*. DNA methylation levels at *TRIM59* have been robustly associated with chronological age and its expression has been noted in multiple cancers [56–60]. The same locus was associated with IEAA in a previous GWAS (lead SNP rs11706810) [8]. In addition to *TRIM59*, the eQTL colocalization also implicated this locus in the expression of the pseudogene *KRT8P12*, along with *SMC4* and *KPNA4*. There is evidence that *SMC4* inhibits cellular senescence, an established hallmark of aging [61]. *SMC4* is a core subunit of condensin complexes, which contribute to senescence processes, possibly through reorganization of genomic structure and transcriptional regulation [62]. *KPNA4* is a member of the importin family of nuclear transport receptors, which work through the nuclear pore complex to selectively transport proteins from the cytoplasm to the nucleus. Like cellular senescence, dysfunction of nuclear transport has been proposed as a marker of aging [63]. Moreover, importin levels have been associated with impaired myocardial angiogenesis in aging [64]. A chromosome 1 locus associated with IEAA (lead SNP rs7550821) colocalized with an eQTL for *CD46*, encoding a regulator of T cell function and the complement system—a key component of the innate immune system where it promotes inflammation [65, 66]. There was also strong evidence for colocalization between an IEAA-associated locus on chromosome 15 (lead SNP rs12903325), and an eQTL for the lipid transporter gene *ATP8B4*, which contains variants that have been reported in relation to centenarian status in Italians and Alzheimer’s disease [67, 68]. An eQTL for *CXXC4*, encoding Idax, an inhibitor of Wnt signalling, colocalized with an IEAA-associated locus on chromosome 4 (lead SNP rs144317085). *CXXC4* levels, along with *KPNA4* and *SMC4*, have been associated with cancer progression [69–71].

Using the findings from the European ancestry meta-analyses, we observed shared genetic contributions between PhenoAge and GrimAge acceleration with education, cognitive ability, adiposity, and smoking. There were also significant epigenetic biomarker-specific genetic correlations with numerous other health and modifiable lifestyle factors. These are the first reported genetic correlations between GrimAge and PhenoAge acceleration and health and utilize much better powered GWAS results for IEAA and Hannum age acceleration. The best epigenetic biomarker of mortality risk, GrimAge acceleration, exhibited strong genetic correlations with parental longevity and lung cancer. As GrimAge uses a DNA methylation-based estimator of smoking pack-years in its definition, it is possible that the significant genetic correlation with lung cancer is mediated by smoking. Whereas many studies have observed phenotypic

correlations between epigenetic clocks and health/aging outcomes, other than previous small-scale efforts [8, 9], none have done so through shared genetics. Furthermore, this is the first large-scale epigenetic clock study to attempt to separate correlation from causation. By comprehensively examining the genetic architectures of each clock, we begin to uncover the shared and unique biological signals that each clock is tracking.

Four different Mendelian randomization methods provide directional evidence of a causal influence of adiposity-related traits on GrimAge acceleration, while smoking cessation was inversely related to GrimAge acceleration. Several MR analyses indicate that increased educational attainment is associated with lower GrimAge acceleration. However, there was no causal evidence for associations with other lifestyle traits, such as alcohol consumption. There was limited evidence from MR to implicate any of the DNAm predictors as playing an important causative role in longevity or disease risk.

This is the first study to present polygenic risk scores for six epigenetic biomarkers of aging. A phenome-wide scan of the six polygenic risk scores (PRS) yielded highly significant associations between the PRS of GrimAge acceleration and adiposity-related traits, education, and parental longevity. Both the PRS analysis and the MR analyses suffer from two limitations: (i) the geographical structure in the UK Biobank cohort might confound these analyses [72] and (ii) low heritability estimates for some of the phenotypes (e.g., longevity or epigenetic biomarkers such as PAI1). The lack of replication of previous PAI1 GWAS findings, despite an equivalent sample size, questions the validity of the DNAm PAI1 proxy for GWAS analyses [25]. Limited statistical power due to low heritability or low sample size may help explain the disconnect between the genetic correlation analysis (which revealed a plethora of significant genetic correlations for GrimAge and PhenoAge) and the MR analysis (which led to a dearth of significant findings). In general, careful interpretation of the MR findings is required. Causal MR analyses that modelled epigenetic biomarkers as exposures and disease states as outcomes suffered from weak genetic instruments (e.g., for GrimAge acceleration and granulocyte proportions, where the variance explained was < 1%) or inadequate power in two-sample analysis. Future multivariate MR analyses will be required to test whether the protective causal effect of education on GrimAge is mediated by smoking, obesity, or other factors. These studies could also explore potential pleiotropy between the clocks, health outcomes, and white blood cell proportions. Furthermore, while the assumption of non-pleiotropic associations can be examined in the MR framework, this is not the case for the genetic correlation and polygenic risk score analyses.

Since carrying out these analyses, more accurate DNAm-based age predictors have been developed [73]. Compared to the original Hannum and Horvath clocks, the Zhang clock is less sensitive to cellular heterogeneity. Future studies should consider this clock for GWAS analysis. Furthermore, whereas we selected the original Hannum clock over a slightly modified version (labelled extrinsic epigenetic age acceleration), the correlation between the two age acceleration measures was > 0.95 in the two subsets ( $n_{\text{Set1}} = 5087$ ;  $n_{\text{Set2}} = 4450$ ) of the Generation Scotland study, the largest single cohort study in the meta-analysis.

## Conclusions

Overall, this study highlights the shared genetic architecture between epigenetic aging, lifestyle factors (smoking, obesity), and parental longevity, which shows that DNAm-based biomarkers are valuable endophenotypes of biological aging.

## Methods

### Study cohort information

The meta-analysis sample comprised 36 datasets from 30 cohorts encompassing 40,905 participants (controls/healthy volunteers). Of these, 28 included individuals of European ancestry comprising 34,710 participants, and seven were of individuals of African American ancestry comprising 6195 participants. The total meta-analysis sample age range was 27.2–79.1 years (mean 54.0 years overall; 54.8 years across European ancestry cohorts; 50.4 years across African American cohorts) and comprised 0–100% females (mean 58.3%; 57.3% across European ancestry cohorts; 62.6% across African American cohorts). A Hispanic ancestry subset of the MESA cohort comprising 287 participants was used to compare effect sizes from trans-ethnic meta-analysis outputs and a European ancestry cohort comprising 1402 individuals (Young Finns Study) was used for polygenic prediction. Cohort-level descriptive data are presented in Additional file 2: Table S1 and described in Additional file 1.

### Data preparation

Age-adjusted DNA methylation-based estimates of Hannum age, Intrinsic Horvath age, PhenoAge, GrimAge, plasminogen activator inhibitor-1 levels, and unadjusted granulocyte proportion were calculated using the Horvath epigenetic age calculator software (<https://dnamage.genetics.ucla.edu/> or standalone scripts provided by Steve Horvath and Ake Lu). The following outputs were assessed: intrinsic epigenetic age acceleration—“IEAA”, Hannum age acceleration—“AgeAccelerationResidualHannum”, PhenoAge acceleration—“AgeAccelPheno”, GrimAge acceleration—“AgeAccelGrim”, estimate levels of Plasminogen Activation Inhibitor 1, adjusted for age—“DNAmPAIadjAge”, and estimated proportion of granulocytes—“Gran”. For each cohort, an outlier threshold for methylation values of  $\pm 5$  standard deviations was applied and outlier samples were excluded from the analysis.

### GWAS and meta-analysis

Quality control and imputation were done by each study separately (Additional file 1). Genotypes were imputed to either Haplotype Reference Consortium (HRC) or 1000 genomes phase 3 panels [74, 75]. In each cohort, association testing was conducted using imputed dosages using an additive model. Linear models were adjusted for sex and genetic principal components. GWAS summary statistics were obtained for between 1,097,816 and 15,221,271 genetic markers. This was the case for all cohorts with the exception of GOLDN (whole-genome sequence data) and the Sister Study (imputed data not available at the time of analysis). For each cohort, summary statistics were processed and harmonized using the R package EasyQC [76]. Multi-allelic variants were filtered to contain the variant with the highest minor allele count. At the individual cohort level, variants that were monomorphic, with a minor allele count  $\leq 25$ , genotyped

in < 30 individuals, or with an imputation quality score < 0.6 were removed. Allele codes and marker names were harmonized, duplicate variants were removed, and allele frequency checks were performed against the appropriate population reference data. Meta-analyses were performed with METAL using an inverse variance fixed effects scheme [77]. Meta-analyses were performed on European ancestry and African American studies separately ( $n = 34,710$  and  $6195$ , respectively). Variants were omitted from the meta-analysis if they were absent from > 50% of the total meta-analysis sample size. Cohort-specific genomic inflation factors ranged from 0.86 to 1.07. Genome-wide significance was defined as  $P < 5 \times 10^{-8}$ . To summarize the associations in terms of index SNP with the strongest association and other SNPs in linkage disequilibrium, we used conditional and joint association analysis of GWAS summary data, including the *HLA* region, in the GCTA-COJO software [22, 23]. A stepwise selection model was used with default settings for SNP LD ( $R^2 < 0.9$ ), analysis window size (10 Mb), and genome-wide significance ( $P < 5 \times 10^{-8}$ ) using HRC imputed genotype data from Generation Scotland, and 1000G imputed genotype data from ARIC as the reference panels for the European ancestry and African American analyses, respectively. Heterogeneity  $I^2$  statistics were obtained from the meta-analyses and plotted against both  $-\log_{10} p$  values and effect sizes to determine if SNPs with heterogeneous effects across cohorts were more statistically significant or had larger effect sizes. Systematic between-study heterogeneity was also investigated [21]. Meta-analyses were re-run after excluding cohorts identified as outliers and effect sizes were visually compared with the full meta-analysis output. Forest plots were prepared for all significant loci.

### Trans-ethnic meta-analysis

A trans-ethnic meta-analysis of all European ancestry and African American cohorts was conducted using default settings in the Meta-Regression of Multi-Ethnic Genetic Association (MR-MEGA) tool [27]. We considered summary output for the first principal component of the meta-regression.

MR-MEGA summary statistics were uploaded to the Functional Annotation of Meta-Analysis Summary Statistics (FUMA) (<http://fuma.ctglab.nl>) software for annotation of the top loci using default settings, selecting 1000 Genomes phase 3 (all populations) as the reference population [79]. Independent lead SNPs had  $P < 5 \times 10^{-8}$  and were independent of each other at  $r^2 < 0.6$ ; lead SNPs within this subset were required to have  $r^2 < 0.1$ . A locus was defined by considering lead SNPs in a 250-kb range and all SNPs in LD ( $r^2 \geq 0.6$ ) with at least one independent SNP.

### Functional annotation of meta-analysis summary statistics

The European ancestry and African American meta-analysis summary statistics were uploaded to FUMA (<http://fuma.ctglab.nl>) for further annotation and functional analysis [78]. Genes were annotated from SNP-level data using the “SNP2GENE” tool, permitting gene set and tissue expression analyses using MAGMA [28]. Positional mapping was performed based on ANNOVAR annotations, applying a maximum distance of 10 kb between SNPs and genes. A Bonferroni-corrected significance threshold (adjusting for 18,606 tested genes) of  $P < 2.69 \times 10^{-6}$  was set for the gene-based GWAS. Genes annotated to significant GWAS loci were further investigated using the

“GENE2FUNC” tool in FUMA for enrichment of GWAS catalog gene sets, omitting the MHC region [79]. Bonferroni-corrected  $P$  value thresholds were applied.

### Functional enrichment

To test if the GWAS meta-analysis findings were associated with regulatory and functional features of interest, enrichment analyses were conducted using GARFIELD [43]. SNPs were first pruned ( $r^2 > 0.1$ ) then annotated to categories (e.g., chromatin states, histone modifications, DNaseI hypersensitive sites, and transcription factor binding sites). Statistical enrichment was then carried out for SNPs at two  $P$  value thresholds ( $P < 1 \times 10^{-5}$  and  $P < 1 \times 10^{-8}$ ) while accounting for MAF, distance to the nearest TSS, and number of LD proxies.

### Colocalization analysis

We hypothesized that some of the lead loci from the meta-analyses will have shared variants (1) across the Epigenetic biomarkers, (2) with DNAm sites in blood, and (3) with gene expression levels in blood. We used GoDMC summary statistics on 190,102 DNAm sites [29] to examine the overlap between loci and epigenetic clock DNAm sites, BMI-associated DNAm sites, and smoking-associated DNAm sites. We used eQTL Gen summary statistics on 19,942 transcripts [80] that were available in the MR-Base database [81]. We used the Rpackage gwasglue (<https://mrcieu.github.io/gwasglue/>) to extract SNPs that were  $\pm 1$  Mb of the lead SNP and to harmonize the datasets. For each epigenetic biomarker–molecular trait pair (or pair of epigenetic biomarkers), we then performed colocalization analysis using the coloc.abf function in the R/coloc package [26], using default parameters. We only kept colocalized pairs with more than 50 shared SNPs and a posterior probability above 0.8 ( $PP > 0.80$ ). We removed the *HLA* region from the eQTL colocalization analysis.

### Disease and phenotype ontology enrichment

The potential role of Mendelian disease genes and associated pathways in influencing the epigenetic biomarkers was investigated with MendelVar [44], independently for each marker phenotype. We did not limit our analysis to any particular phenotype class among the Mendelian disease genes but looked for enrichment of any disease processes found to be strongly linked to genes in the GWAS loci. MendelVar analysis was run using intervals based on  $\pm 0.5$  Mbp window around the lead SNPs using the 1000 Genomes EUR population as LD reference [74]. Inside MendelVar, INRICH was run in “target” enrichment mode, with the target gene set filter set at minimum 5 (-i option) and maximum of 20,000 (-j option), and minimum observed threshold of 2 (-z option) [82]. The nominal  $P$  values were corrected for multiple testing with two rounds of permutation in INRICH.

### Heritability and genetic correlation analysis

LD score regression, using LD scores and weights estimated from European ancestry populations (downloaded from <https://data.broadinstitute.org/alkesgroup/LDSCORE/>), was used to assess genetic correlations between the six epigenetic biomarkers. Genetic correlations were further assessed between the six epigenetic biomarkers and publicly

available GWAS summary statistics using the LDHub web interface (<http://ldsc.broadinstitute.org/ldhub/>) [45]. Meta-analysis results for each epigenetic biomarker were uploaded to the LDHub website, selecting all available traits for genetic correlation analysis. SNP heritability was estimated using univariate LD score regression. As the majority of large-scale GWA studies have been based on European ancestry populations, heritability and genetic correlation analyses were limited to this group to maximize statistical power. Filtering was performed to exclude traits where the LD Hub output came with the following warning messages: “Caution: using these data may yield less robust results due to minor departure of the LD structure” and “Caution: using this data may yield results outside bounds due to relative low Z score of the SNP heritability of the trait.” This left a total of 693 unique traits from 708 to 711 studies per epigenetic biomarker. A Bonferroni-corrected significance threshold of  $P < 0.05/693 = 7.21 \times 10^{-5}$  was applied.

### Polygenic risk scores

To determine the proportion of variance in the six epigenetic biomarkers that can be explained by common genetic variants, we carried out a polygenic risk score analysis using results from the European ancestry meta-analyses. Weights for the additive genetic scores were created by re-running the meta-analyses excluding one cohort (test cohort) at a time. Six weighted PGR scores (one for each epigenetic biomarker) were generated using default settings of the PRSice software (clump-kb = 250, clump- $p$  = 1; clump- $r^2$  = 0.25) [83].  $P$  value thresholds were set at  $< 5 \times 10^{-8}$ ,  $< 0.01$ ,  $< 0.05$ ,  $< 0.1$ ,  $< 0.5$ , and 1. Linear regression models were built to calculate the incremental  $R^2$  between the null model (epigenetic biomarker ~ sex) and full model (epigenetic biomarker ~ sex + polygenic risk score) in the test cohort. The procedure was iterated after excluding different test cohorts one by one (Lothian Birth Cohort 1921, Lothian Birth Cohort 1936, Framingham Heart Study, Born in Bradford, ARIES, and SABRE, respectively) from the meta-analysis. Finally, these steps were repeated, using the full meta-analysis summary statistic output to generate polygenic risk scores in a completely independent cohort (Young Finns Study,  $n = 1320$ ).

For Born in Bradford, ARIES, and SABRE, best-guess genotype files with a MAF cut-off of 1% and info score  $> 0.8$  were generated and the polygenic risk score analyses were corrected for 20 genetic PCs.

A phenome-wide association study of 581 heritable traits from the UK Biobank study was then carried out for polygenic risk scores based on independent SNPs with  $P < 5 \times 10^{-8}$  or  $P < 1$  from each of the six GWAS meta-analyses ([http://mrcieu.mrsoftware.org/PRS\\_atlas/](http://mrcieu.mrsoftware.org/PRS_atlas/)) [52]. The  $P$  value thresholds were based on the leave-one-out cohort PRS analyses described above (GrimAge acceleration and Granulocyte proportions:  $P < 1$ ; IEAA, Hannum Age acceleration, PhenoAge acceleration, IEAA and PAII levels:  $P < 5 \times 10^{-8}$ ). An FDR-corrected  $P$  value ( $P_{\text{FDR}} < 0.05$ ) was applied separately to each set of PheWAS results.

### Mendelian randomization

To investigate if the epigenetic biomarkers were (i) causally influenced by lifestyle factors and (ii) had a causal effect on aging and disease outcomes, Mendelian



randomization (MR) was performed in MR-Base [81]. The epigenetic measures were considered as both exposures (i.e., causally influencing the outcome) and outcomes (i.e., the epigenetic measure being causally influenced by a trait of interest). The analyses were further split into four sections: primary exposures/outcomes (common lifestyle risk factors and aging/disease outcomes from the largest available GWAS in MR-Base); secondary exposures/outcomes (traits identified as relevant via moderate genetic correlations from the LD regression analyses); 34 cell count exposures [24]; and 38 biomarker exposures [84]. SNPs instrumenting each exposure were clumped using a European LD reference panel and an  $r^2 < 0.001$ . Harmonization of the SNP effects with the exposure and outcomes were performed so that palindromic SNPs were aligned when minor allele frequencies (MAFs) were  $< 0.3$  or were otherwise excluded.

Inverse variance weighted (IVW) MR was carried out as the main analysis, with pleiotropy-robust sensitivity analyses featuring MR-Egger [85], weighted median [86], and weighted mode MR [87]. Significant associations were defined by a Bonferroni-corrected  $P$  value  $< 0.05$ . Where there was evidence for a causal effect based of the IVW model, we also assessed the potential for horizontal pleiotropy by means of heterogeneity assessment (Cochran's  $Q$ -statistic) of individual SNP effects in both IVW and MR-Egger analyses and the Egger intercept test for directional pleiotropy [85].

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-021-02398-9>.

**Additional file 1.** Individual cohort descriptions and acknowledgements.

**Additional file 2.** Supplementary Tables -Tables S1-S31.

**Additional file 3.** Supplementary Figures - Figures S1-S31.

**Additional file 4.** Assessment of genomic inflation and heterogeneity.

**Additional file 5.** Colocalization plots.

**Additional file 6.** Review history.

## Peer review information

Andrew Cosgrove was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

## Review history

The review history is available as Additional file 6.

## Authors' contributions

Conception/design of the work: DLM, JLM, RCR, ATL, MKS, GD, CLR, SH, REM. Analysis, interpretation of data: DLM, JLM, RCR, ATL, MKS, GD, LB, XG, AJ, JJ, SKas, SKat, PK, PRM, PPM, MN, TP, AP, LMR, SMR, TGR, OR, MS, DS, PT, MDvdZ, RMW, XW, YW, RX, ZX, JY, WZ. Drafted the work: DLM, JLM, RCR, ATL, MKS, SH, REM. Substantive revisions: all authors. The author(s) read and approved the final manuscript.

## Funding

REM, SH, and AL are supported by a National Institute of Health U01 grant, U01AG060908-01. REM and DLM are supported by Alzheimer's Research UK major project grant, ARUK-PG2017B-10. RCR is a de Pass Vice Chancellor's Research Fellow at the University of Bristol. CR and RCR receive support from a Cancer Research UK Program Grant (C18281/A191169). CLR, JLM, and RCR are members of the UK Medical Research Council Integrative Epidemiology Unit at the University of Bristol (MC\_UU\_00011/5). IJD is supported by Age UK (Disconnected Mind program), UKRI Medical Research Council grant, MR/R0245065/1, and by National Institute of Health R01 grant, 1R01AG054628-01A1. GD is supported by the University of Edinburgh School of Philosophy, Psychology and Language Sciences. Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN268201800001I). We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. Cohort-specific acknowledgements are presented in Additional file 1.

### Availability of data and materials

Meta-analysis summary statistics for each epigenetic biomarker are publicly available at <https://datashare.is.ed.ac.uk/handle/10283/3645> and the GWAS Catalog ([http://ftp.ebi.ac.uk/pub/databases/gwas/summary\\_statistics/GCST90014001-GCST90015000](http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90014001-GCST90015000); accession numbers GCST90014287-GCST90014304). For cohort-specific details, please see Additional file 1, which contains information for each study.

### Declarations

#### Ethics approval and consent to participate

Each of the studies was approved by their local Ethical Committee. All subjects provided written informed consent. For further details please see Additional file 1.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Crewe Road South, Edinburgh EH4 2XU, UK. <sup>2</sup>MRC Integrative Epidemiology Unit University of Bristol, Bristol, UK. <sup>3</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. <sup>4</sup>Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA. <sup>5</sup>Lothian Birth Cohort, Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK. <sup>6</sup>Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands. <sup>7</sup>The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. <sup>8</sup>Swiss Tropical and Public Health Institute, Basel, Switzerland. <sup>9</sup>University of Basel, Basel, Switzerland. <sup>10</sup>National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, USA. <sup>11</sup>Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. <sup>12</sup>Department of Gynecology and Obstetrics, Emory University School of Medicine, Atlanta, GA, USA. <sup>13</sup>Longitudinal Study Section, Translational Gerontology Branch, National Institute on Aging, Baltimore, MD, USA. <sup>14</sup>Research Unit Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Bavaria, Germany. <sup>15</sup>Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Bavaria, Germany. <sup>16</sup>TUM School of Medicine, Technical University of Munich, Munich, Germany. <sup>17</sup>Department of Clinical Chemistry, Fimlab Laboratories, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, 33520 Tampere, Finland. <sup>18</sup>Epidemiology, Biostatistics and Biodemography, Department of Public Health, University of Southern Denmark, Odense, Denmark. <sup>19</sup>Department of Clinical Genetics, Odense University Hospital, Odense, Denmark. <sup>20</sup>Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Helsinki, Finland. <sup>21</sup>Department of Biostatistics, University of Alabama at Birmingham, Birmingham, USA. <sup>22</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. <sup>23</sup>Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, USA. <sup>24</sup>MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. <sup>25</sup>Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark. <sup>26</sup>Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China. <sup>27</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. <sup>28</sup>Department of Biomedical Sciences, Chang Gung University, Taoyuan, Taiwan. <sup>29</sup>Division of Pediatric Infectious Diseases, Department of Pediatrics, Chang Gung Memorial Hospital, Taoyuan City, Taiwan. <sup>30</sup>Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. <sup>31</sup>Amsterdam Public Health Research Institute, Amsterdam, The Netherlands. <sup>32</sup>Cancer Epidemiology Division, Cancer Council Victoria, 615 St Kilda Road, Melbourne, Victoria 3004, Australia. <sup>33</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Solna, Sweden. <sup>34</sup>Brown Foundation Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, USA. <sup>35</sup>National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA. <sup>36</sup>The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. <sup>37</sup>Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. <sup>38</sup>School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA. <sup>39</sup>Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria 3168, Australia. <sup>40</sup>Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, 207 Bouverie Street, Melbourne, Victoria 3010, Australia. <sup>41</sup>Department of Pathology & Laboratory Medicine, Larner College of Medicine, University of Vermont, Burlington, VT 05446, USA. <sup>42</sup>Genetics of DNA methylation Consortium, Bristol, UK. <sup>43</sup>Department of Clinical Physiology, Tampere University Hospital, and Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, 33521 Tampere, Finland. <sup>44</sup>Children's Minnesota Research Institute, Children's Minnesota, Minneapolis, MN 55404, USA. <sup>45</sup>Department of Biostatistics, Boston University School of Public Health, Boston, USA. <sup>46</sup>NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK. <sup>47</sup>Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK. <sup>48</sup>Division of Psychiatry, University of Edinburgh, Edinburgh, UK. <sup>49</sup>Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA. <sup>50</sup>TOPMed Data Coordinating Center, Genetic Analysis Center, Department of Biostatistics, University of Washington, Seattle, USA. <sup>51</sup>Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA. <sup>52</sup>German Center for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany. <sup>53</sup>Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland. <sup>54</sup>Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. <sup>55</sup>Department of Clinical

Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. <sup>56</sup>Department of Public Health Sciences, Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA. <sup>57</sup>Gerontology Research Center, Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland. <sup>58</sup>Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA. <sup>59</sup>Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Bavaria, Germany. <sup>60</sup>Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, 55101 Mainz, Germany. <sup>61</sup>Chair of Genetic Epidemiology, Institute for Medical Information Processing, Biometry, and Epidemiology, Faculty of Medicine, Ludwig-Maximilians-Universität München, Munich, Germany. <sup>62</sup>MRC Unit for Lifelong Health and Ageing at UCL, London, UK. <sup>63</sup>Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands. <sup>64</sup>Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA. <sup>65</sup>Division of Cardiology, Beth Israel Deaconess Medical Center, Boston, MA, USA. <sup>66</sup>Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MS, USA. <sup>67</sup>Department of Bioinformatics, Institute of Health Sciences, Hacettepe University, 06100 Ankara, Turkey. <sup>68</sup>Deans Office, College of Public Health, University of Kentucky, Lexington, UK. <sup>69</sup>Geriatric Unit, Azienda Sanitaria Toscana Centro, Florence, Italy. <sup>70</sup>Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA, USA. <sup>71</sup>Population Sciences in the Pacific Program (Cancer Epidemiology), University of Hawai'i Cancer Center, University of Hawai'i, Honolulu, HI, USA. <sup>72</sup>Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA. <sup>73</sup>Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA. <sup>74</sup>MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Crewe Rd. South, Edinburgh EH4 2XU, UK. <sup>75</sup>Dept of Epidemiology, University of Alabama at Birmingham, Birmingham, USA. <sup>76</sup>Department of Public Health, University of Helsinki, Helsinki, Finland. <sup>77</sup>Bristol NIHR Biomedical Research Centre, Bristol, UK. <sup>78</sup>Department of Epidemiology, University of Washington, Seattle, WA, USA. <sup>79</sup>Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA 90095, USA.

Received: 14 October 2020 Accepted: 3 June 2021

Published online: 29 June 2021

## References

- Niccoli T, Partridge L. Ageing as a risk factor for disease. *Curr Biol*. 2012;22(17):R741–52.
- Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet*. 2018; 19(6):371–84 <https://doi.org/10.1038/s41576-018-0004-3>.
- McCartney DL, Hillary RF, Stevenson AJ, et al. Epigenetic prediction of complex traits and death. *Genome Biol*. 2018; 19(1):136 <https://doi.org/10.1186/s13059-018-1514-1>.
- Hamilton OKL, Zhang Q, McRae AF, et al. An epigenetic score for BMI based on DNA methylation correlates with poor physical health and major disease in the Lothian Birth Cohort. *Int J Obes*. 2019;43(9):1795–802 <https://doi.org/10.1038/s41366-018-0262-3>.
- Quach A, Levine ME, Tanaka T, et al. Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany NY)*. 2017; <https://doi.org/10.18632/aging.101168>.
- Marioni RE, Harris SE, Shah S, McRae AF, von Zglinicki T, Martin-Ruiz C, et al. The epigenetic clock and telomere length are independently associated with chronological age and mortality. *Int J Epidemiol*. 2016;45(2):424–32 <https://doi.org/10.1093/ije/dyw041>.
- Marioni RE, Shah S, McRae AF, et al. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol*. 2015;16(1):25 <https://doi.org/10.1186/s13059-015-0584-6>.
- Lu AT, Xue L, Salfati EL, Chen BH, Ferrucci L, Levy D, et al. GWAS of epigenetic aging rates in blood reveals a critical role for TERT. *Nat Commun*. 2018;9(1):387 <https://doi.org/10.1038/s41467-017-02697-5>.
- Gibson J, Russ TC, Clarke T-K, et al. A meta-analysis of genome-wide association studies of epigenetic age acceleration. *bioRxiv*. 2019; <https://doi.org/10.1101/585299>.
- Hillary RF, Stevenson AJ, Cox SR, McCartney DL, Harris SE, Seeboth A, et al. An epigenetic predictor of death captures multi-modal measures of brain health. *Mol Psychiatry*. 2019; <https://doi.org/10.1038/s41380-019-0616-9>.
- Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, et al. DNA methylation aging clocks: challenges and recommendations. *Genome Biol*. 2019;20(1):249 <https://doi.org/10.1186/s13059-019-1824-y>.
- Bocklandt S, Lin W, Sehl ME, Sánchez FJ, Sinheimer JS, Horvath S, et al. Epigenetic predictor of age. *PLoS One*. 2011; 6(6):e14821.
- Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sada SV, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*. 2013;49(2):359–67 <https://doi.org/10.1016/j.molcel.2012.10.016>.
- Horvath S. DNA methylation age of human tissues and cell types. *Genome Biol*. 2013;14(10):R115 <https://doi.org/10.1186/gb-2013-14-10-r115>.
- Robertson NA, Hillary RF, McCartney DL, Terradas-Terradas M, Higham J, Sproul D, et al. Age-related clonal haemopoiesis is associated with increased epigenetic age. *Curr Biol*. 2019;29(16):R786–7.
- Lu AT, Quach A, Wilson JG, et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. *Aging (Albany NY)*. 2019; <https://doi.org/10.18632/aging.101684>.
- Levine ME, Lu AT, Quach A, et al. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*. 2018; <https://doi.org/10.18632/aging.101414>.
- Hillary RF, Stevenson AJ, McCartney DL, Campbell A, Walker RM, Howard DM, et al. Epigenetic clocks predict prevalence and incidence of leading causes of death and disease burden. *bioRxiv*. 2020.
- Li X, Ploner A, Wang Y, Magnusson PK, Reynolds C, Finkel D, et al. Longitudinal trajectories, correlations and mortality associations of nine biological ages across 20-years follow-up. *Elife*. 2020;9:e51507.
- Horvath S, Ritz BR. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)*. 2015;7(12):1130–42.

21. Magosi LE, Goel A, Hopewell JC, Farrall M. Identifying systematic heterogeneity patterns in genetic association meta-analysis studies. *PLoS Genet.* 2017;13(5):e1006755.
22. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet.* 2012;44(4):369–75.
23. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76–82.
24. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell.* 2016;167(5):1415–29.
25. Huang J, Sabater-Lleal M, Asselbergs FW, Tregouet D, Shin SY, Ding J, et al. Genome-wide association study for circulating levels of PAI-1 provides novel insights into its regulation. *Blood.* 2012;120(24):4873–81.
26. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalization between pairs of genetic association studies using summary statistics. *PLoS Genet.* 2014;10(5):e1004383.
27. Mägi R, Horikoshi M, Sofer T, Mahajan A, Kitajima H, Franceschini N, et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet.* 2017;26(18):3639–50.
28. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* 2015;11(4):e1004219 <https://doi.org/10.1371/journal.pcbi.1004219>.
29. Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, et al. Genomic and phenomic insights from an atlas of genetic effects on DNA methylation. *medRxiv.* 2020.
30. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature.* 2017;541(7635):81–6 <https://doi.org/10.1038/nature20784>.
31. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet.* 2016;9(5):436–47 <https://doi.org/10.1161/CIRCGENETICS.116.001506>.
32. Lu AT, Hannon E, Levine ME, Crimmins EM, Lunnon K, Mill J, et al. Genetic architecture of epigenetic and neuronal ageing rates in human brain regions. *Nat Commun.* 2017;8:15353.
33. Shete S, Hosking FJ, Robertson LB, Dobbins SE, Sanson M, Malmer B, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet.* 2009;41(8):899–904.
34. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet.* 2009;85(5):679–91 <https://doi.org/10.1016/j.ajhg.2009.09.012>.
35. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet.* 2019;104(1):65–75.
36. van Rooij FJA, Qayyum R, Smith AV, Zhou Y, Trompet S, Tanaka T, et al. Genome-wide trans-ethnic meta-analysis identifies seven genetic loci influencing erythrocyte traits and a role for RBPMS in erythropoiesis. *Am J Hum Genet.* 2017;100(1):51–63.
37. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature.* 2012;491(7422):51–63.
38. De Lange KM, Moutsianas L, Lee JC, Lamb CA, Luo Y, Kennedy NA, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet.* 2017;49(2):256–61.
39. Demenais F, Margeritte-Jeannin P, Barnes KC, Cookson WOC, Altmüller J, Ang W, et al. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat Genet.* 2018;50(1):42–53.
40. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med.* 2010;363(13):1211–21.
41. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet.* 2012;44(12):1336–40.
42. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet.* 2010;42(6):508–14.
43. Iotchkova V, Ritchie GRS, Geihs M, Morganella S, Min JL, Walter K, et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet.* 2019;51(2):343–53.
44. Sobczyk MK, Gaunt TR, Paternoster L. MendelVar: gene prioritization at GWAS loci using phenotypic enrichment of Mendelian disease genes. *Bioinformatics.* 2021;37(1):1–8.
45. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, et al. LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics.* 2017;33(2):272–9 <https://doi.org/10.1093/bioinformatics/btw613>.
46. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science.* 2013;340(6139):1467–71.
47. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature.* 2016;533(7604):539–42 <https://doi.org/10.1038/nature17671>.
48. Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E, et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet.* 2017;49(7):1107–12 <https://doi.org/10.1038/ng.3869>.
49. Berndt SI, Gustafsson S, Mägi R, Ganna A, Wheeler E, Feitosa MF, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet.* 2013;45(5):501–12 <https://doi.org/10.1038/ng.2606>.
50. Shungin D, Winkler T, Croteau-Chonka DC, et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature.* 2015;518(7538):187–96 <https://doi.org/10.1038/nature14132>.
51. Patel YM, Park SL, Han Y, Wilkens LR, Bickeböller H, Rosenberger A, et al. Novel association of genetic markers affecting CYP2A6 activity and lung cancer risk. *Cancer Res.* 2016;76(19):5768–76 <https://doi.org/10.1158/0008-5472.CAN-16-0446>.
52. Richardson TG, Harrison S, Hemani G, Smith GD. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenotype. *Elife.* 2019;8:e43657.

53. Marioni RE, Suderman M, Chen BH, Horvath S, Bandinelli S, Morris T, et al. Tracking the epigenetic clock across the human life course: a meta-analysis of longitudinal cohort data. *J Gerontol A Biol Sci Med Sci*. 2019;74(1):57–61 <https://doi.org/10.1093/gerona/gly060>.
54. Gontier G, Iyer M, Shea JM, Bieri G, Wheatley EG, Ramalho-Santos M, et al. Tet2 rescues age-related regenerative decline and enhances cognitive function in the adult mouse brain. *Cell Rep*. 2018;22(8):1974–81 <https://doi.org/10.1016/j.celrep.2018.02.001>.
55. Wang Y, Sano S, Yura Y, Ke Z, Sano M, Oshima K, et al. Tet2-mediated clonal hematopoiesis in nonconditioned mice accelerates age-associated cardiac dysfunction. *JCI Insight*. 2020;5(6) <https://doi.org/10.1172/jci.insight.135204>.
56. Zbieć-Piekarska R, Spólnicka M, Kupiec T, Parys-Proszek A, Makowska Z, Pałeczka A, et al. Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Sci Int Genet*. 2015;17:173–9.
57. McCartney DL, Zhang F, Hillary RF, Zhang Q, Stevenson AJ, Walker RM, et al. An epigenome-wide association study of sex-specific chronological ageing. *Genome Med*. 2019;12(1):1.
58. Sun Y, Ji B, Feng Y, Zhang Y, Ji D, Zhu C, et al. TRIM59 facilitates the proliferation of colorectal cancer and promotes metastasis via the PI3K/AKT pathway. *Oncol Rep*. 2017;38(1):43–52 <https://doi.org/10.3892/or.2017.5654>.
59. Zhan W, Han T, Zhang C, Xie C, Gan M, Deng K, et al. TRIM59 promotes the proliferation and migration of non-small cell lung cancer cells by upregulating cell cycle related proteins. *PLoS One*. 2015;10(11) <https://doi.org/10.1371/journal.pone.0142596>.
60. Zhou Z, Ji Z, Wang Y, Li J, Cao H, Zhu HH, et al. TRIM59 is up-regulated in gastric tumors, promoting ubiquitination and degradation of p53. *Gastroenterology*. 2014;147(5):1043–54 <https://doi.org/10.1053/j.gastro.2014.07.021>.
61. Avelar RA, Ortega JG, Tacutu R, Tyler EJ, Bennett D, Binetti P, et al. A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol*. 2020;21(1):91.
62. Iwasaki O, Tanizawa H, Kim KD, Kosenkov A, Nacarelli T, Tashiro S, et al. Involvement of condensin in cellular senescence through gene regulation and compartmental reorganization. *Nat Commun*. 2019;10(1):5688.
63. Martins F, Sousa J, Pereira CD, da Cruz e Silva OAB, Rebelo S. Nuclear envelope dysfunction and its contribution to the aging process. *Aging Cell*. 2020;19(5):e13143.
64. Ahluwalia A, Narula J, Jones MK, Deng X, Tarnawski AS. Impaired angiogenesis in aging myocardial microvascular endothelial cells is associated with reduced importin  $\alpha$  and decreased nuclear transport of HIF1 $\alpha$ : Mechanistic implications. *J Physiol Pharmacol*. 2010;61(2):133–9.
65. Ni Choileain S, Astier AL. CD46 processing: a means of expression. *Immunobiology*. 2012;217(2):169–75 <https://doi.org/10.1016/j.imbio.2011.06.003>.
66. Astier AL. T-cell regulation by CD46 and its relevance in multiple sclerosis. *Immunology*. 2008;124(2):149–54 <https://doi.org/10.1111/j.1365-2567.2008.02821.x>.
67. Giuliani C, Sazzini M, Pirazzini C, Bacalini MG, Marasco E, Gnecci-Ruscone GA, et al. Impact of demography and population dynamics on the genetic architecture of human longevity. *Aging (Albany NY)*. 2018;10(8):1947–63.
68. Bellenguez C, Kucukali F, Jansen I, Andrade V, Morenau-Grau S, Amin N, et al. Large meta-analysis of genome-wide association studies expands knowledge of the genetic etiology of Alzheimer disease and highlights potential translational opportunities. *medRxiv*. 2020.
69. Kojima T, Shimazui T, Hinotsu S, Joraku A, Oikawa T, Kawai K, et al. Decreased expression of CXXC4 promotes a malignant phenotype in renal cell carcinoma by activating Wnt signaling. *Oncogene*. 2009;28(2):297–305 <https://doi.org/10.1038/onc.2008.391>.
70. Zhang C, Kuang M, Li M, Feng L, Zhang K, Cheng S. SMC4, which is essentially involved in lung development, is associated with lung adenocarcinoma progression. *Sci Rep*. 2016;6:34508.
71. Yang J, Lu C, Wei J, Guo Y, Liu W, Luo L, et al. Inhibition of KPNA4 attenuates prostate cancer metastasis. *Oncogene*. 2017;36(20):2868–78 <https://doi.org/10.1038/onc.2016.440>.
72. Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, et al. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat Commun*. 2019;10(8):333.
73. Zhang Q, Vallegger CL, Walker RM, Lin T, Henders AK, Montgomery GW, et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. *Genome Med*. 2019;11(1):54.
74. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74 <https://doi.org/10.1038/nature15393>.
75. McCarthy S, Das S, Kretschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279–83 <https://doi.org/10.1038/ng.3643>.
76. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9(5):1192–212 <https://doi.org/10.1038/nprot.2014.071>.
77. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190–1.
78. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826 <https://doi.org/10.1038/s41467-017-01261-5>.
79. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1):D1005–12 <https://doi.org/10.1093/nar/gky1120>.
80. Vösa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*. 2018.
81. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408.
82. Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: Interval-based enrichment analysis for genome-wide association studies. *Bioinformatics*. 2012;28(13):1797–9.
83. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics*. 2015;31(9):1466–8 <https://doi.org/10.1093/bioinformatics/btu848>.
84. Sinnott-Armstrong N, Tanigawa Y, Amar D, Mars NJ, Aguirre M, Venkataraman GR, et al. Genetics of 38 blood and urine biomarkers in the UK Biobank. *bioRxiv*. 2019.
85. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512–25 <https://doi.org/10.1093/ije/dyv080>.

86. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol.* 2016;40(4):304–14.
87. Hartwig FP, Smith GD, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol.* 2017;46(6):1985–98 <https://doi.org/10.1093/ije/dyx102>.

### Publisher's Note

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

