



# Mendelian randomization of genetically independent aging phenotypes identifies LPA and VCAM1 as biological targets for human aging

Paul R. H. J. Timmers<sup>1,2</sup>✉, Evgeny S. Tiys<sup>3,4</sup>, Saori Sakaue<sup>5,6,7,8</sup>, Masato Akiyama<sup>9,10</sup>, Tuomo T. J. Kiiskinen<sup>11</sup>, Wei Zhou<sup>8,12,13</sup>, Shih-Jen Hwang<sup>14,15</sup>, Chen Yao<sup>14,15</sup>, Biobank Japan Project\*, FinnGen\*, Joris Deelen<sup>16,17</sup>, Daniel Levy<sup>14,15</sup>, Andrea Ganna<sup>8,11,12</sup>, Yoichiro Kamatani<sup>9,18</sup>, Yukinori Okada<sup>16</sup>, Peter K. Joshi<sup>2</sup>, James F. Wilson<sup>1,2,20</sup> and Yakov A. Tsepilov<sup>3,19,20</sup>

**Length and quality of life are important to us all, yet identification of promising drug targets for human aging using genetics has had limited success. In the present study, we combine six European-ancestry genome-wide association studies of human aging traits—healthspan, father and mother lifespan, exceptional longevity, frailty index and self-rated health—in a principal component framework that maximizes their shared genetic architecture. The first principal component (aging-GIP1) captures both length of life and indices of mental and physical wellbeing. We identify 27 genomic regions associated with aging-GIP1, and provide additional, independent evidence for an effect on human aging for loci near *HTT* and *MAML3* using a study of Finnish and Japanese survival. Using proteome-wide, two-sample, Mendelian randomization and colocalization, we provide robust evidence for a detrimental effect of blood levels of apolipoprotein(a) and vascular cell adhesion molecule 1 on aging-GIP1. Together, our results demonstrate that combining multiple aging traits using genetic principal components enhances the power to detect biological targets for human aging.**

Aging affects us all, from the personal, progressive loss of health to the collective burden of chronic age-related disease and frailty on society. In humans, the body undergoes a systemic functional decline after reaching adulthood, which manifests itself as age-related disease, infirmity and eventually death<sup>1</sup>. The factors determining the rate of aging and death are complex and interlinked, and include genetics, lifestyle, environmental exposures and chance.

Quantifying the aging process is not straightforward. A variety of aging-related phenotypes has been studied as proxies, from chronological measurements such as the length of time from birth until the occurrence of a major disease (healthspan)<sup>2</sup> or death (lifespan)<sup>3,4</sup>, to cellular deterioration measurements such as telomere attrition<sup>5</sup> and loss of the Y chromosome<sup>6,7</sup>, to holistic measurements such as the frailty index<sup>8</sup>, encompassing multiple functional impairment indicators<sup>9</sup>. Although the genetic component of these aging-related traits tends to be estimated at <15%<sup>2,8,10</sup>, recent progress has been

made on characterizing this component using large genome-wide association studies (GWASs), and combining similar GWASs to increase statistical power<sup>11,12</sup>.

The benefit of combining GWASs of several aging phenotypes, especially in different populations, is the ability to detect biological mechanisms that influence multiple core components of aging, while downweighing population- and trait-specific features. For example, a recent multivariate analysis of healthspan, parental lifespan and longevity GWASs found that genetic loci that were not shared between traits often associated with population-specific, behavioral risk factors such as smoking and skin cancer<sup>11</sup>. On the other hand, that analysis showed that genetic loci shared between traits were associated with biological pathways such as cellular homeostasis and heme metabolism<sup>11</sup>.

However, to date, large aging-related trait GWASs have been combined only using multivariate analysis of variance, which detects

<sup>1</sup>MRC Human Genetics Unit, MRC Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. <sup>2</sup>Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh, UK. <sup>3</sup>Laboratory of Theoretical and Applied Population Genomics, Novosibirsk State University, Novosibirsk, Russia. <sup>4</sup>Laboratory of Glycogenomics, Institute of Cytology and Genetics, Novosibirsk, Russia. <sup>5</sup>Center for Data Sciences, Harvard Medical School, Boston, MA, USA. <sup>6</sup>Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan. <sup>7</sup>Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>8</sup>Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA. <sup>9</sup>Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. <sup>10</sup>Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. <sup>11</sup>Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. <sup>12</sup>Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. <sup>13</sup>Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA. <sup>14</sup>Framingham Heart Study, Framingham, MA, USA. <sup>15</sup>Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA. <sup>16</sup>Max Planck Institute for Biology of Ageing, Cologne, Germany. <sup>17</sup>Cologne Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases, University of Cologne, Cologne, Germany. <sup>18</sup>Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan. <sup>19</sup>Laboratory of Recombination and Segregation Analysis, Institute of Cytology and Genetics, Novosibirsk, Russia. <sup>20</sup>These authors contributed equally: James F. Wilson, Yakov A. Tsepilov. \*A list of members and affiliations appears at the end of the paper. ✉e-mail: [paul.timmers@ed.ac.uk](mailto:paul.timmers@ed.ac.uk)

a genetic variation that is either shared between multiple traits or strongly associated with a single trait<sup>13,14</sup>. This mixture generates heterogeneous SNP effect sizes, complicating the downstream analysis<sup>11</sup>. An alternative is to perform principal component analysis on the genetic covariance between traits and use the component loadings to construct new, genetically independent phenotypes (GIPs)<sup>15</sup>. As their name implies, GIPs capture the genetic covariance between phenotypes while being genetically uncorrelated to each other. In practice; this means that the first principal component (GIP1) maximizes the genetic overlap between all traits, whereas each subsequent GIP contains genetic variation distinguishing the traits from each other<sup>15</sup>.

In the present study, we cluster 11 large aging-related trait GWASs by genetic similarity, and explicitly quantify their common and unique genetic architecture using the GIP methodology. We then characterize this common genetic aging phenotype, identify robust genomic loci and highlight proteins that may be potential drug targets for improving the length and quality of life.

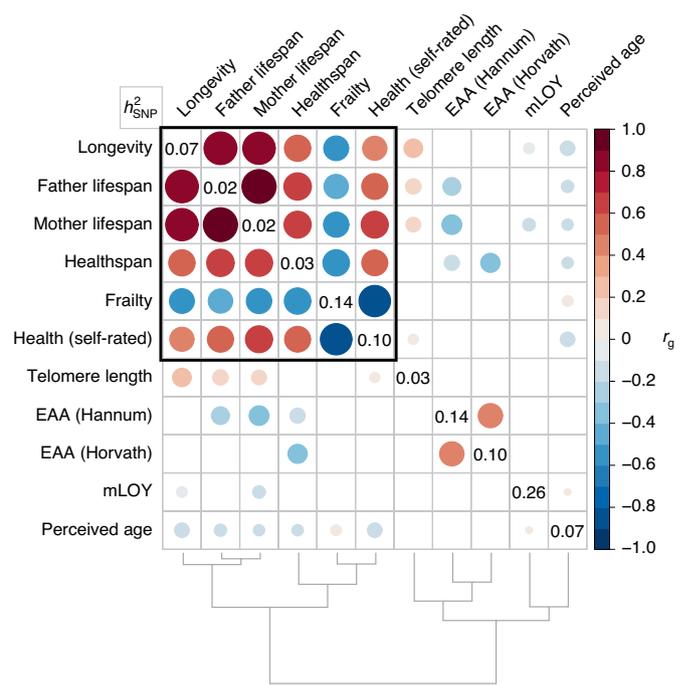
## Results

**Aging-related traits cluster based on genetic correlations.** We gathered publicly available GWAS summary statistics on aging-related traits measured in at least 10,000 European-ancestry individuals. These included self-rated health<sup>16</sup>, healthspan<sup>2</sup>, father lifespan<sup>4</sup>, mother lifespan<sup>4</sup>, exceptional longevity<sup>17</sup>, frailty index<sup>8</sup>, perceived age<sup>18</sup>, Hannum epigenetic age acceleration<sup>19</sup>, Horvath epigenetic age acceleration<sup>19</sup>, telomere length<sup>5</sup> and mosaic loss of the Y chromosome<sup>7</sup> (Supplementary Table 1). A variety of UK and European individuals is represented between these studies, from children (aged 10+) to centenarians (aged 100+), with birth years spanning the twentieth and early twenty-first century. The largest sample consists of UK Biobank participants and their parents (see Supplementary Note for details of each GWAS, and Extended Data Fig. 1 for correlations between studies due to sample overlap).

Calculating genetic correlations ( $r_g$ ) from summary statistics, and performing hierarchical clustering based on the magnitude of these correlations, we find that the first six traits—mostly chronological and holistic measures of aging—form a cluster of high genetic similarity ( $|r_g| \geq 0.5$ ;  $P < 5 \times 10^{-15}$ ). In contrast, the cellular measures of aging show low or no correlations with other traits ( $|r_g| \leq 0.3$ ), although the epigenetic age acceleration phenotypes correlate strongly with each other ( $r_g = 0.5$ ; 95% confidence interval (CI) = 0.2–0.8; Fig. 1 and Supplementary Data 1).

Testing the 6 aging-related traits in the main cluster for correlations with 728 other traits from the GWAS-MAP platform<sup>20</sup>, we find that their genetic similarity may be largely explained through strong, shared genetic correlations (meta  $|r_g| \geq 0.5$ , false discovery rate (FDR)  $< 0.05$ ,  $P_{\text{het}} > 0.05$ ,  $I^2 < 0.50$ ) with chest pain, cardiovascular disorders, smoking-related disease, type 2 diabetes, and general illness or medication use in UK Biobank (Supplementary Data 2). However, each core aging trait also has several genetic correlations that differ substantially (Methods) from the other aging-related phenotypes and may reflect population- or trait-specific risk factors. For example, self-rated health correlates more strongly with physical fitness, body mass index and a noisy work environment, healthspan correlates more strongly with skin and breast cancers, father lifespan correlates more strongly with hypertension, mother lifespan uniquely correlates with lower childhood height, longevity uniquely lacks a negative correlation with chronic knee pain, and frailty correlates more strongly with hearing aid usage, daytime napping and allergic disease (eczema/dermatitis and hayfever/rhinitis) (Supplementary Data 2).

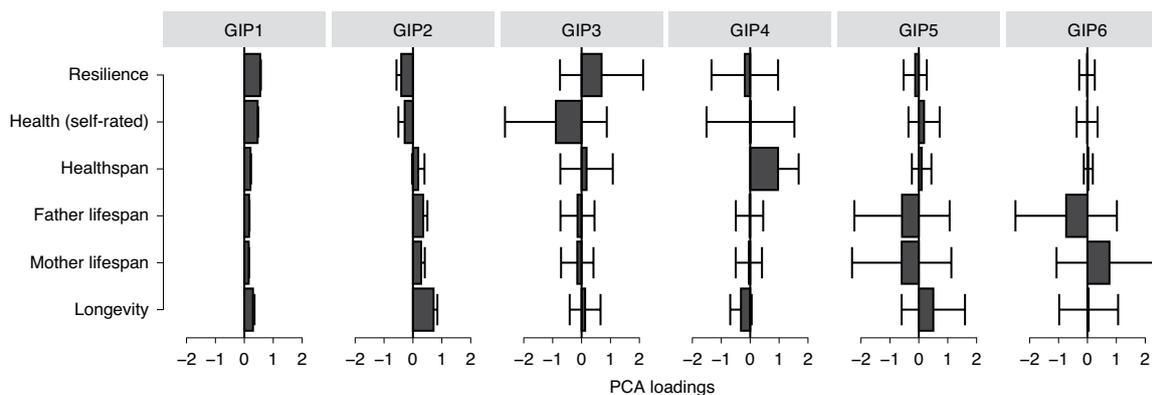
**Aging-GIPs capture distinct elements of wellbeing.** We combined the six GWASs in the main correlation cluster using the loadings from the principal components of the genetic correla-



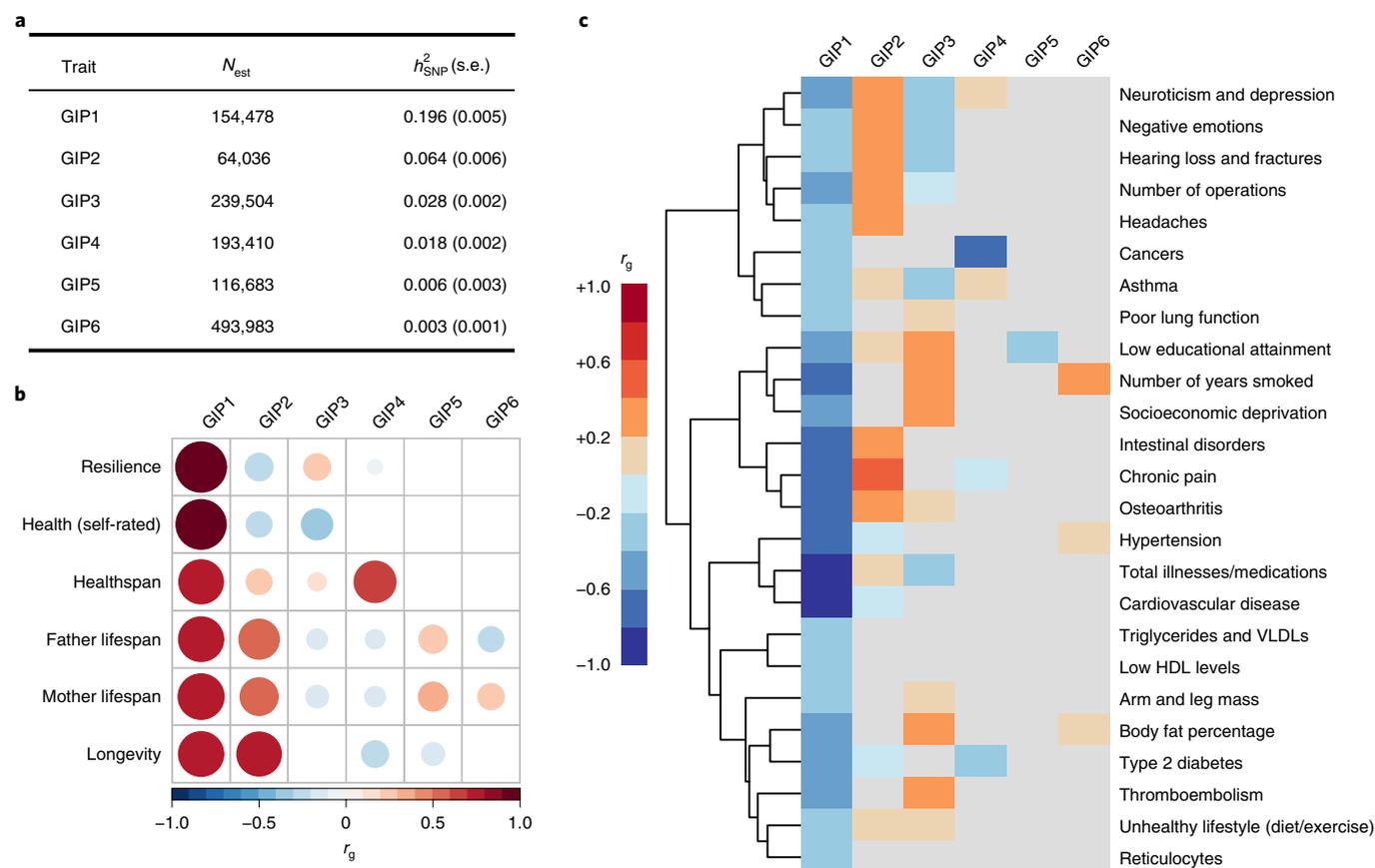
**Fig. 1 | There are strong genetic correlations between measures of the length and quality of life.** Diagonal values show the observed-scale SNP heritability of each aging-related phenotype and off-diagonal circles show the genetic correlation among traits, with both values calculated using HDL<sup>21</sup>. Blank squares did not pass multiple testing corrections (FDR  $> 5\%$ ). The bottom dendrogram shows the hierarchical relationship between traits, based on the magnitude of their genetic correlations. The black box highlights the first cluster of aging-related phenotypes used in follow-up analyses. EAA, epigenetic age acceleration; mLOY, mosaic loss of Y chromosome.

tion matrix (Fig. 2), yielding association summary statistics for six aging-GIPs (available at <https://doi.org/10.7488/ds/2972>). As expected, high-definition likelihood (HDL)<sup>21</sup> estimated aging-GIP1 to be the most heritable of the aging-GIPs ( $h^2_{\text{SNP}} = 0.20$ ; standard error (s.e.) = 0.005), capturing  $> 70\%$  of the genetics of healthspan, parental lifespan and longevity, and 90% of the genetics of self-rated health and the inverse of frailty (henceforth referred to as ‘resilience’) (Fig. 3). A leave-one-out analysis confirms that the GIP analysis is highly robust to the selection of GWAS, with the genetic architecture of aging-GIP1 remaining largely the same, after excluding any one of the six chronological and holistic aging-related trait GWASs (estimates range from  $r_g = 0.950$  (s.e. = 0.024) when excluding resilience, to  $r_g = 0.996$  (0.023) when excluding healthspan; Supplementary Table 2).

Apart from the genetic correlations with its component traits, aging-GIP1 shows significant correlations, with 459 of 729 traits tested ( $P_{\text{adj}} < 0.05$ , Bonferroni-adjusted for 6 GIPs and 729 traits). These suggest that, in addition to length of life, aging-GIP1 also captures mental and physical wellbeing. For example, among the largest negative aging-GIP1 correlations are mental illness, taking medications and diseases of old age (such as cardiometabolic disorders, cancers and osteoarthritis;  $r_g \leq -0.5$ ;  $P_{\text{adj}} < 1 \times 10^{-9}$ ). Inversely, fitness and education are among the largest positive correlations ( $r_g \geq 0.5$ ;  $P_{\text{adj}} < 1 \times 10^{-10}$ ). Aging-GIP1 also shows moderate negative correlations with infectious diseases, including N39 (International Classification of Diseases, 10th revision<sup>22</sup>) urinary tract infections ( $r_g = -0.66$ ; 95% CI =  $-0.42$  to  $-0.90$ ;  $P_{\text{adj}} = 3 \times 10^{-4}$ ), coughing on most days ( $r_g = -0.39$ ; 95% CI =  $-0.30$



**Fig. 2 | Principal component loadings used to create the genetically independent phenotypes for human aging.** Resilience is the inverse of the frailty index. The error bars represent 95% CIs of the loading estimate, inferred using 1,000 rounds of Monte Carlo simulations (see Supplementary Note for details). GIP, genetically independent phenotype; PCA, principal component analysis.



**Fig. 3 | Phenome-wide genetic correlations show that aging-GIP1 captures both physical and mental wellbeing.** **a**, Table showing the estimated sample size ( $N_{est}$ ) and SNP heritability ( $h^2_{SNP}$ ) of each aging-GIP GWAS, with s.e. in parentheses. **b**, Genetic correlations between aging-GIP GWASs and the six aging-related trait GWASs used to construct them. Blank values failed to pass nominal significance ( $P \geq 0.05$ ). **c**, GWASs of 402 phenotypes measured in European-ancestry individuals from the GWAS-MAP platform with strong and significant associations with GIP1 ( $P < 1.5 \times 10^{-5}$  and  $|r_g| > 0.25$ ). These 402 traits are summarized here in 25 groups, based on hierarchical clustering of the magnitude of pairwise genetic correlations. Each group is manually annotated with a label describing the traits within the cluster, and displays the values of the most informative trait (that is, the trait with the highest total z-score across aging-GIPs). Values failing to pass nominal significance ( $P \geq 0.05$ ) are grayed out. The dendrogram displayed here shows the hierarchical relationship between the most informative trait in each cluster. HDL, high-density lipoprotein; VLDL, very-low-density lipoprotein. See Supplementary Data 3 for the full list of correlations. Genetic correlation  $P$  values are derived from two-sided Wald’s test with one degree of freedom.

to  $-0.49$ ;  $P_{adj} = 1 \times 10^{-11}$ ) and severe COVID-19 hospitalization (that is, resulting in respiratory support or death;  $r_g = -0.33$ ; 95% CI =  $-0.20$  to  $-0.46$ ;  $P_{adj} = 0.004$ ).

However, aging-GIP1 also retains some strong correlations with socioeconomic factors, such as smoking behaviors (for example, current tobacco smoking:  $r_g = -0.50$ ; 95% CI =  $-0.45$  to  $-0.55$ ;

$P_{\text{adj}} = 4 \times 10^{-92}$ ) and having a job involving manual or physical work ( $r_g = -0.49$ ; 95% CI =  $-0.44$  to  $-0.54$ ;  $P_{\text{adj}} = 2 \times 10^{-89}$ ; Fig. 3 and Supplementary Data 3). Although such socioeconomic risk factors are partly heritable<sup>23</sup>, they could also reflect social stratification or geographic confounding shared between the original six studies used to construct aging-GIP1 (refs. 24,25). Therefore, as a sensitivity analysis, we used the GIP methodology to remove all genetic correlations with UK Biobank GWASs of household income and socioeconomic deprivation from the aging-GIPs—recognizing that this may remove some of the true signal as well (Extended Data Fig. 2). The SNP heritability of the adjusted aging-GIP1 phenotype is substantially attenuated ( $h^2_{\text{SNP}} = 0.09$ ; s.e. = 0.003) and, consequently, genetic correlations with all phenotypes are reduced. However, most traits that were significantly associated at 5% FDR remain associated with the residual phenotype (Extended Data Fig. 3). As expected, the largest reductions in genetic correlations are with socioeconomic, educational and smoking/drinking traits, which are reduced by  $\geq 0.30$  (Supplementary Data 3).

The remaining aging-GIPs have much lower heritability ( $h^2_{\text{SNP}} < 0.065$ ) and are of less interest to the present study because they capture genetic variance shared between fewer traits and do so with higher degrees of uncertainty. In short, aging-GIP2 correlates mainly with nonlethal causes of poor self-rated health, such as chronic pain (neck, back, stomach), negative emotions and hearing problems (59 traits;  $P_{\text{adj}} < 0.05$ ); aging-GIP3 correlates mainly with measures of socioeconomic deprivation and body composition (73 traits); and aging-GIP4 captures healthspan-specific correlations with cancer and diabetes (7 traits). Although aging-GIP5 and aging-GIP6 are underpowered due to their low heritabilities, they appear to distinguish parental lifespan from longevity (possibly through educational attainment), and between father lifespan and mother lifespan (possibly through risk taking and cardiovascular factors), respectively (Fig. 3).

**Characterizing the genomics of aging-GIP1.** Across the genome, 27 loci in the aging-GIP1 GWAS pass the genome-wide significance threshold, Bonferroni's adjusted for 6 GIPs ( $P < 8.3 \times 10^{-9}$ ). The strongest lead SNPs in these loci are rs429358 ( $P = 2 \times 10^{-40}$ ) and rs660895 ( $P = 2 \times 10^{-22}$ ), located nearest to the *APOE* and *HLA-DRB1/DQA1* genes, respectively (Fig. 4 and Table 1). Genetic fine-mapping with SuSiE<sup>26</sup> highlights three independent 95% credible sets of causal variants for the *APOE* locus: the *APOE*  $\epsilon 4$  and  $\epsilon 2$  alleles at single variant resolution, and a set of three variants intronic or near to *APOC1*. Three other loci have a credible set containing only one variant. These variants are rs9271300 (intergenic between *HLA-DRB1* and *HLA-DQA1*), rs34811474 (a nonsynonymous *ANAPC4* exon variant) and rs2165702 (an intergenic variant near *PHB*; Supplementary Table 3 and Supplementary Data 4).

Most lead SNPs are strongly associated with self-rated health and resilience, in line with the large loadings of these traits in the construction of aging-GIP1 (Fig. 4 and Supplementary Table 4). Seventeen aging-GIP1 loci overlap with genome-wide significant ( $P < 5 \times 10^{-8}$ ) loci from the original aging-related trait GWAS, showing evidence of a shared causal variant (coloc posterior probability (PP)  $\geq 80\%$ ; Supplementary Table 5). For the other loci, the GIP framework either increases power ( $n = 4$ ) or indicates that there may be a different causal variant (coloc PP  $< 80\%$ ;  $n = 6$ ; Supplementary Data 5).

Loci near *APOE*, *HLA-DRB1/DQA1*, *LPA* and *CDKN2B/-AS1* have previously been validated using the same trait in an external cohort<sup>2-4</sup>. For the remaining 23 loci, we measured lead SNP effects on participant survival in FinnGen (release 5;  $n = 203,244$ ; 6.94% deceased) and BioBank Japan ( $n = 135,983$ ; 24.1% deceased), to provide additional evidence of their association with human aging traits in independent samples. Combining both cohorts to achieve adequate power, we find that aging-GIP1-increasing alleles of lead

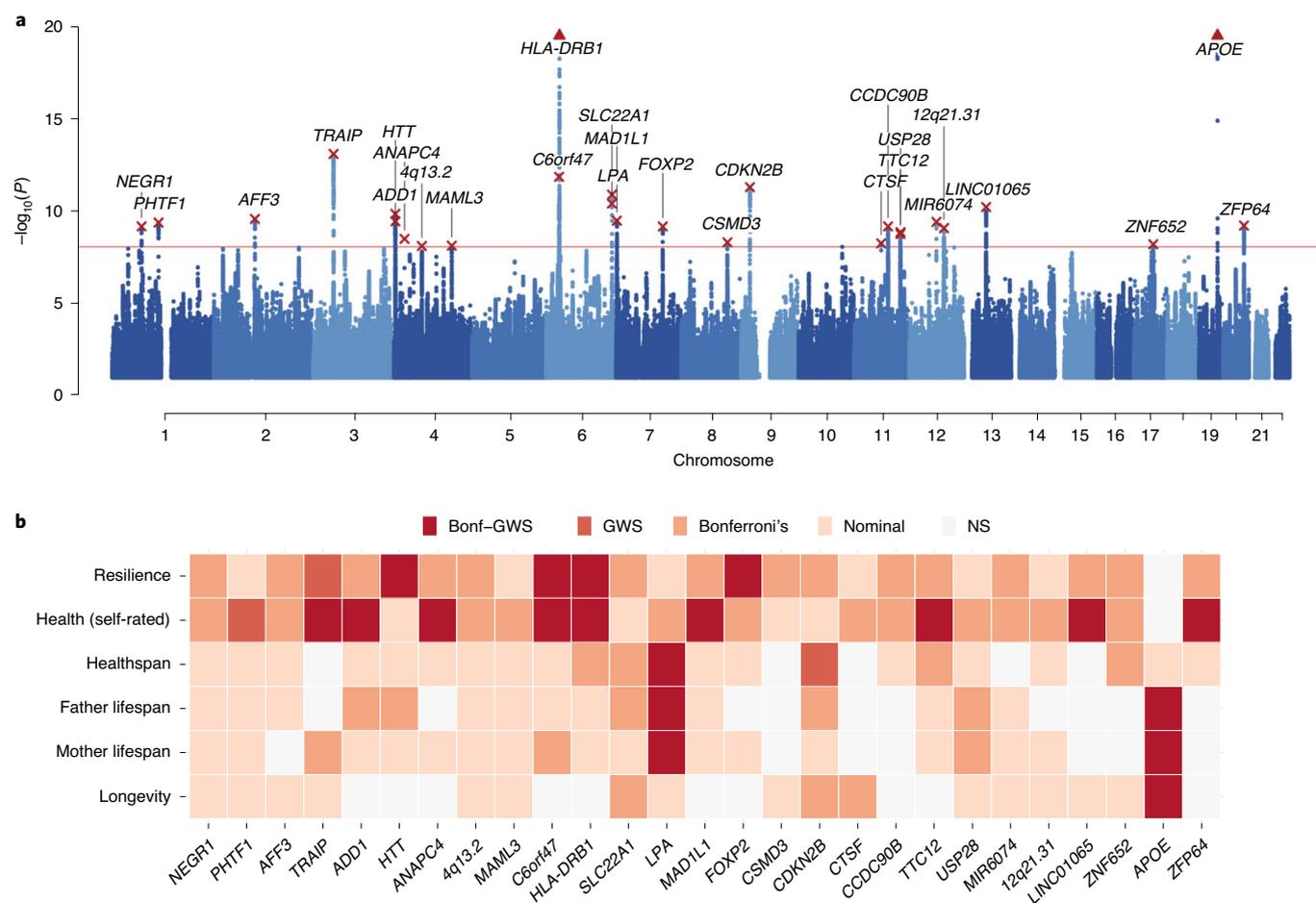
SNPs near *HTT* and *MAML3* have a protective effect on survival in these cohorts (one-sided  $P < 0.0022$ , that is, taking into account the 23 loci tested), increasing average lifespan by around 2.53 (95% CI = 0.91–4.15; one-sided  $P = 0.001$ ) and 2.51 (95% CI = 0.79–4.23; one-sided  $P = 0.002$ ) months per allele, respectively. Although we are underpowered to confirm the remaining 21 loci individually, we find that, collectively, their aging-GIP1-increasing alleles are also associated with increased Finnish and Japanese survival (one-sided  $P = 0.044$ ; Supplementary Table 6).

Again, we find aging-GIP1 genetics are stable when performing a leave-one-out analysis: lead SNP effects of most aging-GIP1 loci do not change significantly when excluding one of the six aging traits. The exceptions are for previously replicated loci and *SLC22A1/A2*, where exclusion of one of the traits can reduce the aging-GIP1 effect size, although loci near *HLA-DRB1/DQA1*, *LPA* and *CDKN2B/-AS1* remain genome-wide significant in all leave-one-out scenarios (Supplementary Data 6). Similarly, when completely removing genetic correlations with household income and socioeconomic deprivation, we find that aging-GIP1 effect sizes are attenuated but, in all cases, remain associated with the residual trait ( $P < 0.0019$ , that is, taking into account the 27 loci tested; Extended Data Fig. 4 and Supplementary Data 7).

We further looked up all lead aging-GIP1 SNPs and close proxies ( $r^2_{\text{EUR}} \geq 0.8$ ) in PhenoScanner<sup>27</sup> and the GWAS catalog<sup>28</sup>, excluding associations discovered solely in UK Biobank, which showed that 24 of 27 aging-GIP1 loci had previously been associated with one or more traits at genome-wide significance. Most of these loci were associated with cardiometabolic, immune-related or neuropsychiatric disorders, although several were also associated with measures of educational attainment and household income. Of specific interest are loci near *APOE*, *HLA-DRB1/DQA1*, *CDKN2B/-AS1* and *ZNF652/PHB*, which show aging-GIP1-increasing alleles are associated with a reduction in multiple diseases, but do not appear to associate with socioeconomic factors, suggesting that these loci largely capture intrinsic sources of aging (Supplementary Data 8).

Aggregating SNP association statistics across the genome into gene scores using the Pathway Scoring Algorithm (PASCAL)<sup>29</sup>, we find that high-scoring genes for aging-GIP1 (Supplementary Data 9) appear to be overrepresented in the heme metabolism Hallmark gene set, as well as 300 gene ontology (GO) pathways (FDR = 5%), regardless of adjustment for socioeconomic correlations. These GO pathways cluster into 20 groups, related to: neuronal development, organization and function; transcriptional regulation; chemical homeostasis; cellular growth, differentiation and apoptosis; proteolysis; protein phosphorylation; intracellular signaling and transport; immune system development; the muscle system; and lipoprotein metabolism (Supplementary Data 10). Similarly, aging-GIP1 heritability appears to be enriched in genomic regions containing histone marks associated with the central nervous system (Supplementary Table 7).

**Causal inference of blood protein levels on aging-GIP1.** Mendelian randomization (MR) uses genetic variation as instrumental variables to estimate the causal effect of exposures of interest on an outcome<sup>30</sup>. We used a set of well-validated blood protein quantitative trait loci (pQTLs) for 857 proteins<sup>30</sup> as genetic instruments in a two-sample MR<sup>31</sup> and colocalization<sup>32</sup> framework to infer putative causal links between protein levels and aging-GIP1 (Supplementary Data 11). We find robust evidence for a detrimental effect on aging-GIP1 (FDR  $< 5\%$ ) for the levels of four proteins in blood (Table 2), with pQTL instruments passing sensitivity and causal directionality tests, and additionally colocalizing with the aging-GIP1 signal (Methods). Three of these proteins—apolipoprotein(a) (LPA), olfactomedin-1 (OLFM1) and low-density lipoprotein (LDL) receptor-related protein 12 (LRP12)—were instrumented by a *cis*-pQTL and encoded by genes that appeared significantly enriched in the



**Fig. 4 | Twenty-seven independent genomic loci are associated with the shared genetic component of aging-related GWASs. a**, Manhattan plot showing associations of SNPs across the genome with aging-GIP1, with the y axis showing the strength of the association and the x axis the genomic position of the SNP. The line represents the genome-wide significance threshold, Bonferroni's adjusted for 6 GIPs ( $P=8.3 \times 10^{-9}$ ). The lead SNP of each independent locus is marked with a cross and annotated with the nearest (upstream) gene, or the cytogenetic band if there are no genes within 250 kb. The y axis has been capped at  $P=1 \times 10^{-20}$ . The following loci exceed this cap and are represented as triangles: *APOE* ( $P=1.5 \times 10^{-40}$ ) and *HLA-DRB1* ( $P=1.4 \times 10^{-22}$ ). **b**, Significance of the association of each lead SNP with the aging-related trait GWASs used to construct aging-GIP1. Bonf-GWS, Bonferroni's adjusted genome-wide significance ( $P < 8.3 \times 10^{-9}$ ); GWS, genome-wide significance ( $P < 5 \times 10^{-8}$ ); Bonferroni's,  $P < 0.05$  adjusted for 6 traits and 27 SNPs ( $P < 3 \times 10^{-4}$ ); nominal,  $P < 0.05$ ; NS, not significant. *P* values are derived from two-sided Wald's test with one degree of freedom.

gene score analysis. The remaining protein, vascular cell adhesion molecule 1 (VCAM1), was instrumented by a *trans*-pQTL shared with  $\beta_2$ -microglobulin ( $\beta_2$ M); however, only VCAM1 protein levels colocalized with the signal at this locus (Supplementary Data 11). When performing the same MR analysis on aging-GIP1 adjusted for household income and socioeconomic deprivation, protein effects remained significant (Extended Data Fig. 5).

Among the discovered proteins, LPA shows the most significant effect on aging-GIP1 ( $P_{MR}=2 \times 10^{-8}$ ), with an increase of 1 s.d. in genetically predicted blood protein levels, causing a decrease of 0.035 (95% CI=0.025–0.045) s.d. in aging-GIP1. This significance appears to be driven by a consistent detrimental effect across all six aging-GIP1 component traits ( $\beta_{MR}$  range 0.013–0.035; all nominal  $P < 0.05$ ; Fig. 5). For a sense of scale, when performing the same MR analysis on the unstandardized parental lifespan GWAS, this equates to a loss of approximately 7 months of life (95% CI=5–9 months) per s.d. increase in LPA blood levels (Supplementary Table 8). The effects of the remaining proteins on aging-GIP1 are larger in magnitude but appear unequally distributed across aging-GIP1 component traits. We estimate that a decrease of 1 s.d. in genetically predicted VCAM1 blood protein levels is associated with an increase of approximately 18 months of life (13–23 months);

however, comparing VCAM1 effects on standardized aging-GIP1 components, we find its effect to be largest on mid-to-late-life aging traits (lifespan, longevity, resilience) compared with early life aging traits (healthspan). Similarly, the effects of OLFM1 and LRP12 proteins appear mostly mediated through late-life aging traits (resilience and/or longevity; Fig. 5).

## Discussion

We combined European-ancestry GWASs of healthspan, father lifespan, mother lifespan, longevity, frailty and self-rated health using a framework that maximized power to detect associations with their shared genetic component while downweighting trait-specific genetic associations. The resulting aging-GIP1 trait captured the genetics underlying physical and mental wellbeing, and showed strong inverse genetic correlations with cardiovascular, inflammatory and neuropsychiatric disease traits. We highlight 27 loci with genome-wide significant effects on aging-GIP1, including two new loci near *HTT* and *MAML3* which showed directionally consistent evidence of an effect on survival in two additional, independent samples. Across the genome, we found genes enriched for association with aging-GIP1 to be overrepresented in heme metabolism and pathways related to (among others) neurogenesis, homeostasis,

**Table 1 | Twenty-seven independent genomic loci associated with the first genetic principal component of aging-related trait GWASs (aging-GIP1)**

Nearest gene(s)	rsID	Chromosome	Position	A1	Freq1	$\beta$ 1 (s.e.)	P	Het
<i>NEGR1</i>	rs2815748	1	72816147	G	0.20	0.028 (0.004)	$7 \times 10^{-10}$	
<i>PHTF1</i>	rs1230682	1	114293526	A	0.63	0.023 (0.004)	$4 \times 10^{-10}$	
<i>AFF3</i>	rs7609078	2	100490363	A	0.37	0.023 (0.004)	$3 \times 10^{-10}$	
<i>TRAIP</i>	rs2271961	3	49878113	T	0.50	0.027 (0.004)	$8 \times 10^{-14}$	
<i>ADD1</i>	rs16843603	4	2928577	C	0.28	0.025 (0.004)	$1 \times 10^{-10}$	
<i>HTT</i>	rs362273	4	3227419	G	0.32	0.024 (0.004)	$4 \times 10^{-10}$	
<i>ANAPC4</i>	rs34811474	4	25408838	A	0.23	0.025 (0.004)	$3 \times 10^{-9}$	
4q13.2	rs10434248	4	67842921	A	0.57	0.021 (0.004)	$8 \times 10^{-9}$	
<i>MAML3</i>	rs56172573	4	140919381	C	0.38	0.021 (0.004)	$7 \times 10^{-9}$	
<i>C6orf47/GPANK1</i>	rs805262	6	31628733	C	0.52	0.027 (0.004)	$1 \times 10^{-12}$	
<i>HLA-DRB1/DQA1</i>	rs660895	6	32577380	A	0.79	0.043 (0.004)	$1 \times 10^{-22}$	*
<i>SLC22A1/A2</i>	rs9456508	6	160598596	T	0.98	0.082 (0.012)	$4 \times 10^{-11}$	**
<i>LPA</i>	rs118039278	6	160985526	G	0.92	0.045 (0.007)	$1 \times 10^{-11}$	***
<i>MAD1L1/SNORA114</i>	rs11764780	7	2020904	C	0.19	0.029 (0.005)	$3 \times 10^{-10}$	
<i>FOXP2</i>	rs12705966	7	114248851	G	0.34	0.023 (0.004)	$7 \times 10^{-10}$	
<i>CSMD3</i>	rs560719	8	113032374	T	0.50	0.021 (0.004)	$5 \times 10^{-9}$	
<i>CDKN2B/-AS1</i>	rs9632885	9	22072638	G	0.52	0.025 (0.004)	$5 \times 10^{-12}$	**
<i>CTSF</i>	rs2924807	11	66341005	G	0.50	0.021 (0.004)	$5 \times 10^{-9}$	
<i>CCDC90B/DLG2</i>	rs2512690	11	83145469	C	0.34	0.023 (0.004)	$7 \times 10^{-10}$	
<i>TTC12/ANKK1</i>	rs2186800	11	113242860	A	0.54	0.022 (0.004)	$2 \times 10^{-9}$	
<i>USP28/HTR3B</i>	rs61907878	11	113751052	C	0.90	0.037 (0.006)	$1 \times 10^{-9}$	
<i>MIR6074</i>	rs7306710	12	66376091	T	0.48	0.023 (0.004)	$4 \times 10^{-10}$	
12q21.31	rs6539846	12	84811217	A	0.49	0.022 (0.004)	$8 \times 10^{-10}$	
<i>LINC01065</i>	rs8002970	13	53924489	C	0.45	0.024 (0.004)	$6 \times 10^{-11}$	
<i>ZNF652/PHB</i>	rs28394864	17	47450775	G	0.54	0.021 (0.004)	$6 \times 10^{-9}$	
<i>APOE</i>	rs429358	19	45411941	T	0.84	0.066 (0.005)	$2 \times 10^{-40}$	***
<i>ZFP64</i>	rs67442863	20	51031131	T	0.18	0.029 (0.005)	$6 \times 10^{-10}$	

Loci were defined as 500-kb regions centered on a lead genome-wide significant SNP ( $P < 8.3 \times 10^{-9}$ ) in linkage equilibrium ( $r^2_{\text{EUR}} < 0.1$ ) with other lead locus SNPs. Nearest gene(s), closest genes upstream/downstream to the lead SNP (within 250 kb) or, if none, the closest cytogenetic band; rsID, the lead SNP within the locus; Position, base-pair position (GRCh37); A1, effect allele, associated with higher aging-GIP1; Freq1, allele frequency of the effect allele in UK Biobank;  $\beta$ 1, effect estimate (and s.e.) of the A1 allele on aging-GIP1 in s.d. units; P, two-sided nominal P value from Wald's test; Het, evidence of heterogeneity; asterisks indicate that aging-GIP1 effect size changes significantly when leaving out one of the core aging traits from the GIP1 calculation (\*all but one leave-one-out effects are the same; \*\*all but two effects are the same; \*\*\* $\leq 3$  effects are the same).

**Table 2 | MR of genetically predicted blood levels of four proteins suggesting that they have a causal detrimental effect on aging-GIP1**

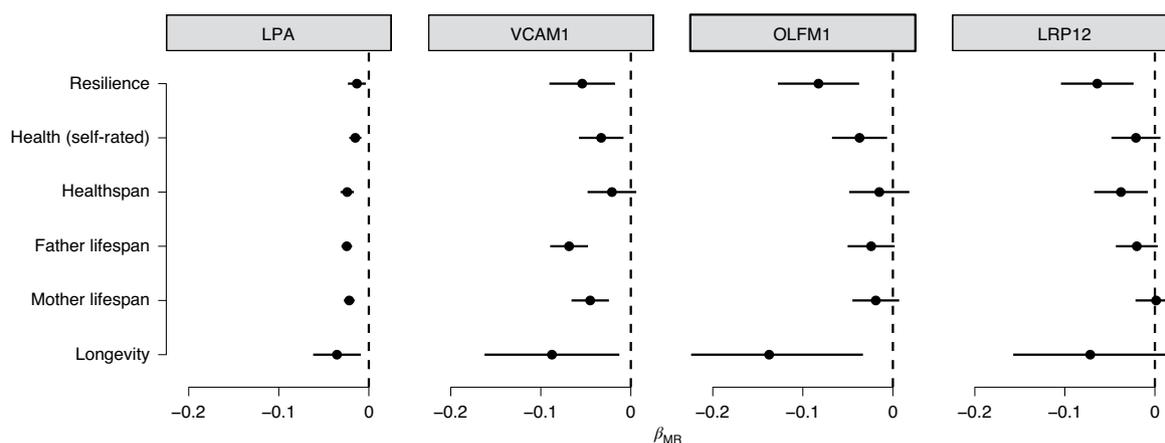
Exposure	$\beta_{\text{MR}}$ (s.e.)	P	FDR	$P_{\text{steiger}}$	$P_{\text{reverse}}$	Coloc PP (%)	LD check
<b>LPA</b>	<b>-0.035 (0.005)</b>	<b><math>2 \times 10^{-11}</math></b>	<b><math>2 \times 10^{-8}</math></b>	<b><math>6 \times 10^{-274}</math></b>	<b>0.556</b>	<b>100</b>	<b>Pass</b>
<b>VCAM1</b>	<b>-0.095 (0.019)</b>	<b><math>7 \times 10^{-7}</math></b>	<b><math>2 \times 10^{-4}</math></b>	<b><math>9 \times 10^{-12}</math></b>	<b>0.063</b>	<b>99</b>	<b>Pass</b>
OLFM1	-0.111 (0.024)	$3 \times 10^{-6}$	$6 \times 10^{-4}$	$1 \times 10^{-7}$	-	-	Pass
LRP12	-0.078 (0.021)	$2 \times 10^{-4}$	0.033	$4 \times 10^{-10}$	-	-	Pass

In bold are exposures passing all quality checks, including reverse MR and coloc, which required access to genome-wide summary statistics.  $\beta_{\text{MR}}$ , MR effect with s.e. in parentheses. P, two-sided nominal Wald's test P value for the MR effect. FDR, FDR-adjusted P value taking into account the 857 proteins tested.  $P_{\text{steiger}}$ , two-sided nominal Wald's test P value for the MR-Steiger test assessing whether the exposure-outcome pair has the correct causal direction.  $P_{\text{reverse}}$ , two-sided nominal Wald's test P value for the MR effect of aging-GIP1 on the exposure, that is, evidence of reverse causality. Coloc PP, posterior probability of colocalization estimated by coloc. LD check, a secondary check for colocalization, which requires at least 1 of the 30 strongest aging-GIP1 SNPs within 500 kb of the pQTL to be in strong LD ( $r^2_{\text{EUR}} \geq 0.8$ ) with the pQTL itself.

proteolysis, immunity and the muscle system in human aging. Last, we performed MR of predicted blood protein levels on aging-GIP1, which revealed that the levels of LPA, VCAM1, OLFM1 and LRP12 proteins may be detrimental to multiple indices of healthy aging.

LPA is a glycoprotein making up the main component of large lipoprotein(a) particles. It is a well-known risk factor for atherosclerotic disease<sup>33,34</sup> and a target of ongoing clinical trials with regard

to cardiovascular outcomes<sup>35</sup>. A recent MR study used 27 genetic instruments for LPA and found a link between genetically elevated LPA levels and reduced healthspan and parental lifespan<sup>36</sup>, in line with our results. Our analysis suggested that the detrimental effect of LPA may apply to aging more generally, and stringent colocalization and reverse MR tests provided additional evidence for causality. It is interesting that human endothelial cell culture experiments



**Fig. 5 | MR of blood protein expression levels on aging-related GWASs.** Exposures that showed causal evidence of an effect on aging-GIP1 were tested for association with each of the aging-related traits used to construct aging-GIP1. The x axis shows the MR effect estimate, with lines representing 95% CIs, as estimated by TwoSampleMR<sup>3</sup>. The sample sizes of exposure GWASs were 6,861 for LPA, 3,301 for VCAM1, and 3,200 for OLFM1 and LRP12.

show that the addition of LPA increases cell-surface expression of VCAM1 (ref. <sup>37</sup>) and can increase endothelial cell contraction and permeability<sup>38</sup>.

VCAM1 is a cell adhesion glycoprotein localized predominantly on endothelial cell surfaces and its expression is upregulated in response to inflammatory signals, which mediate adhesion and transduction of leukocytes across endothelial walls<sup>39</sup>. The link that we established between VCAM1 and human aging relied on a *trans*-pQTL instrument shared with  $\beta_2M$  and is therefore more susceptible to horizontal pleiotropy, that is, the genetic variant may influence VCAM1 levels indirectly, and its effect on human aging traits could be caused by factors independent of VCAM1 levels. However, only VCAM1 colocalized with the pQTL signal, and experimental evidence from mouse studies suggests that the effect of VCAM1 on aging is likely to be causal. Specifically, VCAM1 levels in blood are known to increase with age in both humans and mice<sup>40</sup>, and treatment with anti-VCAM1 antibodies or an inducible deletion of *Vcam1* improves cognitive performance of aged mice<sup>40</sup>. Of note, similar results have been found for  $\beta_2M$  abundance and mouse knockouts<sup>41</sup> and, as such, identification of robust genetic instruments for  $\beta_2M$  levels is also warranted.

We were unable to robustly assess colocalization and reverse causality for the LRP12 and OLFM1 signals because genome-wide association summary statistics were not available, so the effects of these proteins on human aging should be interpreted with additional caution. LRP12 belongs to the LDL-receptor superfamily and may play a role in brain development<sup>42</sup> and both tumor proliferation and suppression, depending on the tissue<sup>43,44</sup>. Similarly, OLFM1 is a glycoprotein involved in neuronal development and maintenance<sup>45</sup>, which has also been shown to suppress colorectal tumor metastasis<sup>46</sup>. In mouse models, *Olfm1* knockouts showed reduced cerebral infarction and fertility<sup>47</sup>. However, it is unclear whether reduction of LRP12 or OLFM1 blood levels will have a beneficial effect in humans.

Importantly, our MR analysis was restricted to blood pQTLs, precluding detection of causal effects of protein levels in other tissues. Although it is likely that there are proteins with tissue-specific effects on aging, particularly in the brain, the samples needed for detection of such pQTLs are less readily available and therefore more difficult to study at scale (although progress is being made<sup>48</sup>). Still, blood may be a particularly suitable tissue to identify aging-related proteins. Connecting the circulatory systems of two mice of different ages has been shown to accelerate signs of aging in the brain, muscle and liver of the young mouse, and reverse similar signs in the old mouse<sup>49,50</sup>. Likewise, the detrimental cognitive

effect of injecting old blood in mice is counteracted when anti-VCAM1 antibodies are concomitantly injected<sup>40</sup>. As such, the blood currently remains one of the most promising tissues to detect aging-related proteins.

Heme metabolism and iron levels were previously hypothesized to play a role in human aging<sup>11</sup>, and in the present study we identify the same pathway using new methods and additional data. It is of interest that both the heme pathway and the proteins we uncovered using MR are strongly linked to vascular and endothelial damage<sup>51</sup>. Across the genome, we also found an enrichment for brain tissues and pathways related to neuronal integrity. Given that endothelial cells are central to both the cardiovascular system and the blood-brain barrier<sup>52,53</sup>, and that endothelial function declines with age<sup>54</sup>, progressive endothelial dysfunction may manifest itself as an age-related disease. Indeed, recent findings suggest that the detrimental effects of the *APOE* *e4* allele—the largest genetic determinant of human aging—are mediated by an accelerated breakdown of the blood-brain barrier, independent of amyloid- $\beta$  and tau accumulation<sup>55</sup>. We therefore speculate that molecules involved in maintaining or repairing endothelial integrity may be key to avoiding both age-related cardiovascular injury and neurodegeneration, and recommend further research into this area.

Our study demonstrates that GIP analysis of genetically correlated GWASs can increase power to detect shared genetic architecture and can identify genetic loci that would have been missed by any individual GWAS. For example, regions near *AFF3*, *MAML3*, *USP28/HTR3B* and *MIR6074* reached genome-wide significance only in the combined analysis, and validation of *MAML3* in an external survival cohort suggests that these findings may hold across populations and aging measurements. In addition, *USP28* within the *USP28/HTR3B* locus resembles *CDKN2A* within the well-known *CDKN2B/-AS1* locus, because both have been implicated in cellular proliferation<sup>56,57</sup> and senescence<sup>58,59</sup>. In fact, short hairpin RNA knockdown of *USP28* in vitro results in decreased *CDKN2A* expression<sup>58</sup>. However, additional validation and functional investigation of aging-GIP1 loci are warranted.

A secondary advantage of the GIP method is the high stability of the resultant aging-GIP1 GWAS, which appears to be largely robust to the selection of component traits. Nevertheless, we note that our analysis focused on chronological and disease measures of aging and did not include the cellular measures of aging due to their modest genetic correlations and lack of power. Whether aging-GIP1 loci and pathways accurately reflect the aging process overall, or only capture specific domains of aging, is unknown. Inclusion of future

aging-related GWASs performed on Biobank-scale samples should provide insight into how well aging-GIP1 captures human aging.

An important limitation of the aging-GIP1 GWASs, and GWASs of aging-related traits more generally, is the potential confounding of aging genetics with social stratification. Socioeconomic factors exhibiting geographical clustering can induce gene–environment correlations that could inflate trait heritability and may bias results, especially for UK Biobank studies<sup>25</sup>. In our study, we found that genetic correlations of the six aging-related GWASs with socioeconomic deprivation were moderately high (up to 61% for self-rated health). This confounding can be inadvertently propagated and even amplified in the shared genetic component captured by aging-GIP1. Indeed, when explicitly removing genetic correlations with two UK Biobank GWASs of socioeconomic factors, aging-GIP1 heritability is halved. After adjustment, genetic correlations with mental and physical health traits are also somewhat reduced, although the overall patterns remain stable. Similarly, effect sizes of blood protein levels and aging-GIP1 loci—including loci that have been confidently linked to aging-related traits previously—were only slightly reduced after adjustment.

The shift of these effect sizes after adjustment was almost equal for all SNPs and MR signals. Together with the highly decreased heritability and stable genetic correlations, this suggests that the adjustment removed the unspecific polygenic background induced by UK Biobank microstructure and social confounding. We therefore consider this procedure to be a suitable approach for confounder correction, although it may be somewhat conservative for also removing the genetic background of the adjusting traits, which may genuinely be associated with aging. An alternative solution to the confounding problem could be the inclusion of a more ancestrally diverse set of aging-related GWASs from a larger variety of populations.

Despite these limitations, modeling the shared genetic component of human aging proxies has allowed us to downweigh nonbiological features, and propose pathways and proteins that may causally influence the human aging process. We share the full aging-GIP1 summary statistics without restrictions<sup>60</sup> to encourage further MR analysis using other biomarkers, and accelerate the discovery of drug targets able to prolong mental and physical wellbeing throughout life.

## Methods

**Data sources.** We searched PubMed and Google Scholar in April 2020 for GWASs of aging measures, including only studies for which we could obtain autosomal genome-wide summary statistics measured in at least 10,000 European-ancestry individuals. If multiple studies were performed on similar traits, we kept the study with the largest sample size. GWASs meeting inclusion criteria included extreme longevity<sup>17</sup> (survival past 90th percentile), father and mother lifespan<sup>4</sup>, healthspan<sup>61</sup>, self-reported health<sup>16</sup>, frailty index<sup>8</sup>, epigenetic age acceleration<sup>19</sup>, telomere length<sup>5</sup>, mosaic loss of the Y chromosome<sup>7</sup> and perceived age<sup>18</sup>. As the self-reported health GWAS used the first release of UK Biobank data ( $n \sim 150,000$ ), we looked up the same phenotype in the Neale Lab GWAS collection ( $n \sim 500,000$ ), which had a larger sample size but was otherwise measured identically<sup>62</sup>. The original derivation of each set of summary statistics is briefly described in the Supplementary Note.

For each set of summary statistics, we discarded poorly imputed (INFO < 40%), rare (minor allele frequency (MAF) < 0.5%) or poorly measured SNPs ( $n$  individuals < 1% of total). The remaining SNPs were aligned to genome build GRCh37 and harmonized to match UK Biobank SNP IDs (discarding any duplicates). We then estimated the phenotypic variance of the trait (residuals) from a selection of independent SNPs provided by the MultiABEL R package<sup>13</sup> using equation (1) from Winkler et al.<sup>63</sup>

$$\text{Var}(Y) = \text{Median} \left( 2pqn \times \text{Var}(\hat{\beta}) \right) \quad (1)$$

where  $\text{Var}(Y)$  is the phenotypic variance,  $p$  and  $q$  are the major and minor allele frequencies,  $n$  is the total sample size (cases + controls, if applicable) and  $\text{Var}(\hat{\beta})$  is the variance of the effect size estimate. If  $n$  differed by SNP, we calculated the phenotypic variance separately for quintiles of  $n$  and took the mean estimate. SNP statistics were then standardized by dividing effect sizes and s.e. by  $\sqrt{\text{Var}(Y)}$ .

**Estimation of genetic and nongenetic correlations.** The HDL R package<sup>21</sup> v.1.3.4 was used to calculate SNP heritabilities of aging-related GWASs and their genetic correlations with each other, using the default European-ancestry

linkage disequilibrium (LD) reference panel and non-major histocompatibility complex (MHC) SNPs with MAF  $\geq 0.01$ . All GWASs had  $\geq 99.9\%$  of the SNPs in this reference panel, except for the Hannum and Horvath epigenetic age acceleration GWASs, which had 96.29%. Pearson's correlations between GWASs due to phenotypic similarity and sample overlap were calculated from  $z$ -scores of independent SNPs provided by the MultiABEL R package<sup>13</sup> which were nonsignificant in both studies ( $|z| < 1.96$ ). The number of SNPs used for this calculation is reported in Supplementary Data 1.

LD-score (LDSC) regression<sup>64</sup> v.1.0.0 was used to calculate genetic correlations between selected GWAS summary statistics from the GWAS-MAP platform<sup>20</sup> and the aging-related GWAS and aging-GIP summary statistics used in our study, where the longer computation time of the HDL software would have been impractical. The GWAS-MAP platform contains summary statistics for 1,329,912 complex traits and gene expression levels, and 3,642 binary traits, derived from UK Biobank<sup>62,65</sup> and large European-ancestry consortia (for example, MAGIC, CARDIoGRAM, SSGAC and GIANT; for an up-to-date list of phenotypes, see <https://pheligo.com>)<sup>66</sup>. We included GWASs with  $\geq 1$  million SNPs, measured in  $\geq 10,000$  individuals (if continuous) or  $\geq 2,000$  cases and controls (if binary). We further excluded the healthspan and self-rated health GWASs from the GWAS-MAP platform to avoid duplication, after which 728 traits remained. Out of specific interest<sup>67</sup>, we also calculated the genetic correlation between aging-GIPs and a case–control GWAS of COVID-19 hospitalization, with cases defined as laboratory-confirmed COVID-19 patients experiencing a severe outcome and controls defined as the rest of the population (A2\_ALL excluding 23andMe, release 4)<sup>68</sup>. GWASs with counterintuitive effect directions were reversed (for example, satisfaction traits are coded from high to low in UK Biobank), and each set of GWAS summary statistics was filtered before analysis by excluding SNPs in the MHC, SNPs with MAF < 1% and, if measured, SNPs with INFO < 90%. A full list of 729 GWASs and references can be found in Supplementary Data 3.  $P$  values were adjusted for multiple testing using Bonferroni's correction (729 traits and 6 GIPs).

The same software was used to calculate pairwise correlations between GWAS-MAP statistics, which allowed traits to be clustered based on the magnitude of their genetic similarity. For computational tractability, we included only GWASs that showed large and significant effects on aging-GIP1 ( $|r_{g1}| > 0.25$ ;  $P_{adj} < 0.05$ ). Clusters were identified hierarchically by maximizing the Bayesian information criterion using the mclust R package<sup>69</sup> v.5.4.1, up to a maximum of 100 clusters.

**Identification of shared and unique genetic correlations.** For each selected GWAS-MAP phenotype, genetic correlations with aging-related traits were meta-analyzed using fixed-effect, inverse-variance weighting, with heterogeneity quantified by Cochran's  $Q$  and  $I^2$  statistics, implemented in the meta R package<sup>70</sup> v.4.15-1. GWAS-MAP phenotypes that were significantly correlated with all six aging-related GWAS at FDR of 5% and lacking substantial heterogeneity ( $P_{het} > 0.05$  and  $I^2 < 50\%$ ) were considered to be shared. For GWAS-MAP phenotypes with evidence of heterogeneity, we performed a leave-one-out sensitivity analysis to assess which aging-related trait(s) contributed most to this heterogeneity. If heterogeneity could be completely removed by excluding a single GWAS ( $I^2 = 0\%$ ), and exclusion of any other GWAS did not substantially reduce heterogeneity ( $I^2 \geq 50\%$ ), the aging-related trait outlier was considered to have a unique genetic correlation with the GWAS-MAP phenotype.

**Genetically independent phenotype analysis of aging GWAS.** Principal component loadings for six aging-GIPs were estimated from the genetic covariance matrix of the six chronological and holistic aging GWASs, analogous to a principal component analysis of phenotypic correlations. Specifically, eigenvectors from the genetic covariance matrix were transformed into loadings by dividing them by the square root of the phenotypic variance of the GIP. This phenotypic GIP variance was calculated as follows:

$$\text{Var}(\text{GIP}_i) = \sum [(\mathbf{a}_i \otimes \mathbf{a}_i) \circ \Sigma_{\text{ph}}] \quad (2)$$

where  $\mathbf{a}_i$  is the eigenvector of the  $i$ th aging-GIP and  $\Sigma_{\text{ph}}$  is the phenotypic variance–covariance matrix of the core aging traits. As GWASs were standardized,  $\Sigma_{\text{ph}}$  is equivalent to the phenotypic correlation matrix, with off-diagonal correlations estimated from the correlation observed between independent null  $z$  statistics (described above).

The six chronological and holistic aging GWAS statistics were then combined on an SNP-by-SNP basis using the principal component loadings to construct genome-wide summary statistics for the six aging-GIPs. GIP effect estimates were calculated by summing effect estimates from the individual aging-related trait GWAS, each multiplied by their corresponding loading, and standardized using the expected GIP variance (equation (2)). The s.e.s of the GIP effect estimate were calculated by performing the equivalent calculation using variance arithmetic, also taking into account the phenotypic covariance between GWASs to adjust for sample overlap:

$$\text{s.e.}(\hat{\beta}) = \sqrt{\sum [(\mathbf{a}_i \otimes \mathbf{a}_i) \circ \Sigma_{\text{ph}} \circ (\text{s.e.}(\mathbf{B}) \otimes \text{s.e.}(\mathbf{B}))]} \quad (3)$$

where  $\text{s.e.}(\mathbf{B})$  is the vector of SNP s.e.s of the core aging trait effect estimates. Effective sample sizes were then estimated based on the median  $z$  statistic and allele frequencies, that is, solving equation (1) for  $n$ . Full technical details of the GIP method are described in Supplementary Note.

Aging-GIP summary statistics were calculated for the 7,324,133 SNPs shared between quality-controlled GWAS statistics, of which 5,353,660 were common ( $\text{MAF} \geq 5\%$ ) and 1,970,474 were rare ( $\text{MAF} < 5\%$ ). Finally, s.e.s of each aging-GIP GWAS were adjusted to account for the LDSC regression intercept, which ranged from 0.99 (GIP1) to 1.03 (GIP4).

**Adjusting for genetic correlations with socioeconomic factors.** UK Biobank GWAS summary statistics for household income and Townsend's deprivation index were downloaded from the Neale Lab GWAS collection<sup>62</sup> and subjected to the same SNP filters used for aging-related trait quality control. Genetic correlations between GWASs of aging-GIPs and the two socioeconomic GWASs were calculated using the HDL R package<sup>21</sup> v.1.3.4. As before, phenotypic correlations were estimated from the correlation between independent null  $z$  statistics.

For each aging-GIP, adjustment loadings were then calculated as:

$$\mathbf{a}_i = \{1, -\beta_1, -\beta_2\}$$

where  $\beta$  refers to  $\beta = \text{Cov}(\mathbf{X})^{-1} \times \text{Cov}(\text{GIP}, \mathbf{X})$ . Here,  $\text{Cov}(\mathbf{X})$  is the genetic covariance matrix of UK Biobank GWAS of household income and Townsend's deprivation index, and  $\text{Cov}(\text{GIP}, \mathbf{X})$  is the genetic covariance matrix between the aging-GIP and the two GWAS.

These loadings were used as weights in the linear combination framework to calculate new aging-GIP summary statistics, which—by definition—were uncorrelated to the UK Biobank GWAS of household income and Townsend's deprivation index. Last, s.e.s were adjusted to account for the LDSC regression intercept, which ranged from 0.99 (GIP3) to 1.02 (GIP1).

**Characterization of genome-wide significant aging-GIP1 loci.** Genome-wide significant loci were defined as 500-kb regions centered on a lead genome-wide significant SNP ( $P < 8.3 \times 10^{-9}$ ) in linkage equilibrium ( $r^2 < 0.1$ ) with other lead locus SNPs. The LD between SNPs was calculated using a random 10,000 subset of unrelated, genomically British individuals from UK Biobank<sup>71</sup>. The susieR package<sup>26</sup> v.0.11.8 was used to perform genetic fine-mapping of the association signals within each aging-GIP1 locus, allowing for up to ten causal variants. We report posterior inclusion probabilities for SNPs within 95% credible sets. Positional annotations of credible set SNPs were retrieved using ANNOVAR<sup>72</sup> v.2019Oct24.

SNPs within each 500-kb locus were looked up in the six aging-related trait GWASs used to construct the aging-GIPs. Any region containing an SNP associated with an aging-related trait at genome-wide significance ( $P < 5 \times 10^{-8}$ ) was tested for colocalization between the trait and the aging-GIP1 signal. For this, we used the coloc R package<sup>73</sup> v.4.0.4 with its default parameters, denoting a posterior colocalization probability (PP) of 80% as evidence of a shared GWAS signal.

**Association of loci with Finnish and Japanese survival.** Loci were considered to be previously replicated if they had been associated at genome-wide significance with one of the core aging traits and also had evidence of an effect in an independent cohort on the same trait ( $P < 0.05$ ). We attempted to find additional evidence for an effect on aging for the aging-GIP1 loci, which had not been previously replicated.

Effects were first looked up in a GWAS of survival of FinnGen study participants<sup>73</sup>. The FinnGen study associated SNPs across the genome with the survival of 218,396 Finnish-ancestry individuals (203,244 censored, 15,152 deceased) using Genetic Analysis of Time-to-Event phenotypes (GATE) v.0.40. SNP effects were log(hazard ratios), calculated from a mixed-effect frailty model that adjusted for sex, genotyping batch, birth year and the first ten genomic principal components as fixed effects, and cryptic relatedness using the genetic relatedness matrix as random effects.

Analogously, the same SNPs (if polymorphic) were regressed against the survival of 135,983, unrelated, Japanese-ancestry individuals (97,365 censored, 30,976 deceased) from BioBank Japan<sup>74,75</sup>. In this analysis, a fixed-effect Cox's proportional hazards model was fitted using the survival R package v.2.41:

$$h(t) = h_0(t) \exp(\mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \dots + \mathbf{X}_n\beta_n + \mathbf{G}\gamma) \quad (4)$$

where  $h(t)$  is the hazard at time  $t$ , given that the subject is alive at time  $t$  and  $h_0(t)$  is the baseline hazard at time  $t$ ;  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  are the vectors of covariates with fixed effects  $\beta_1, \beta_2, \dots, \beta_n$ ; and  $\gamma$  is the effect of the vector of SNP dosages  $\mathbf{G}$ . Covariates were sex, disease status and the first 20 principal components, where disease status refers to 1 of 47 common diseases in Japan used to recruit the individuals. Each SNP was tested in a separate model.

The SNP effects from both studies were converted from log(hazard ratios) to approximate years of life by inverting the sign and multiplying the effect estimate and s.e.s by ten<sup>3</sup>. For each SNP, a combined effect was calculated by meta-analyzing the cohort-specific effects in a fixed-effect framework (weighted using inverse

variance), implemented in the meta R package<sup>70</sup> v.4.15-1. One-sided  $P$  values were adjusted for multiple testing of 23 loci using Bonferroni's correction. The collective effect of the 21 remaining loci was calculated using a random-effect framework, to allow for heterogeneity in effect size estimates.

**Aging-GIP1 leave-one-out sensitivity analyses.** Leave-one-out sensitivity analyses were performed for aging-GIP1, where, one at a time, a core aging trait was excluded and aging-GIP1 loadings and summary statistics were recalculated using the remaining five traits. Genetic and nongenetic correlations were calculated between the original aging-GIP1 and each leave-one-out GIP using HDL inference and null SNP  $z$  statistics, as described above.

To test for heterogeneity in genome-wide significant aging-GIP1 loci, we estimated the difference between the lead SNP effect in aging-GIP1 and the effect in the leave-one-out GIP1 GWAS, taking into account null correlations between traits. The s.e. of the difference in effects was calculated as follows:

$$\begin{aligned} & \text{s.e.}(\hat{\beta}_1 - \hat{\beta}_2) \\ &= \sqrt{\text{s.e.}(\hat{\beta}_1)^2 + \text{s.e.}(\hat{\beta}_2)^2 - 2r \times \text{s.e.}(\hat{\beta}_1) \times \text{s.e.}(\hat{\beta}_2)} \end{aligned} \quad (5)$$

where  $\text{s.e.}(\hat{\beta}_1)$  and  $\text{s.e.}(\hat{\beta}_2)$  are the s.e.s of the SNP for aging-GIP1 and the leave-one-out GIP1, respectively, and  $r$  is the nongenetic correlation between GWASs. Statistical significance of the difference was assessed using Wald's test and adjusted for multiple testing of 27 loci using Bonferroni's correction.

**Lookup of known SNP associations.** Lead SNP and close proxies ( $r^2_{\text{EUR}} \geq 0.8$ ) of the aging-GIP1 loci were looked up in PhenoScanner<sup>27</sup> and the GWAS catalog<sup>28</sup> (accessed 3 December 2020), keeping only the traits with genome-wide significance ( $P < 5 \times 10^{-8}$ ). Triallelic SNPs and associations with treatments or medications were discarded, before converting associations with the lack of a phenotype into the phenotype itself by inverting the sign (for example, 'Qualifications: none' to 'Qualifications'). We then further grouped the traits based on similarities in trait names, keeping the strongest association in the group. This grouping was done by partial matching of trait names—verified manually—and keeping the shortest name. For example, 'Melanoma', 'Malignant melanoma' and 'Malignant melanoma of skin' were grouped and renamed to 'Melanoma'.

**Tissue enrichment.** Stratified LDSC regression v.1.0.0 was used to stratify aging-GIP1 SNPs into categories and test whether the proportion of SNP heritability in a category exceeded that expected from the proportion of SNPs in the category<sup>64</sup>. We kept only HapMap3 SNPs, excluding the MHC region and SNPs with  $\text{MAF} < 0.05$ , and used the 1000 Genomes Phase 3 LDSC reference as weights. Categories tested included the 10 groups summarizing 220 cell-type specific annotations from Finucane et al.<sup>64</sup>, adjusting for the baseline model (v.1.2).

**Gene and pathway enrichment.** PASCAL<sup>29</sup> was used to aggregate aging-GIP1 SNP-level  $P$  values (with and without adjustment for socioeconomic correlations) into gene scores and test these scores for enrichment against predefined gene sets. Gene sets were Hallmark (C1) and Gene Ontology Biological Process (C5.BP) sets from v.7.2 of the Molecular Signatures Database<sup>76</sup>.

Aging-GIP1 summary statistics were first aligned to the 1000 Genomes SNP build (matching the PASCAL LD reference), before being tested with default PASCAL parameters, which includes discarding SNPs with  $\text{MAF} < 5\%$  and SNPs in the MHC region. Gene results passing a 5% FDR threshold were considered to be significant. For each pathway in the C1 and C5.BP datasets, PASCAL calculated two measures of significance based on  $\chi^2$  and permutation statistics. We separately adjusted C1 and C5.BP for multiple testing and considered a pathway with both  $\chi^2$  and permutation statistics passing a 5% FDR threshold to be significant.

Significant C5.BP pathways (with and without adjustment for socioeconomic correlations) were clustered based on their Jaccard's similarity coefficient (that is, the size of the intersection of genes divided by the size of the union of genes). The mclust R package<sup>69</sup> v.5.4.1 was used to maximize the Bayesian information criterion of Jaccard's similarity matrix to identify the optimal number of clusters (up to a maximum of 100). We used the number of clusters selected by most mclust models to group the pathways.

**MR of blood protein levels.** Genetic instruments for blood protein levels (pQTLs) were retrieved from Zheng et al.<sup>30</sup>. We included all tier 1 instruments: *cis*- and *trans*-pQTLs shown to influence  $\leq 5$  proteins (specificity) with no evidence of heterogeneity in effect sizes between multiple protein expression studies (consistency). A total of 857 proteins had nonpalindromic SNP instruments (898 total pQTLs) present in the aging-GIP1 summary statistics. Two-sample MR of blood protein levels as exposures and aging-GIP1 as outcome was performed using the TwoSampleMR R package<sup>31</sup> v.0.5.5. If multiple pQTL instruments were available for a protein, heterogeneity and MR-Egger sensitivity tests were also performed. This analysis was repeated with the six standardized aging-GIP1 component traits as outcome, as well as the unstandardized, combined parental lifespan GWAS from Timmers et al.<sup>4</sup> to provide an intuitive measure of the effect.

The MR effects and s.e.s from the latter were multiplied by 10 to convert them from units of negative log(hazard ratio) to approximate years of life<sup>3</sup>.

Aging-GIP1 MR results passing a 5% FDR threshold and sensitivity tests ( $P_{\text{Hci}} > 0.05$  and  $P_{\text{Egger}} > 0.05$ , if applicable) were taken forward for follow-up colocalization tests to rule out LD linkage. For proteins for which we had access to genome-wide summary statistics (for example, LPA,  $\beta_2\text{M}$  and VCAM1), we used the coloc R package<sup>32</sup> v.4.0-4 to perform colocalization analysis using the default parameters and denoted  $\text{PP} \geq 80\%$  as evidence of a shared signal. For instruments without summary statistics, we performed an LD check as described in Zheng et al.<sup>30</sup>, which involved checking whether any of the 30 strongest aging-GIP1 SNPs in a 1-Mb region centered on each pQTL were in high LD ( $r^2_{\text{EUR}} \geq 0.8$ ) with that pQTL.

Finally, proteins passing sensitivity and colocalization tests were subjected to a reverse-causality test using MR-Steiger<sup>77</sup> and, if genome-wide summary statistics were available, a bidirectional MR analysis, both implemented in TwoSampleMR. For the bidirectional MR, we used up to 27 genome-wide significant lead SNPs from aging-GIP1 (shared between GWASs and replacing missing or palindromic SNPs with the next most significant SNP) as instruments and the protein expression statistics as outcome. Proteins significant for the MR-Steiger test ( $P < 0.05$ ) and showing no evidence of reverse causality in the bidirectional MR ( $P > 0.05$ ), if applicable, were considered to have robust causal effects on aging-GIP1.

**Ethical oversight.** BioBank Japan participants provided written informed consent and survival study protocols were approved by BioBank Japan Project ethical review boards from the Institute of Medical Sciences, University of Tokyo and the RIKEN Center for Integrative Medical Sciences. All other data were publicly available and approved by ethical committees as described in their respective publications. See Supplementary Note for FinnGen ethical approval.

**Statistics and reproducibility.** The statistical methods used to analyze the data are described fully in Methods, with basic data processing done using R v.3.6.0 unless specified otherwise. We used the independent FinnGen and BioBank Japan cohorts to successfully replicate 2 of the 23 new aging-related loci. Predetermination of study sample size, randomization of experiments and blinding of investigators to experiments were not applicable for our type of study.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Aging-related trait GWAS summary statistics can be retrieved or requested from the study authors at <http://www.nealelab.is/uk-biobank> (self-rated health; field 2178), <https://www.longevitygenomics.org/downloads> (longevity), and the following: <https://doi.org/10.5281/zenodo.1302861> (healthspan), <https://doi.org/10.6084/m9.figshare.9204998.v3> (frailty index), <https://doi.org/10.5523/bris.21crwsnj4xwjm2g4qi8chathha> (perceived age), <https://doi.org/10.7488/ds/2463> (parental lifespan), <https://doi.org/10.7488/ds/2631> (epigenetic age acceleration), <https://doi.org/10.1016/j.ajhg.2020.02.006> (telomere length) and <https://doi.org/10.1038/ng.3821> (mLOY). Summary statistics for aging-GIP1 calculated in the present study have been deposited in the Edinburgh DataShare repository, available without restrictions at <https://doi.org/10.7488/ds/2972>. Summary statistics of the GWAS-MAP phenotypes used to calculate phenome-wide genetic correlations are available from GeneAtlas (<http://geneatlas.roslin.ed.ac.uk>), NealeLab (<http://www.nealelab.is/uk-biobank>) or their respective publications. The COVID-19 GWAS summary statistics have been made available by the COVID-19 Host Genetics Initiative at <https://www.covid19hg.org/results>. GWAS catalog and PhenoScanner associations can be found at <https://www.ebi.ac.uk/gwas> and <http://www.phenoscaner.medschl.cam.ac.uk>, respectively. Curated gene sets (Hallmark and GO) are available from the Molecular Signatures Database (<https://www.gsea-msigdb.org>). Source data are provided with this paper, or are available in the supplementary documents and upon request from the corresponding author.

## Code availability

The following codes are available: HDL: <https://github.com/zhenin/HDL>; LDSC: <https://github.com/bulik/ldsc>; SuSiE: <https://stephenslab.github.io/susieR/>; ANNOVAR: <https://annovar.openbioinformatics.org/>; PASCAL: <https://www2.unil.ch/cbg/index.php?title=Pascal>; TwoSampleMR: <https://mrceu.github.io/TwoSampleMR>. This code can be accessed without restrictions.

Received: 15 March 2021; Accepted: 25 November 2021;  
Published online: 20 January 2022

## References

- López-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194–1217 (2013).
- Zenin, A. et al. Identification of 12 genetic loci associated with human healthspan. *Commun. Biol.* **2**, 41 (2019).
- Joshi, P. K. et al. Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat. Commun.* **8**, 910 (2017).
- Timmers, P. R. H. J. et al. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *eLife* **8**, e39856 (2019).
- Li, C. et al. Genome-wide association analysis in humans links nucleotide metabolism to leukocyte telomere length. *Am. J. Hum. Genet.* **106**, 389–404 (2020).
- Forsberg, L. A. et al. Mosaic loss of chromosome Y in peripheral blood is associated with shorter survival and higher risk of cancer. *Nat. Genet.* **46**, 624–628 (2014).
- Wright, D. J. et al. Genetic variants associated with mosaic Y chromosome loss highlight cell cycle genes and overlap with cancer susceptibility. *Nat. Genet.* **49**, 674–679 (2017).
- Atkins, J. et al. A genome-wide association study of the frailty index highlights synaptic pathways in aging. *Aging Cell* **20**, e13459 (2019).
- Mitnitski, A. B., Mogilner, A. J. & Rockwood, K. Accumulation of deficits as a proxy measure of aging. *Sci. World J.* **1**, 323–336 (2001).
- Ruby, J. G. et al. Estimates of the heritability of human longevity are substantially inflated due to assortative mating. *Genetics* **210**, 1109–1124 (2018).
- Timmers, P. R. H. J., Wilson, J. F., Joshi, P. K. & Deelen, J. Multivariate genomic scan implicates novel loci and haem metabolism in human ageing. *Nat. Commun.* **11**, 3570 (2020).
- Melzer, D., Pilling, L. C. & Ferrucci, L. The genetics of human ageing. *Nat. Rev. Genet.* **21**, 88–101 (2020).
- Shen, X. et al. Multivariate discovery and replication of five novel loci associated with immunoglobulin G N-glycosylation. *Nat. Commun.* **8**, 447 (2017).
- Ning, Z. et al. Beyond power: multivariate discovery, replication, and interpretation of pleiotropic loci using summary association statistics. Preprint at *bioRxiv* <https://doi.org/10.1101/022269> (2015).
- Tsepilov, Y. A. et al. Analysis of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain conditions. *Commun. Biol.* **3**, 329 (2020).
- Harris, S. E. et al. Molecular genetic contributions to self-rated health. *Int. J. Epidemiol.* **46**, 994–1009 (2017).
- Deelen, J. et al. A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat. Commun.* **10**, 3669 (2019).
- Roberts, V., Main, B., Timpson, N. J. & Haworth, S. Genome-wide association study identifies genetic associations with perceived age. *J. Invest. Dermatol.* **140**, 2380–2385 (2020).
- Gibson, J. et al. A meta-analysis of genome-wide association studies of epigenetic age acceleration. *PLoS Genet.* **15**, e1008104 (2019).
- Shashkova, T. I. et al. The GWAS-MAP platform for aggregation of results of genome-wide association studies and the GWAS-MAP| homo database of 70 billion genetic associations of human traits. *Vavilov J. Genetics Breeding* **24**, 876–884 (2020).
- Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations across human complex traits. *Nat. Genet.* **52**, 859–864 (2020).
- ICD-10: *International Statistical Classification of Diseases and Related Health Problems: Tenth Revision* 2nd edn (World Health Organization, 2004); <https://apps.who.int/iris/handle/10665/42980>
- Hill, W. D. et al. Molecular genetic contributions to social deprivation and household income in UK Biobank. *Curr. Biol.* **26**, 3083–3089 (2016).
- Abdellaoui, A., Verweij, K. J. H. & Nivard, M. G. Geographic confounding in genome-wide association studies. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.18.435971> (2021).
- Abdellaoui, A. et al. Genetic correlates of social stratification in Great Britain. *Nat. Hum. Behav.* **3**, 1332–1342 (2019).
- Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. B Stat. Methodol.* **82**, 1273–1300 (2020).
- Kamat, M. A. et al. PhenoScanner V2: an expanded tool for searching human genotype–phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).
- Buniello, A. et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
- Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).
- Hemani, G. et al. The MR-base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
- Zheng, J. et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).

32. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
33. Boffa, M. B., Marcovina, S. M. & Koschinsky, M. L. Lipoprotein(a) as a risk factor for atherosclerosis and thrombosis: mechanistic insights from animal models. *Clin. Biochem.* **37**, 333–343 (2004).
34. Nordestgaard, B. G. et al. Lipoprotein(a) as a cardiovascular risk factor: current status. *Eur. Heart J.* **31**, 2844–2853 (2010).
35. Viney, N. J. et al. Antisense oligonucleotides targeting apolipoprotein(a) in people with raised lipoprotein(a): two randomised, double-blind, placebo-controlled, dose-ranging trials. *Lancet* **388**, 2239–2253 (2016).
36. Arsenault, B. J. et al. Association of long-term exposure to elevated lipoprotein(a) levels with parental life span, chronic disease-free survival, and mortality risk: a Mendelian randomization analysis. *JAMA Netw. Open* **3**, e200129 (2020).
37. Allen, S. et al. Expression of adhesion molecules by Lp(a): a potential novel mechanism for its atherogenicity. *FASEB J.* **12**, 1765–1776 (1998).
38. Cho, T., Jung, Y. & Koschinsky, M. L. Apolipoprotein(a), through its strong lysine-binding site in KIV 10, mediates increased endothelial cell contraction and permeability via a rho/rho kinase/MYPT1-dependent pathway. *J. Biol. Chem.* **283**, 30503–30512 (2008).
39. Kong, D. H., Kim, Y. K., Kim, M. R., Jang, J. H. & Lee, S. Emerging roles of vascular cell adhesion molecule-1 (VCAM-1) in immunological disorders and cancer. *Int. J. Mol. Sci.* **19**, 13–17 (2018).
40. Yousef, H. et al. Aged blood impairs hippocampal neural precursor activity and activates microglia via brain endothelial cell VCAM1. *Nat. Med.* **25**, 988–1000 (2019).
41. Smith, L. K. et al. B2-microglobulin is a systemic pro-aging factor that impairs cognitive function and neurogenesis. *Nat. Med.* **21**, 932–937 (2015).
42. Grote, A. et al. LRP12 silencing during brain development results in cortical dyslamination and seizure sensitization. *Neurobiol. Dis.* **86**, 170–176 (2016).
43. Bethge, N. et al. A gene panel, including LRP12, is frequently hypermethylated in major types of B-cell lymphoma. *PLoS ONE* **9**, e104249 (2014).
44. Garnis, C., Coe, B. P., Zhang, L., Rosin, M. P. & Lam, W. L. Overexpression of LRP12, a gene contained within an 8q22 amplicon identified by high-resolution array CGH analysis of oral squamous cell carcinomas. *Oncogene* **23**, 2582–2586 (2004).
45. Nakaya, N., Sultana, A., Lee, H. S. & Tomarev, S. I. Olfactomedin 1 interacts with the Nogo A receptor complex to regulate axon growth. *J. Biol. Chem.* **287**, 37171–37184 (2012).
46. Shi, W. et al. Olfactomedin 1 negatively regulates NF- $\kappa$ B signalling and suppresses the growth and metastasis of colorectal cancer cells. *J. Pathol.* **240**, 352–365 (2016).
47. Cheng, A. et al. Pancortin-2 interacts with WAVE1 and Bcl-xL in a mitochondria-associated protein complex that mediates ischemic neuronal death. *J. Neurosci.* **27**, 1519–1528 (2007).
48. Yang, C. et al. Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat. Neurosci.* **24**, 1302–1312 (2021).
49. Rebo, J. et al. A single heterochronic blood exchange reveals rapid inhibition of multiple tissues by old blood. *Nat. Commun.* **7**, 13363 (2016).
50. Katsimpardi, L. et al. Vascular and neurogenic rejuvenation of the aging mouse brain by young systemic factors. *Science* **344**, 630–634 (2014).
51. Balla, J. et al. Haem, haem oxygenase and ferritin in vascular endothelial cell injury. *Nephrol. Dial. Transplant.* **18**, v8–v12 (2003).
52. Higashi, Y. & Yoshizumi, M. Endothelial function. *Jpn. J. Clin. Med.* **61**, 1138–1144 (2003).
53. Engelhardt, B. Development of the blood–brain barrier. *Cell Tissue Res.* **314**, 119–129 (2003).
54. Lakatta, E. G. & Levy, D. Arterial and cardiac aging: major shareholders in cardiovascular disease enterprises. Part I: aging arteries: a ‘set up’ for vascular disease. *Circulation* **107**, 139–146 (2003).
55. Montagne, A. et al. APOE4 leads to blood–brain barrier dysfunction predicting cognitive decline. *Nature* **581**, 71–76 (2020).
56. Popov, N. et al. The ubiquitin-specific protease USP28 is required for MYC stability. *Nat. Cell Biol.* **9**, 765–774 (2007).
57. Serrano, M. The tumor suppressor protein p16INK4a. *Exp. Cell. Res.* **237**, 7–13 (1997).
58. Li, F. et al. USP28 regulates deubiquitination of histone H2A and cell proliferation. *Exp. Cell. Res.* **379**, 11–18 (2019).
59. Baker, D. J. et al. Naturally occurring p16INK4a-positive cells shorten healthy lifespan. *Nature* **530**, 184–189 (2016).
60. Timmers, P. R. H. J. et al. Mendelian randomization of genetically independent aging phenotypes identifies LPA and VCAM1 as biological targets for human aging [dataset]. *University of Edinburgh* <https://doi.org/10.7488/ds/2972> (2020).
61. Zenin, A. et al. Genome-wide association summary statistics for human healthspan (Version 1) [dataset]. *Zenodo* <https://doi.org/10.5281/zenodo.1302861> (2018).
62. Neale, B. UK Biobank—Neale lab. *Neale Lab* <http://www.nealelab.io/uk-biobank/%0A> (2018).
63. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
64. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
65. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).
66. Shashkova, T. I. et al. PheLiGe: an interactive database of billions of human genotype–phenotype associations. *Nucleic Acids Res.* **49**, D1347–D1350 (2021).
67. Ying, K. et al. Genetic and phenotypic analysis of the causal relationship between aging and COVID-19. *Commun. Med.* **1**, 35 (2021).
68. The COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* **28**, 715–718 (2020).
69. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. Mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *R J.* **8**, 289–317 (2016).
70. Schwarzer, G., Carpenter, J. R. & Rücker, G. *Meta-Analysis with R* (Springer International Publishing, 2015).
71. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
72. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
73. Dey, R. et al. An efficient and accurate frailty model approach for genome-wide survival association analysis controlling for population structure and relatedness in large-scale biobanks. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.10.31.358234> (2020).
74. Hirata, M. et al. Overview of BioBank Japan follow-up data in 32 diseases. *J. Epidemiol.* **27**, S22–S28 (2017).
75. Sakaue, S. et al. Trans-biobank analysis with 676,000 individuals elucidates the association of polygenic risk scores of complex traits with human lifespan. *Nat. Med.* **26**, 542–548 (2020).
76. Liberzon, A. et al. The molecular signatures database Hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
77. Hemani, G., Tilling, K. & Smith, G. D. Orienting the causal relationship between imprecisely measured traits using genetic instruments. Preprint at *bioRxiv* <https://doi.org/10.1101/117101> (2017).

## Acknowledgements

We thank the UK Biobank Resource, approved under application 19655. We thank the Medical Research Council Human Genetics Unit for their funding (to P.R.H.J.T. and J.F.W.: grant no. MC\_UU\_00007/10), and also the University of Edinburgh (to P.K.J.), the Russian Ministry of Education and Science 5–100 Excellence Programme (to E.S.T. and Y.A.T.) and the Ministry of Education and Science of the RF via the Institute of Cytology and Genetics SB RAS (to E.S.T. and Y.A.T.: grant no. 0259-2021-0009/AAAA-A17-117092070032-4). The FinnGen project and authors (T.T.J.K., W.Z. and A.G.) acknowledge funding from Business Finland (nos. HUS 4685/31/2016 and UH 4386/31/2016) as well as the industry partners AbbVie Inc., AstraZeneca UK Ltd., Biogen MA Inc., Celgene Corporation, Celgene International II Sàrl, Genentech Inc., Merck Sharp & Dohme Corp, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc. and Novartis AG. The following biobanks are acknowledged for their samples in the FinnGen project: Auria Biobank ([www.auria.fi/biopankki](http://www.auria.fi/biopankki)), THL Biobank ([www.thl.fi/biobank](http://www.thl.fi/biobank)), Helsinki Biobank ([www.helsinginbiopankki.fi](http://www.helsinginbiopankki.fi)), Biobank Borealis of Northern Finland (<https://www.ppsph.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx>), Finnish Clinical Biobank Tampere ([www.tays.fi/en-US/Research\\_and\\_development/Finnish\\_Clinical\\_Biobank\\_Tampere](http://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)), Biobank of Eastern Finland ([www.ita-suomenbiopankki.fi/en](http://www.ita-suomenbiopankki.fi/en)), Central Finland Biobank ([www.ksshp.fi/fi-FI/Potilaalle/Biopankki](http://www.ksshp.fi/fi-FI/Potilaalle/Biopankki)), Finnish Red Cross Blood Service Biobank ([www.veripalvelu.fi/verenluovutus/biopankkitoiminta](http://www.veripalvelu.fi/verenluovutus/biopankkitoiminta)) and Terveystalo Biobank ([www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki](http://www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki)). All Finnish Biobanks are members of BBMRI infrastructure ([www.bbmri.fi](http://www.bbmri.fi)) and FinBB (<https://finbb.fi>). The Framingham Heart Study is funded by a National Institutes of Health (NIH) contract (nos. N01-HC-25195, HHSN268201500001 and 75N92019D00031; Boston University). This project was funded in part by the Division of Intramural Research, National Heart, Lung and Blood Institute (NHLBI), NIH, Bethesda, MD. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the NHLBI, the NIH or the US Department of Health and Human Services. None of the funders had any role in study design, data collection and analysis, decision to publish or preparation of the manuscript. For the purpose of Open Access, the authors have applied a CC-BY public copyright license to any author accepted manuscript version arising from this submission.

## Author contributions

P.R.H.J.T. conceived the project, provided the methodology, software, validation, formal analysis and investigation, wrote the original draft and reviewed and edited it, and visualized and administered the project. E.S.T. and J.D. provided formal analysis, and

reviewed and edited the manuscript. S.S., M.A., T.T.J.K. and W.Z. provided validation and formal analysis. S.-J.H. and C.Y. provided formal analysis. D.L. provided resources, reviewed and edited the manuscript, and supervised the project. A.G. and Y.O. provided resources and supervised the project. Y.K. provided resources and curated the data. P.K.J. reviewed and edited the manuscript, and supervised the project. J.F.W. provided resources, reviewed and edited the manuscript, supervised the project and acquired funding. Y.A.T. conceived the project, provided the methodology, software and formal analysis, reviewed and edited the manuscript, and supervised the project.

### Competing interests

P.R.H.J.T. is a salaried employee of BioAge Labs Inc. P.K.J. is a paid consultant for Humanity Inc. and Global Gene Corporation. The remaining authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43587-021-00159-8>.

**Correspondence and requests for materials** should be addressed to Paul R. H. J. Timmers.

**Peer review information** *Nature Aging* thanks Xiao Dong and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

## Biobank Japan Project

Yoichiro Kamatani<sup>9,18</sup>

---

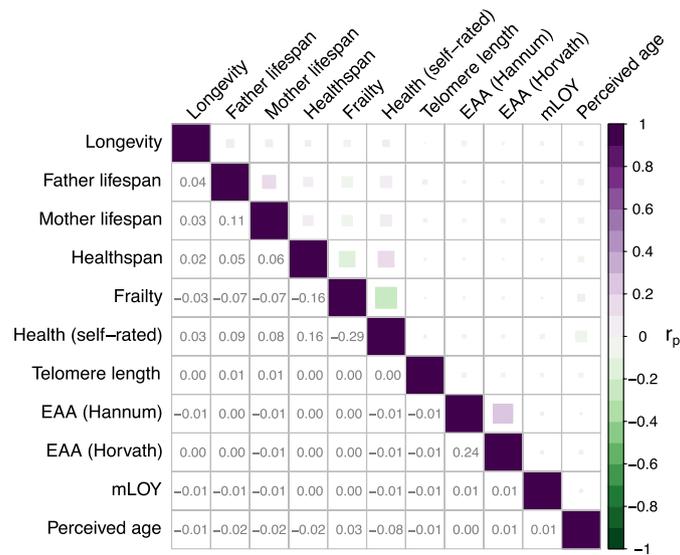
A full list of Biobank Japan Project members and their affiliations appears in the Supplementary Information.

## FinnGen

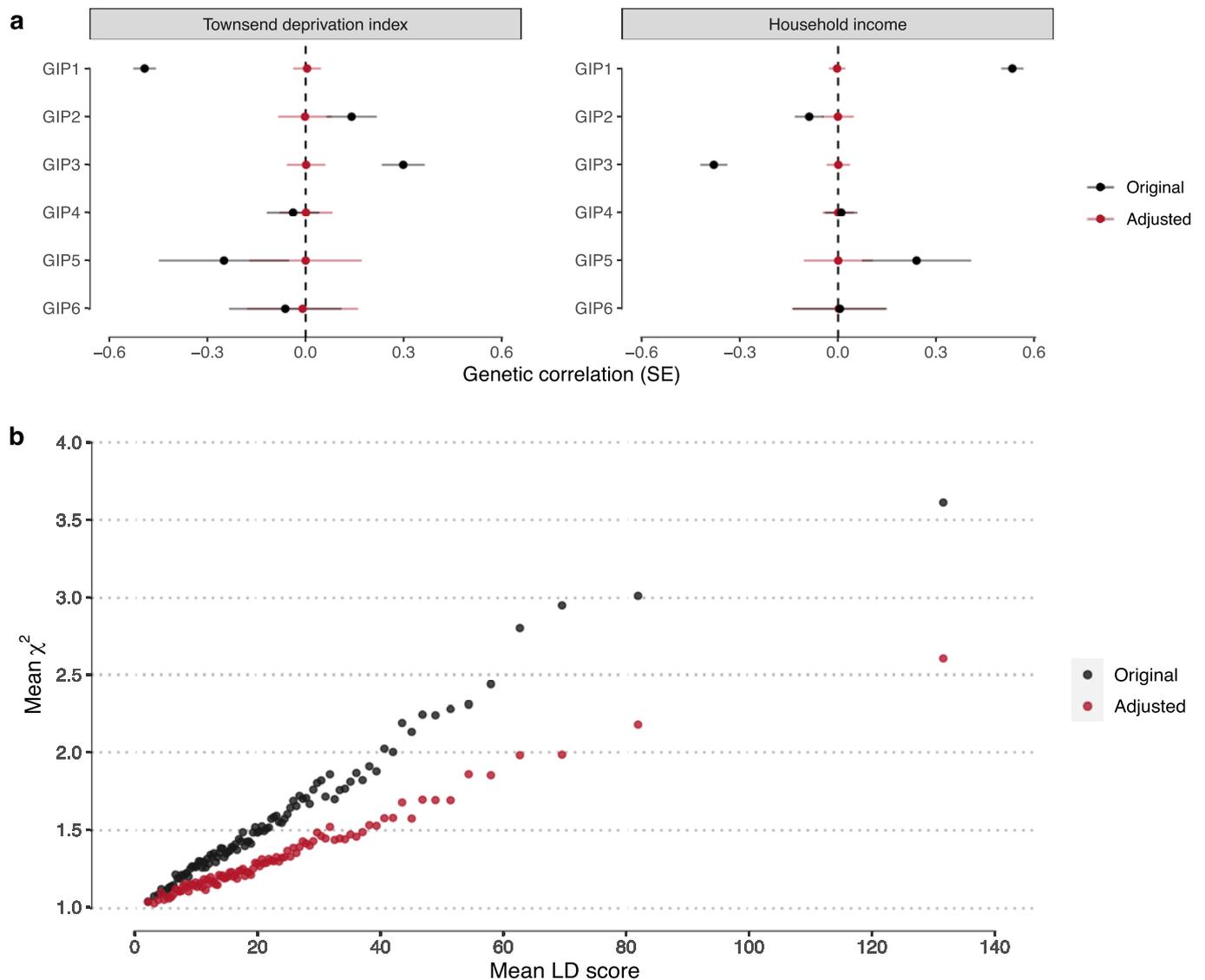
Tuomo T. J. Kiiskinen<sup>8,11</sup>, Wei Zhou<sup>8,12,13</sup> and Andrea Ganna<sup>8,11,12</sup>

---

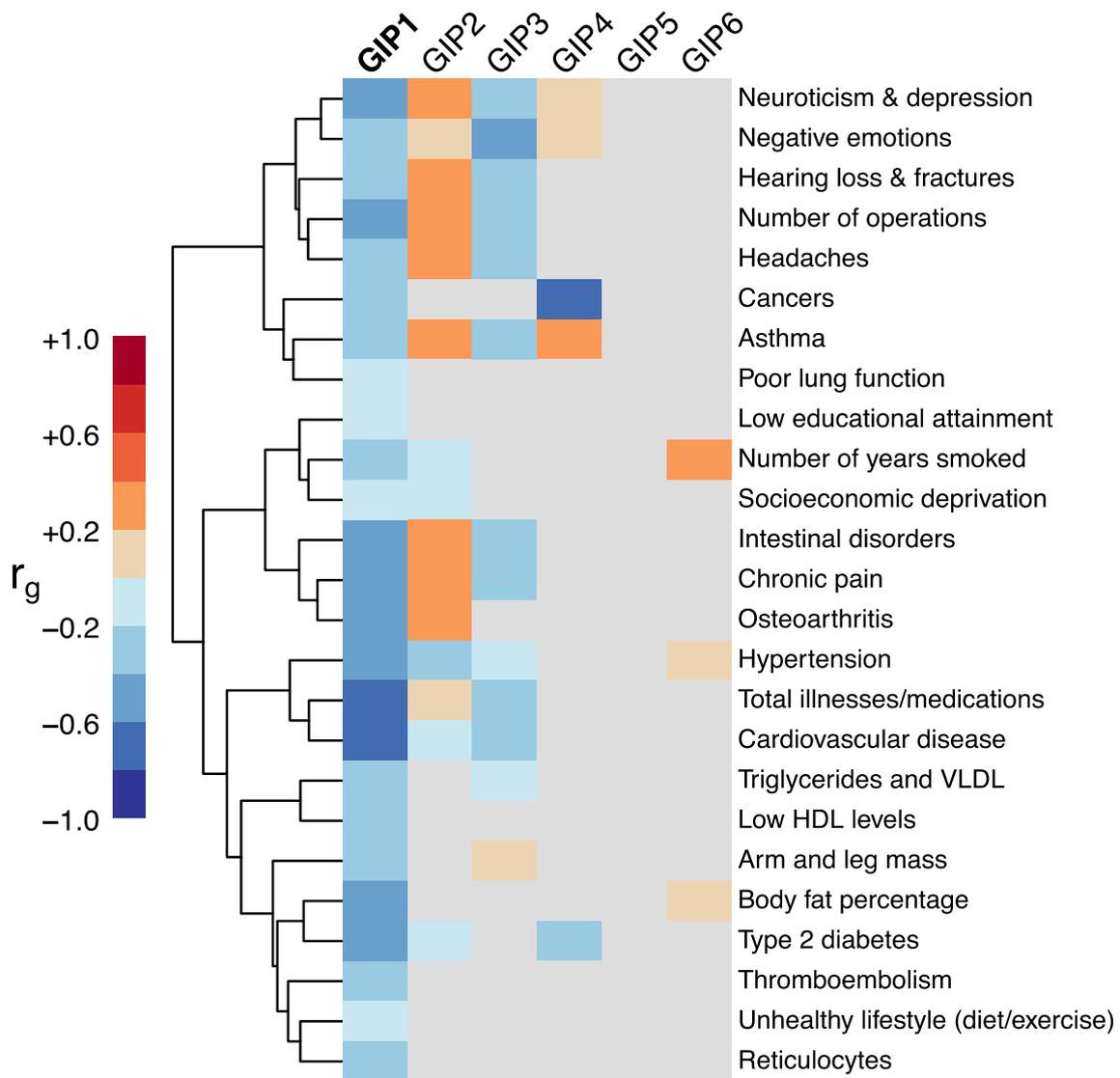
A full list of FinnGen members and their affiliations appears in the Supplementary Information.



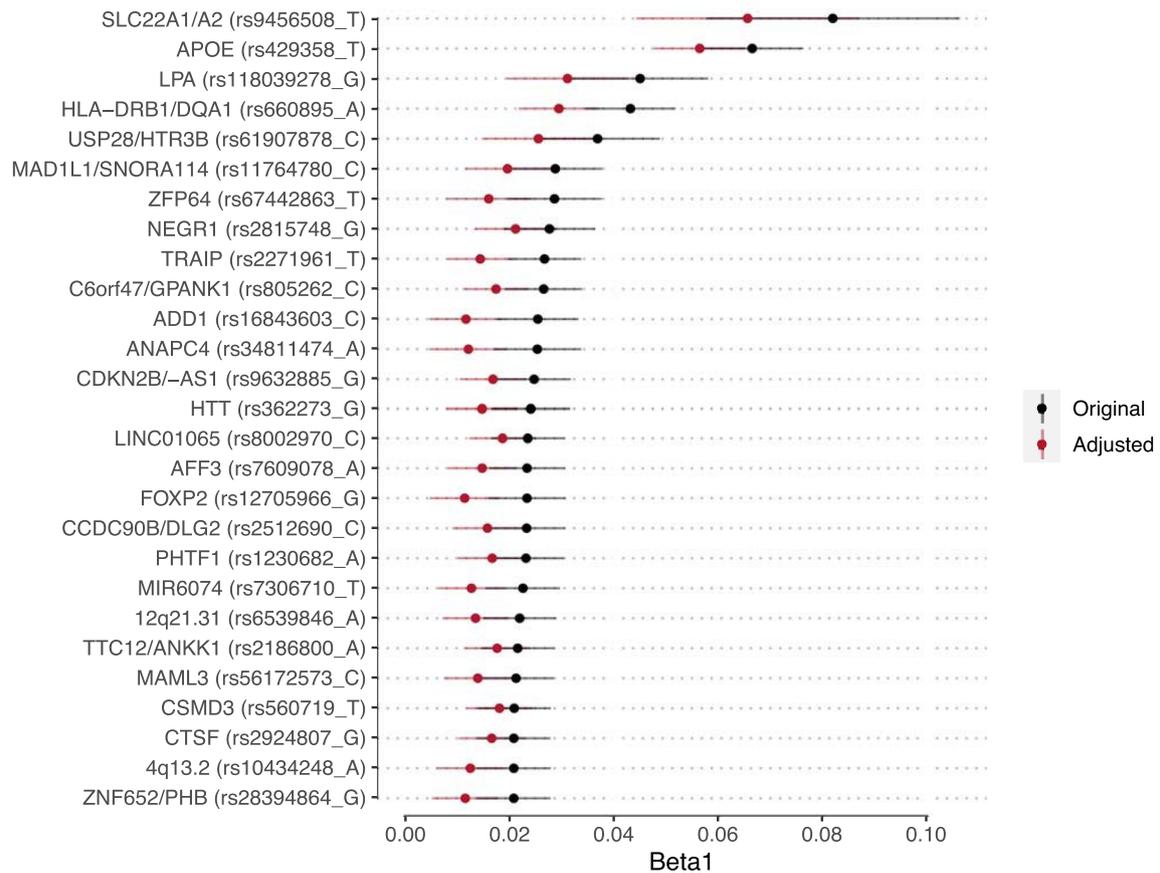
**Extended Data Fig. 1 | Non-genetic correlations between ageing-related GWASs.** Off-diagonal values show the phenotypic correlation between GWASs observed due to sample overlap. Correlations were calculated using the Z scores of independent SNPs (provided by the MultiABEL R package)<sup>13</sup> that were non-significant in both studies ( $|z| < 1.96$ ). The number of SNPs used for each pairwise correlation is reported in Supplementary Data 1. EAA, Epigenetic Age Acceleration; mLOY, mosaic Loss of Y chromosome.



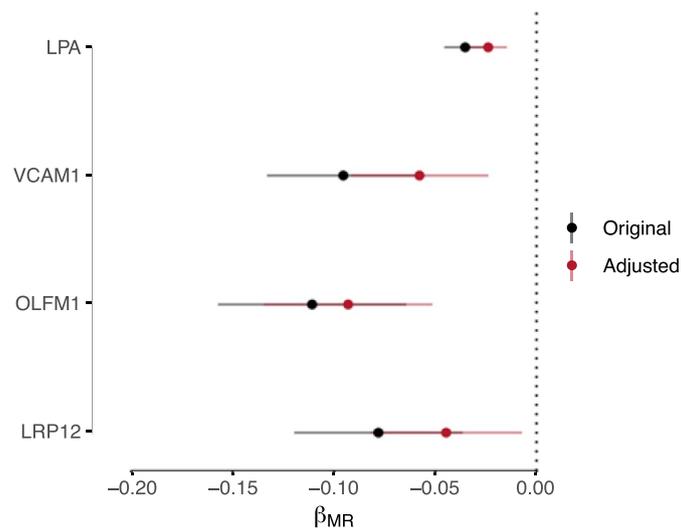
**Extended Data Fig. 2 | Adjustment of aging-GIPs to remove genetic correlations with socioeconomic status and household income.** a) Genetic correlations of aging-GIPs with UK Biobank GWASs of Townsend's deprivation index and household income. In black are the original correlations before adjustment, in red are the correlations after adjustment. Error bars represent 95% confidence intervals based on 1,027,336 SNPs (99.75%) from the HDL R package<sup>2</sup> LD reference panel. b) LD score plot showing the relationship between LD scores and  $\chi^2$  statistics of aging-GIP1 SNPs, before and after adjustment. Each dot represents an LD score percentile. The x-axis shows the mean LD score of variants in each LD score percentile, and the y-axis shows the mean  $\chi^2$  statistics of SNPs in that LD score percentile. The weighted slope of the regression can be used as a measure of the trait heritability.



**Extended Data Fig. 3 | Phenome-wide genetic correlations of aging-GIPs adjusted for household income and Townsend’s deprivation index.** Shown are the genetic correlations of the 25 GWAS-MAP clusters (defined in Fig. 3) with aging-GIPs adjusted for socioeconomic status and household income (as measured in UK Biobank). Values failing to pass nominal significance (two-sided Wald test  $P \geq 0.05$ ) are greyed out. See Supplementary Data 3 for the full list of correlations.



**Extended Data Fig. 4 | Effect sizes of lead aging-GIP1 SNPs adjusted for household income and Townsend's deprivation index.** The x-axis shows the mean effect size estimate of the lead SNPs associated with aging-GIP1 at Bonferroni-corrected genome-wide significance ( $P < 8.3 \times 10^{-9}$ ), with 95% confidence intervals represented by lines. The estimated sample sizes are 154,478 and 195,015 for the original and adjusted aging-GIP1 GWAS, respectively. The y-axis is annotated with the nearest gene(s) or cytogenetic band to each locus, with the lead SNP and the aging-GIP1-increasing allele in parentheses. In black are the original aging-GIP1 estimates, while in red are the estimates after removing all genetic correlations with GWASs of UK Biobank socioeconomic status and household income.



**Extended Data Fig. 5 | MR of blood protein levels on aging-GIP1 adjusted for household income and Townsend's deprivation index.** Exposures significant for a causal effect on aging-GIP1 were tested for their effect on the aging-GIP1 residual phenotype, which was adjusted to remove all genetic correlations with GWASs of socioeconomic status and household income in UK Biobank. The x-axis shows the MR effect estimates, with lines representing 95% confidence intervals, as estimated by TwoSampleMR<sup>31</sup>. The sample sizes of exposure GWASs were 6,861 for LPA; 3,301 for VCAM1; and 3,200 for OLFM1 and LRP12. In black is the original aging-GIP1 causal effect estimate, while in red is the effect with the adjusted aging-GIP1 phenotype. LPA—apolipoprotein(a). VCAM1—vascular cell adhesion molecule 1. OLFM1—olfactomedin 1. LRP12— LDL receptor related protein 12.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection No software was used to collect the data.

Data analysis High-definition likelihood (HDL) R package v1.3.4; LD-score regression (LDSC) v1.0.0; meta R package v4.15-1; Genetic Analysis of Time-to-Event phenotypes (GATE) v0.40; survival R package v2.41; Stratified LD-score regression v1.0.0; PASCAL v1; mclust R package v5.4.1; susieR package v0.11.8; ANNOVAR v.2019Oct24; TwoSampleMR R package v0.5.5; R v3.6.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Ageing-related trait GWAS summary statistics can be retrieved or requested from study authors at <http://www.nealelab.is/uk-biobank/> (self-rated health; field 2178), <https://www.longevitygenomics.org/downloads> (longevity), and the following DOI: 10.5281/zenodo.1302861 (healthspan), 10.6084/m9.figshare.9204998.v3 (frailty index), 10.5523/bris.21crwsnj4xwjm2g4qi8chathha (perceived age), 10.7488/ds/2463 (parental lifespan), 10.7488/ds/2631 (epigenetic age acceleration), 10.1016/j.ajhg.2020.02.006 (telomere length), 10.1038/ng.3821 (mLOY). Summary statistics for Ageing-GIP1 calculated in this study have been deposited in the Edinburgh DataShare repository, available without restrictions at <https://doi.org/10.7488/ds/2972>. Summary statistics of the GWAS-MAP phenotypes used to calculate phenome-wide genetic correlations are available from GeneAtlas (<http://geneatlas.roslin.ed.ac.uk/>), NealeLab (<http://www.nealelab.is/uk-biobank/>), or

their respective publications. The COVID-19 GWAS summary statistics have been made available by the COVID-19 Host Genetics Initiative at <https://www.covid19hg.org/results/>. GWAS catalog and PhenoScanner associations can be found at <https://www.ebi.ac.uk/gwas/> and <http://www.phenoscaner.medschl.cam.ac.uk/>, respectively. Curated gene sets (hallmark and gene ontology) are available from the Molecular Signatures Database (<https://www.gsea-msigdb.org/>). Source data for figures in this study are available in the supplementary documents and upon request from the corresponding author.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were not predetermined. Instead, sample sizes (i.e. the total number of individuals in each study) were reported in the ageing-related GWAS publications or their summary statistics. See <a href="http://www.nealelab.is/uk-biobank">http://www.nealelab.is/uk-biobank</a> , <a href="https://doi.org/10.7554/eLife.39856">https://doi.org/10.7554/eLife.39856</a> , <a href="https://doi.org/10.1101/19007559">https://doi.org/10.1101/19007559</a> , <a href="https://doi.org/10.1371/journal.pgen.1008104">https://doi.org/10.1371/journal.pgen.1008104</a> , <a href="https://doi.org/10.1038/s42003-019-0290-0">https://doi.org/10.1038/s42003-019-0290-0</a> , <a href="https://doi.org/10.1038/s41467-019-11558-2">https://doi.org/10.1038/s41467-019-11558-2</a> , <a href="https://doi.org/10.1038/ng.3821">https://doi.org/10.1038/ng.3821</a> , <a href="https://doi.org/10.1016/j.jid.2020.03.970">https://doi.org/10.1016/j.jid.2020.03.970</a> , <a href="https://doi.org/10.1016/j.ajhg.2020.02.006">https://doi.org/10.1016/j.ajhg.2020.02.006</a> . Sample sizes for the Ageing-GIPs were estimated based on the median Z statistic and the allele frequencies of null SNPs.
Data exclusions	For each set of summary statistics, we discarded SNPs that were poorly imputed (INFO < 40%), rare (MAF < 0.5%), or poorly measured (N individuals < 1% of total). This was to restrict our analysis to reliable SNPs and prevent poorly measured outliers from biasing downstream analyses.
Replication	We measured lead SNP effects on FinnGen participant survival (Release 5; N = 203,244; 6.94% deceased) and BioBank Japan participant survival (N = 135,983; 24.1% deceased), to provide additional evidence of their association with human ageing traits in independent samples. We found evidence for a directionally concordant and statistically significant effect for two of the 23 loci tested. As is common for genome-wide association analyses, larger samples may be needed to confirm or reject the remaining loci.
Randomization	Recruitment and analysis of the cohorts used to generate the GWAS summary statistics are described in the Supplementary Note and referenced publications. As is standard, GWAS analyses rely on observational data which is not randomised but is corrected for population stratification and cryptic relatedness using genetic principal components as fixed effects and/or genetic kinship as random effects. Our study does not compare experimental groups, so additional randomisation is not applicable.
Blinding	Blinding was not relevant to our study: analysis of genetically independent phenotypes does not use group allocation. GWAS of genetically independent phenotypes is a hypothesis-free scan of the genome that does not require sample randomisation.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Description of the population characteristics of ageing-related GWAS are described in the Supplementary Note and referenced publications. For individual-level Biobank Japan (BBJ) data, detailed sample descriptives can be found in Hirata et al (2017) J Epidemiol. <a href="https://doi.org/10.1016/j.je.2016.12.006">https://doi.org/10.1016/j.je.2016.12.006</a> . A total of 135,983 unrelated Japanese-ancestry individuals (56% male) were used for our analysis, recruited at a median age of 64, of which 30,976 had a death record at time of analysis.
----------------------------	---

## Recruitment

BBJ individuals were recruited from June 2003 to March 2008 at 12 medical institutes if they presented with one of 47 common diseases in Japan. Diseases were identified by attending physicians and independent medical coordinators provided information to patients. As diseases did not include cognitive impairments, we expect little self-selection bias. However, while the survival model was adjusted for this disease status, the selection of unhealthy individuals could decrease SNP hazard ratios (as the baseline mortality hazard is higher). This potential bias is towards the null, causing our replication of GWAS findings to be conservative.

## Ethics oversight

Biobank Japan participants provided written informed consent and survival study protocols were approved by Biobank Japan Project ethical review boards from the Institute of Medical Sciences, the University of Tokyo, and the RIKEN Center for Integrative Medical Sciences. All other data was publicly available and approved by ethical committees as described in their respective publications. See Supplementary Note for FinnGen ethical approval

Note that full information on the approval of the study protocol must also be provided in the manuscript.