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

Mendelian randomisation of genetically independent ageing phenotypes identifies LPA and VCAM1 as biological targets for human ageing

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

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

Corresponding author: Paul Timmers, paul.timmers@ed.ac.uk
Keywords: GWAS, GIP, ageing, frailty, healthspan, lifespan, Mendelian randomisation.

Abstract

Length and quality of life is important to us all, yet identification of promising drug targets for human ageing using genetics has had limited success. Here, we combine six European-ancestry genome-wide association studies of human ageing traits—healthspan, father and mother lifespan, exceptional longevity, frailty index, and self-rated health—in a principal component framework that maximises their shared genetic architecture. The first principal component (Ageing-GIP1) captures both length of life and indices of mental and physical wellbeing. We identify 27 genomic regions associated with Ageing-GIP1, and provide additional, independent evidence for an effect on human ageing for loci near *HTT* and *MAML3* using a study of Finnish and Japanese survival. Using proteome-wide two-sample Mendelian randomisation and colocalisation, we provide robust evidence for a detrimental effect of blood levels of apolipoprotein(a) (LPA) and vascular cell adhesion molecule 1 (VCAM1) on Ageing-GIP1. Together, our results demonstrate that combining multiple ageing traits using genetic principal components enhances power to detect biological targets for human ageing.

Introduction

Ageing 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[1]. The factors determining the rate of ageing and death are complex and interlinked, and include genetics, lifestyle, environmental exposures, and chance.

Quantifying the ageing process is not straightforward. A variety of ageing-related phenotypes have been studied as proxies, from chronological measurements such as the length of time from birth until occurrence of a major disease (healthspan)[2] or death (lifespan)[3,4], to cellular deterioration measurements such as telomere attrition[5] and loss of Y chromosome[6,7], to holistic measurements such as the frailty index[8], encompassing multiple functional impairment indicators[9]. While the genetic component of these ageing-related traits tends to be estimated at less than 15%[2,8,10], recent progress has been made on characterising this component using large genome-wide association studies (GWAS), and combining similar GWAS to increase statistical power[11,12].

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

However, to date, large ageing-related trait GWAS have only been combined using MANOVA, which detects genetic variation that is either shared between multiple traits or strongly associated with a single trait[13,14]. This mixture generates heterogeneous SNP effect sizes, complicating the downstream analysis[11]. 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)[15]. As their name implies, GIPs capture the genetic covariance between phenotypes while being genetically uncorrelated to each other. In practice, this means the first principal component (GIP1) maximises the genetic overlap between all traits, while each subsequent GIP contains genetic variation distinguishing the traits from each other[15].

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

Results

Ageing-related traits cluster based on genetic correlations

We gathered publicly available GWAS summary statistics on ageing-related traits measured in at least 10,000 European-ancestry individuals. These included self-rated health[16], healthspan[2], father lifespan[4], mother lifespan[4], exceptional longevity[17], frailty index[8], perceived age[18], Hannum epigenetic age acceleration[19], Horvath epigenetic age acceleration[19], telomere length[5], and mosaic loss of Y chromosome[7] (**Supplementary Table 1**). A variety of UK and European individuals are represented between these studies, from children (aged 10+) to centenarians (aged 100+), with birth years spanning the 20th and early 21st 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 the first six traits—mostly chronological and holistic measures of ageing—form a cluster of high genetic similarity ($|r_g| \geq 0.5$; $P < 5x10^{-15}$). In contrast, the cellular measures of ageing 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% CI 0.2–0.8) (**Figure 1; Supplementary Data 1**).

Testing the six ageing-related traits in the main cluster for correlations with 728 other traits from the GWAS-MAP platform[21], we find their genetic similarity may largely be explained through strong, shared genetic correlations (meta $|r_g| \geq 0.5$; FDR < 0.05; $P_{het} > 0.05$; $I^2 < 0.50$) 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 ageing trait also has several genetic correlations that differ substantially (see Methods) from the other ageing-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**).

Ageing-GIPs capture distinct elements of wellbeing

We combined the six GWAS in the main correlation cluster using the loadings from the principal components of the genetic correlation matrix (**Figure 2**), yielding association summary statistics for six Ageing-GIPs (available at https://doi.org/10.7488/ds/2972). As expected, High-Definition Likelihood[20] estimated Ageing-GIP1 to be the most heritable of the Ageing-GIPs ($h^2_{SNP} = 0.20$; SE = 0.005), capturing over 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") (**Figure 3**). A leave-one-out analysis confirms that the GIP analysis is highly robust to the selection of GWAS, with the genetic architecture

of Ageing-GIP1 remaining largely the same after excluding any one of the six chronological and holistic ageing-related trait GWAS (range r_g GIP1 with GIP1-resilience = 0.950 [SE 0.024] to r_g GIP1 with GIP1-healthspan = 0.996 [0.023]) (**Supplementary Table 2**).

Apart from the genetic correlations with its component traits, Ageing-GIP1 shows significant correlations with 459 out of 729 traits tested ($P_{adj} < 0.05$, Bonferroni-adjusted for six GIPs and 729 traits). These suggest that, in addition to length of life, Ageing-GIP1 also captures mental and physical wellbeing. For example, among the largest negative Ageing-GIP1 correlations are mental illness, taking medications, and diseases of old age (such as cardiometabolic disorders, cancers, and osteoarthritis) ($r_g \leq -0.5$; $P_{adj} < 1x10^{-9}$). Inversely, fitness and education are among the largest positive correlations ($r_g \geq 0.5$; $P_{adj} < 1x10^{-10}$). Ageing-GIP1 also shows moderate negative correlations with infectious diseases, including N39 (International Classification of Diseases 10th Revision) urinary tract infections ($r_g = -0.66$; 95% CI –0.42 to –0.90; $P_{adj} = 3x10^{-4}$), coughing on most days ($r_g = -0.39$; –0.30 to –0.49; $P_{adj} = 1x10^{-11}$), and severe COVID-19 hospitalisation (i.e. resulting in respiratory support or death) ($r_g = -0.33$; –0.20 to –0.46; $P_{adj} = 0.004$).

However, Ageing-GIP1 also retains some strong correlations with socioeconomic factors, such as smoking behaviours (e.g. current tobacco smoking $r_g = -0.50$; –0.45 to –0.55; $P_{adj} = 4x10^{-92}$) and having a job involving manual or physical work ($r_g = -0.49$; –0.44 to –0.54; $P_{adj} = 2x10^{-89}$) (**Figure 3; Supplementary Data 3**). While such socioeconomic risk factors are partly heritable[22], they could also reflect social stratification or geographic confounding shared between the original six studies used to construct Ageing-GIP1[23,24]. Therefore, as a sensitivity analysis, we used the GIP methodology to remove all genetic correlations with UK Biobank GWAS of household income and socioeconomic deprivation from the Ageing-GIPs—recognising this may remove some of the true signal as well (**Extended Data Fig. 2**). The SNP heritability of the adjusted Ageing-GIP1 phenotype is substantially attenuated ($h^2_{SNP} = 0.09$; SE = 0.003), and consequently, genetic correlations with all phenotypes are reduced. However, most traits which were significant at 5% FDR remain associated with the residual phenotype (**Extended Data Fig. 3**). As expected, the largest reductions in genetic correlations are with socioeconomic, education, and smoking/drinking traits, which are reduced by 0.30 or more (**Supplementary Data 3**).

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

Characterising the genomics of Ageing-GIP1

Across the genome, 27 loci pass the Bonferroni-adjusted genome-wide significance threshold ($P < 5x10^{-8}/6$) in the Ageing-GIP1 GWAS. The strongest lead SNPs in these loci are rs429358 ($P = 2x10^{-40}$) and rs660895 ($P = 2x10^{-22}$), located nearest to the *APOE* and *HLA-DRB1/DQA1* genes, respectively (**Figure 4; Table 1**). Genetic fine-mapping with SuSiE[25] highlights three independent 95% credible sets of causal variants for the *APOE* locus: the *APOE* ε4 and ε2 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 non-synonymous *ANAPC4* exon variant), and rs2165702 (an intergenic variant near *PHB*) (**Supplementary Table 3; Supplementary Data 4**).

The majority of lead SNPs are strongly associated with self-rated health and resilience, in line with the large loadings of these traits in the construction of Ageing-GIP1 (**Figure 4; Supplementary Table 4)**. Seventeen Ageing-GIP1 loci overlap with genome-wide significant ($P < 5x10^{-8}$) loci from the original ageing-related trait GWAS, showing evidence of a shared causal variant (coloc PP ≥ 80%) (**Supplementary Table 5**). For the other loci, the GIP framework either increases power (N = 4) or indicates 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[2–4]. 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 ageing traits in independent samples. Combining both cohorts to achieve adequate power, we find Ageing-GIP1-increasing alleles of lead SNPs near *HTT* and *MAML3* have a protective effect on survival in these cohorts (one-sided $P < 0.05/23$, taking into account the 23 loci tested), increasing average lifespan by around 2.53 (95% CI 0.91 to 4.15; one-sided P = 0.001) and 2.51 (0.79 to 4.23; one-sided P = 0.002) months per allele, respectively. While we are underpowered to confirm the remaining 21 loci individually, we find that collectively, their Ageing-GIP1-increasing alleles are also associated with increased Finnish and Japanese survival (one-sided P = 0.044) (**Supplementary Table 6**).

Again, we find Ageing-GIP1 genetics are stable when performing a leave-one-out analysis: lead SNP effects of most Ageing-GIP1 loci do not change significantly when excluding one of the six ageing traits. The exceptions are for previously replicated loci and *SLC22A1/A2*, where exclusion of one of the traits can reduce the Ageing-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 Ageing-GIP1 effect sizes are attenuated but, in all cases, remain associated with the residual trait ($P < 0.05/27$) (**Extended Data Fig. 4; Supplementary Data 7**).

We further looked up all lead Ageing-GIP1 SNPs and close proxies ($r^2_{EUR} \geq 0.8$) in PhenoScanner[26] and the GWAS catalog[27], excluding associations discovered solely in UK Biobank, which showed 24 out of 27 Ageing-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 Ageing-GIP1-increasing alleles are associated with a reduction in multiple diseases but do not appear to associate with socioeconomic factors, suggesting these loci largely capture intrinsic sources of ageing (**Supplementary Data 8**).

Aggregating SNP association statistics across the genome into gene scores using the Pathway Scoring Algorithm (PASCAL)[28], we find high scoring genes for Ageing-GIP1 (**Supplementary Data 9**) appear overrepresented in the haem metabolism hallmark gene set, as well as 300 gene ontology pathways (FDR 5%), regardless of adjustment for socioeconomic correlations. These gene ontology pathways cluster into 20 groups, related to neuronal development, organisation, and function; transcriptional regulation; chemical homeostasis; cellular growth, differentiation, and apoptosis; proteolysis; protein phosphorylation; intracellular signalling and transport; immune system development; the muscle system; and lipoprotein metabolism (**Supplementary Data 10**). Similarly, Ageing-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 Ageing-GIP1

Mendelian randomisation (MR) uses genetic variation as instrumental variables to estimate the casual effect of exposures of interest on an outcome[29]. We used a set of well-validated blood protein quantitative trait loci (pQTL) for 857 proteins[30] as genetic instruments in a two-sample MR[29] and colocalisation[31] framework to infer putative causal links between protein levels and Ageing-GIP1 (**Supplementary Data 11)**. We find robust evidence for a detrimental effect on Ageing-GIP1 (FDR < 5%) for the levels of four proteins in blood (**Table 2**), with pQTL instruments passing sensitivity and causal directionality tests, and additionally colocalising with the Ageing-GIP1 signal (see Methods). Three of these proteins—apolipoprotein(a) (LPA), olfactomedin-1 (OLFM1), and LDL receptor related protein 12 (LRP12)—were instrumented by a cis-pQTL and were encoded by genes that appeared significantly enriched in the gene score analysis. The remaining protein, vascular cell adhesion molecule 1 (VCAM1), was instrumented by a trans-pQTL shared with beta-2-microglobulin (B2M); however, only VCAM1 protein levels colocalised with the signal at this locus (**Supplementary Data 11**). When performing the same MR analysis on Ageing-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 Ageing-GIP1 ($P_{MR} = 2x10^{-8}$), with an increase of one standard deviation in genetically predicted blood protein levels causing a decrease of 0.035 (95% CI 0.025–0.045) standard deviations in Ageing-GIP1.

This significance appears to be driven by a consistent detrimental effect across all six Ageing-GIP1 component traits (β_{MR} range 0.013 to 0.035; all nominal $P < 0.05$) (**Figure 5**). For a sense of scale, when performing the same MR analysis on the unstandardised parental lifespan GWAS, this equates to a loss of approximately 7 months of life (95% CI 5 months to 9 months) per standard deviation increase in LPA blood levels (**Supplementary Table 8**). The effects of the remaining proteins on Ageing-GIP1 are larger in magnitude but appear unequally distributed across Ageing-GIP1 component traits. We estimate a decrease of one standard deviation in genetically predicted VCAM1 blood protein levels is associated with an increase of approximately 18 months of life (13 months to 23 months); however, comparing VCAM1 effects on standardised Ageing-GIP1 components, we find its effect to be largest on mid-to-late life ageing traits (lifespan, longevity, resilience) compared to early-life ageing traits (healthspan). Similarly, the effects of OLFM1 and LRP12 appear mostly mediated through late-life ageing traits (resilience and/or longevity). (**Figure 5**).

Discussion

We combined European-ancestry GWAS of healthspan, father lifespan, mother lifespan, longevity, frailty, and self-rated health using a framework which maximised power to detect associations with their shared genetic component while down-weighing trait-specific genetic associations. The resulting Ageing-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 Ageing-GIP1, including two novel loci near *HTT* and *MAML3* that showed directionally consistent evidence of an effect on survival in two additional, independent samples. Across the genome, we found genes enriched for association with Ageing-GIP1 to be overrepresented in haem metabolism and pathways related to (among others) neurogenesis, homeostasis, proteolysis, immunity, and the muscle system in human ageing. Lastly, we performed MR of predicted blood protein levels on Ageing-GIP1, which revealed the levels of LPA, VCAM1, OLFM1, and LRP12 proteins may be detrimental to multiple indices of healthy ageing.

LPA is a glycoprotein making up the main component of large lipoprotein(a) particles. It is a well-known risk factor for atherosclerotic disease[32] and a target of ongoing clinical trials with regard to cardiovascular outcomes[33]. 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[34], in line with our results. Our analysis suggested the detrimental effect of LPA may apply to ageing more generally, and stringent colocalisation and reverse MR tests provided additional evidence for causality. Interestingly, human endothelial cell culture experiments show addition of LPA increases cell surface expression of VCAM1[36] and can increase endothelial cell contraction and permeability[37].

VCAM1 is a cell adhesion glycoprotein localised predominantly on endothelial cell surfaces and its expression is upregulated in response to inflammatory signals, which mediates adhesion and transduction of leukocytes across endothelial walls[38]. The link we established between VCAM1 and human ageing relied on a trans-pQTL instrument shared with B2M and is therefore more susceptible to horizontal pleiotropy. That is, the genetic variant may influence VCAM1 levels indirectly, and its effect on human ageing traits could be caused by factors independent of VCAM1 levels. However, only VCAM1 colocalised with the pQTL signal, and experimental evidence from mouse studies suggests the effect of VCAM1 on ageing is likely to be causal. Specifically, VCAM1 levels in blood are known to increase with age in both humans and mice[39] and treatment with anti-VCAM1 antibodies or an inducible deletion of *Vcam1* improves cognitive performance of aged mice[39]. Of note, similar results have been found for B2M abundance and mouse knockouts[40], and as such, identification of robust genetic instruments for B2M levels is also warranted.

We were unable to robustly assess colocalisation and reverse causality for the LRP12 and OLFM1 signals as genome-wide association summary statistics were not available, so the effects of these proteins on human ageing should be interpreted with additional caution. LRP12

belongs to the LDL receptor superfamily and may play a role in brain development[41] and both tumour proliferation and suppression, depending on the tissue[42,43]. Similarly, OLFM1 is a glycoprotein involved in neuronal development and maintenance[44] which has also been shown to suppress colorectal tumour metastasis[45]. In mouse models, *Olfm1*-knockouts showed reduced cerebral infarction and fertility[46]. However, it is unclear if reduction of LRP12 or OLFM1 blood levels will have a beneficial effect in humans.

Importantly, our MR analysis was restricted to blood pQTL, precluding detection of causal effects of protein levels in other tissues. While it is likely that there are proteins with tissue-specific effects on ageing, particularly in the brain, the samples needed for detection of such pQTL are less readily available and therefore more difficult to study at scale (although progress is being made[47]). Still, the blood may be a particularly suitable tissue to identify ageing-related proteins. Connecting the circulatory systems of two mice of different ages has been shown to accelerate signs of ageing in the brain, muscle, and liver of the young mouse and reverse similar signs in the old mouse[48,49]. Likewise, the detrimental cognitive effect of injecting old blood in mice is counteracted when anti-VCAM1 antibodies are concomitantly injected[39]. As such, the blood currently remains one of the most promising tissues to detect ageing-related proteins.

Haem metabolism and iron levels were previously hypothesised to play role in human ageing[11], and here we identify the same pathway using new methods and additional data. Interestingly, both the haem pathway and the proteins we uncovered using MR are strongly linked to vascular and endothelial damage[50]. Across the genome we also found an enrichment for brain tissues and pathways related to neuronal integrity. Given endothelial cells are central to both the cardiovascular system and the blood-brain barrier[51,52], and endothelial function declines with age[53], progressive endothelial dysfunction may manifest itself as an age-related disease. Indeed, recent findings suggest the detrimental effects of the *APOE* ε4 allele—the largest genetic determinant of human ageing—are mediated by an accelerated breakdown of the blood-brain barrier, independently from amyloid-β and tau accumulation[54]. 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 GWAS 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* only reached genome-wide significance in the combined analysis, and validation of *MAML3* in an external survival cohort suggests these findings may hold across populations and ageing measurements. In addition, *USP28* within the *USP28/HTR3B* locus resembles *CDKN2A* within the well-known *CDKN2B/-AS1* locus, as both have been implicated in cellular proliferation[55,56] and senescence[57,58]. In fact, shRNA knockdown of *USP28 in vitro* results in decreased *CDKN2A* expression[57]. However, additional validation and functional investigation of Ageing-GIP1 loci are warranted.

A secondary advantage of the GIP method is the high stability of the resultant Ageing-GIP1 GWAS, which appears largely robust to the selection of component traits. Nevertheless, we note our analysis focused on chronological and disease measures of ageing and did not include the cellular measures of ageing due to their modest genetic correlations and lack of power. Whether Ageing-GIP1 loci and pathways accurately reflect the ageing process overall or only capture specific domains of ageing is unknown. Inclusion of future ageing-related GWAS performed on biobank-scale samples should provide insight into how well Ageing-GIP1 captures human ageing.

An important limitation of the Ageing-GIP1 GWAS, and GWAS of ageing-related traits more generally, is the potential confounding of ageing 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[24]. In our study, we found genetic correlations of the six ageing-related GWAS 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 Ageing-GIP1. Indeed, when explicitly removing genetic correlations with two UK Biobank GWAS of socioeconomic factors, Ageing-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 Ageing-GIP1 loci—including loci which have been confidently linked to ageing-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 the adjustment removed the unspecific polygenic background induced by UK Biobank microstructure and social confounding. We therefore consider this procedure a suitable approach for confounder correction, although it may be somewhat conservative for also removing the genetic background of the adjusting traits that may genuinely be associated with ageing. An alternative solution to the confounding problem could be the inclusion of a more ancestrally diverse set of ageing-related GWAS from a larger variety of populations.

Despite these limitations, modelling the shared genetic component of human ageing proxies has allowed us to down-weigh non-biological features and propose pathways and proteins which may causally influence the human ageing process. We share the full Ageing-GIP1 summary statistics without restrictions[59] 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 GWAS of ageing 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. GWAS meeting inclusion criteria included extreme longevity[17] (survival past 90th percentile), father and mother lifespan[4], healthspan[60], self-reported health[16], frailty index[8], epigenetic age acceleration[19], telomere length[5], mosaic loss of Y chromosome[7], and perceived age[18]. As the self-reported health GWAS used the first release of UK Biobank data (N ~ 150,000), we looked up the same phenotype in the Neale Lab GWAS collection (N ~ 500,000), which had a larger sample size but was otherwise measured identically[61]. 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 (MAF < 0.5%), or poorly measured SNPs (N individuals < 1% of total). The remaining SNPs were aligned to genome build GRCh37 and harmonised 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[13] using Equation 1 from Winkler *et al.*[62]:

$$Var(Y) = median\left(2pqN * Var\left(\hat{\beta}\right)\right) \text{ (Equation 1)}$$

Where $Var(Y)$ 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 $Var\left(\hat{\beta}\right)$ 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 standardised by dividing effect sizes and standard errors by $\sqrt{Var(Y)}$.

Estimation of genetic and non-genetic correlations

The High-definition likelihood (HDL) R package[20] v1.3.4 was used to calculate SNP heritabilities of ageing-related GWAS and their genetic correlations with each other, using the default European-ancestry LD reference panel and non-MHC SNPs with MAF ≥ 0.01. All GWAS had ≥ 99.9% of the SNPs in this reference panel, except for the Hannum epigenetic age acceleration GWAS which had 96.29%. Pearson correlations between GWAS due to phenotypic similarity and sample overlap were calculated from Z scores of independent SNPs provided by the MultiABEL R package[13] that were non-significant in both studies ($|Z| < 1.96$). The number of SNPs used for this calculation is reported in **Supplementary Data 1**.

LD-score regression (LDSC)[63] v1.0.0 was used to calculate genetic correlations between selected GWAS summary statistics from the GWAS-MAP platform[21] and the ageing-related GWAS and Ageing-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[61,64] and large European-ancestry consortia (e.g. MAGIC, CARDIoGRAM, SSGAC, GIANT) (for an up-to-date list of phenotypes, see https://phelige.com/)[65]. We included GWAS with ≥1 million SNPs, measured in ≥10,000 individuals (if continuous) or ≥2,000 cases and controls (if binary). We further excluded the healthspan and self-rated health GWAS from the GWAS-MAP platform to avoid duplication, after which 728 traits remained. Out of specific interest[66], we also calculated the genetic correlation between Ageing-GIPs and a case-control GWAS of COVID-19 hospitalisation, 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)[67]. GWAS with counterintuitive effect directions were reversed (e.g. satisfaction traits are coded from high to low in UK Biobank), and each set of GWAS summary statistics was filtered prior to analysis by excluding SNPs in the Major Histocompatibility Complex, SNPs with MAF < 1% and, if measured, SNPs with INFO < 90%. A full list of 729 GWAS and references can be found in **Supplementary Data 3**. P values were adjusted for multiple testing using Bonferroni 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 only included GWAS that showed large and significant effects on Ageing-GIP1 ($|r_g| > 0.25$; $P < 0.05/729/6$). Clusters were identified hierarchically by maximising the Bayesian information criterion using the mclust R package[68] v5.4.1, up to a maximum of 100 clusters.

Identification of shared and unique genetic correlations

For each selected GWAS-MAP phenotype, genetic correlations with ageing-related traits were meta-analysed using fixed-effect inverse-variance weighting, with heterogeneity quantified by the Cochran's Q and I^2 statistics, implemented in the meta R package[69] v4.15-1. GWAS-MAP phenotypes which were significantly correlated with all six ageing-related GWAS at FDR 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 ageing-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 ageing-related trait outlier was considered to have a unique genetic correlation with the GWAS-MAP phenotype.

Genetically independent phenotype analysis of ageing GWAS

Principal component loadings for six Ageing-GIP were estimated from the genetic covariance matrix of the six chronological and holistic ageing GWAS, 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:

$$\mathrm{Var}(GIP_i) = \sum[(a_i \otimes a_i) \circ \Sigma_{ph}] \text{ (Equation 2)}$$

Where a_i is the eigenvector of the i-th Ageing-GIP, and Σ_{ph} is the phenotypic variance-covariance matrix of the core ageing traits. As GWAS were standardised, Σ_{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 ageing GWAS statistics were then combined on a SNP-by-SNP basis using the principal component loadings to construct genome-wide summary statistics for the six Ageing-GIPs. GIP effect estimates were calculated by summing effect estimates from the individual ageing-related trait GWAS, each multiplied by their corresponding loading, and standardised by the expected GIP variance (**Equation 2**). Standard errors of the GIP effect estimate were calculated by performing the equivalent calculation using variance arithmetic, also taking into account the phenotypic covariance between GWAS to adjust for sample overlap:

$$\mathrm{SE}(\hat{\beta}) = \sqrt{\sum[(a_i \otimes a_i) \circ \boldsymbol{\Sigma_{ph}} \circ (\mathrm{SE}(\boldsymbol{B}) \otimes \mathrm{SE}(\boldsymbol{B}))]} \text{ (Equation 3)}$$

Where SE($\boldsymbol{B}$) is the vector of SNP standard errors of the core ageing trait effect estimates. Effective sample sizes were then estimated based on the median Z statistic and allele frequencies, i.e. solving Equation 1 for N. Full technical details of the GIP method are described in the **Supplementary Note**.

Ageing-GIP summary statistics were calculated for the 7,324,133 SNPs shared between QC'd GWAS statistics, of which 5,353,660 were common (MAF ≥ 5%) and 1,970,474 were rare (MAF < 5%). Finally, standard errors of each Ageing-GIP GWAS were adjusted to account for the LD-score 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 deprivation index were downloaded from the Neale Lab GWAS collection[61] and subjected to the same SNP filters used for ageing-related trait quality control. Genetic correlations between GWAS of Ageing-GIPs and the two socioeconomic GWAS were calculated using the HDL R package[20] v1.3.4. As before, phenotypic correlations were estimated from the correlation between independent null Z statistics.

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

$$a_i = \{1, -\beta_1, -\beta_2\}$$

Where β refers to $\beta = Cov(\boldsymbol{X})^{-1} \times Cov(GIP, \boldsymbol{X})$. Here, $Cov(\boldsymbol{X})$ is the genetic covariance matrix of UK Biobank GWAS of household income and Townsend deprivation index, and $Cov(GIP, \boldsymbol{X})$ is the genetic covariance matrix between the Ageing-GIP and the two GWAS.

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

Characterisation of genome-wide significant Ageing-GIP1 loci

Genome-wide significant loci were defined as 500 kb regions centred on a lead genome-wide significant SNP ($P < 5x10^{-8}/6$) in linkage equilibrium ($r^2 < 0.1$) with other lead locus SNPs. LD between SNPs was calculated using a random 10k subset of unrelated, genomically British individuals from UK Biobank[70].The susieR R package[25] v.0.11.8 was used to perform genetic fine-mapping of the association signals within each Ageing-GIP1 locus, allowing for up to 10 causal variants. We report posterior inclusion probabilities for SNPs within 95% credible sets. Positional annotations of credible set SNPs were retrieved using ANNOVAR[71] v.2019Oct24.

SNPs within each 500kb locus were looked up in the six ageing-related trait GWAS used to construct the Ageing-GIPs. Any region containing a SNP associated with an ageing-related trait at genome-wide significance ($P < 5x10^{-8}$) was tested for colocalisation between the trait and the Ageing-GIP1 signal. For this, we used the coloc R package[31] v4.0-4 with its default parameters, denoting a posterior colocalisation 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 ageing 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 ageing for the Ageing-GIP1 loci which had not been previously replicated.

Effects were first looked up in a GWAS of survival of FinnGen study participants[72]. This 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) v0.40. SNP effects were log hazard ratios, calculated from a mixed effect frailty model which adjusted for sex, genotyping batch, birth year, and the first ten genomic PCs 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[73,74]. In this analysis, a fixed effect Cox proportional hazards model was fitted using the survival R package v2.41:

$$h(t) = h_0(t)exp(\boldsymbol{X}_1\beta_1 + \boldsymbol{X}_2\beta_2 + \cdots + \boldsymbol{X}_n\beta_n + \boldsymbol{G\gamma})\text{(Equation 4)}$$

Where h(t) is the hazard at time t, given the subject is alive at time t; $h_0(t)$ is the baseline hazard at time t; X_1, X_2, …, X_n are the vectors of covariates with fixed effects β_1, β_2, …, β_n; and ɣ is the effect of the vector of SNP dosages **G**. Covariates were sex, disease status, and the first 20 PCs, where disease status refers to one 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 standard errors by ten[3]. For each SNP, a combined effect was calculated by meta-analysing the cohort-specific effects in a fixed-effect framework (weighted using inverse variance), implemented in the meta R package[69] v4.15-1. One-sided P values were adjusted for multiple testing of 23 loci using Bonferroni correction. The collective effect of the 21 remaining loci was calculated using a random-effect framework, to allow for heterogeneity in effect size estimates.

Ageing-GIP1 leave-one-out sensitivity analyses

Leave-one-out sensitivity analyses were performed for Ageing-GIP1, where one at a time, a core ageing trait was excluded and Ageing-GIP1 loadings and summary statistics were recalculated using the remaining five traits. Genetic and non-genetic correlations were calculated between the original Ageing-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 Ageing-GIP1 loci, we estimated the difference between the lead SNP effect in Ageing-GIP1 and the effect in the leave-one-out GIP1 GWAS, taking into account null correlations between traits. The standard error of the difference in effects was calculated as follows:

$$\mathrm{SE}(\hat{\beta}1 - \hat{\beta}2) = \sqrt{SE(\hat{\beta}1)^2 + SE(\hat{\beta}2)^2 - 2r * SE(\hat{\beta}1) * SE(\hat{\beta}2)} \text{ (Equation 5)}$$

Where is SE(β1) and SE(β2) are the standard errors of the SNP for Ageing-GIP1 and the leave-one-out GIP1, respectively, and r is the non-genetic correlation between GWAS. Statistical significance of the difference was assessed using a Wald test and adjusted for multiple testing of 27 loci using Bonferroni correction.

Lookup of known SNP associations

Lead SNP and close proxies ($r^2_{EUR} \geq 0.8$) of the Ageing-GIP1 loci were looked up in PhenoScanner[26] and the GWAS catalog[27] (accessed 3 December 2020), keeping only the traits with genome-wide significance ($P < 5x10^{-8}$). 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 (e.g. "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 LD-score regression v1.0.0 was used to stratify Ageing-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[63]. We kept only HapMap3 SNPs, excluding the MHC region and SNPs with MAF < 0.05, and used the 1000 Genomes Phase 3 LD score reference as weights. Categories tested included the 10 groups summarising 220 cell-type specific annotations from Finucane *et al.* [63], adjusting for the baseline model (v1.2).

Gene and pathway enrichment

PASCAL[28] was used to aggregate Ageing-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 version 7.2 of the Molecular Signatures Database[75].

Ageing-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 MAF < 5% and SNPs in the MHC region. Gene results passing a 5% FDR threshold were considered significant. For each pathway in the C1 and C5.BP datasets, PASCAL calculated two measures of significance based on chi-squared and permutation statistics. We separately adjusted C1 and C5.BP for multiple testing and considered a pathway with both chi-squared 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 similarity coefficient (i.e. size of the intersection of genes divided by the size of the union of genes). The mclust R package[68] v5.4.1 was used to maximise the Bayesian information criterion of the Jaccard similarity matrix to identify the optimal number of clusters (up to a maximum of 100). We used the number of clusters selected by the majority of mclust models to group the pathways.

Mendelian randomisation of blood protein levels

Genetic instruments for blood protein levels (pQTL) were retrieved from Zheng *et al.*[30]. We included all Tier 1 instruments: cis- and trans-pQTL shown to influence five or fewer proteins (specificity) with no evidence of heterogeneity in effect sizes between multiple protein expression studies (consistency). A total of 857 proteins had non-palindromic SNP instruments (898 total pQTL) present in the Ageing-GIP1 summary statistics. Two-sample MR of blood protein levels as exposures and Ageing-GIP1 as outcome was performed using the TwoSampleMR R package[29] v0.5.5. 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 standardised Ageing-GIP1 component traits as outcome, as well as the unstandardised, combined parental lifespan GWAS from Timmers *et al.*[4] to provide an intuitive measure of the effect. The MR effects and standard errors from the latter were multiplied by 10 to convert them from units of negative log hazard ratio to approximate years of life[3].

Ageing-GIP1 MR results passing a 5% FDR threshold and sensitivity tests ($P_{Het} > 0.05$ and $P_{Egger} > 0.05$; if applicable) were taken forward for follow-up colocalisation tests to rule out LD linkage. For proteins for which we had access to genome-wide summary statistics (e.g. LPA, B2M, and VCAM1), we used the coloc R package[31] v4.0-4 to perform colocalisation analysis using the default parameters, and denoted PP ≥ 80% as evidence of a shared signal. For instruments without summary statistics, we performed an LD check as described in Zheng

et al.[30], which involved checking whether any of the 30 strongest Ageing-GIP1 SNPs in a 1 Mb region centred on each pQTL were in high LD ($r^2_{EUR} \geq 0.8$) with that pQTL.

Finally, proteins passing sensitivity and colocalisation tests were subjected to a reverse-causality test using MR-Steiger[76], and, if genome-wide summary statistics were available, a bi-directional MR analysis, both implemented in TwoSampleMR. For the bi-directional MR, we used up to 27 genome-wide significant lead SNPs from Ageing-GIP1 (shared between GWAS 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 bi-directional MR ($P > 0.05$) (if applicable) were considered to have robust causal effects on Ageing-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, 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.

Statistics and reproducibility

The statistical methods used to analyse the data are described fully in the Methods subsections above, with basic data processing done using R version 3.6.0 unless specified otherwise. We used the independent FinnGen and Biobank Japan cohorts to successfully replicate two of the 23 new ageing-related loci. Predetermination of study sample size, randomisation of experiments, and blinding of investigators to experiments were not applicable for our type of study.

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.phenoscanner.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.

Code availability

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://mrcieu.github.io/TwoSampleMR/

This code can be accessed without restrictions.

Acknowledgements

We thank the UK Biobank Resource, approved under application 19655. We would like to acknowledge funding from the Medical Research Council Human Genetics Unit (P.R.H.J.T., J.F.W.: MC_UU_00007/10), the University of Edinburgh (P.K.J.), the Russian Ministry of Education and Science 5-100 Excellence Programme (E.S.T., Y.A.T.) and the Ministry of Education and Science of the RF via the Institute of Cytology and Genetics SB RAS (E.S.T., Y.A.T.: 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 (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), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.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), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta) and Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi) and FinBB (https://finbb.fi/). The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195, HHSN268201500001I, and 75N92019D00031 (Boston University). This project was funded in part by the Division of Intramural Research, National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH), Bethesda, MD. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. 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 licence to any Author Accepted Manuscript version arising from this submission.

Author Contributions

P.R.H.J.T.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing—Original draft preparation, Writing—Review & editing, Visualization, Project administration. E.S.T.: Formal analysis, Writing—Review & editing. S.S.: Validation, Formal analysis. M.A.: Validation, Formal

Analysis. T.T.J.K.: Validation, Formal analysis. W.Z.: Validation, Formal Analysis. S-J.H. Formal analysis. C.Y.: Formal analysis. J.D.: Formal Analysis, Writing—Review & editing. D.L.: Resources, Writing—Review & Editing, Supervision. A.G.: Resources, Supervision. Y.K.: Resources, Data Curation. Y.O.: Resources, Supervision. P.K.J.: Writing—Review & editing, Supervision. J.F.W.: Resources, Writing—Review & editing, Supervision, Funding acquisition. Y.A.T.: Conceptualization, Methodology, Software, Formal analysis, Writing—Review & editing, Supervision.

Competing interests

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

Tables

Nearest gene(s)	rsID	Chr	Position	A1	Freq1	β1 (SE)	P	Het
NEGR1	rs2815748	1	72816147	G	0.20	0.028 (0.004)	$7x10^{-10}$	
PHTF1	rs1230682	1	114293526	A	0.63	0.023 (0.004)	$4x10^{-10}$	
AFF3	rs7609078	2	100490363	A	0.37	0.023 (0.004)	$3x10^{-10}$	
TRAIP	rs2271961	3	49878113	T	0.50	0.027 (0.004)	$8x10^{-14}$	
ADD1	rs16843603	4	2928577	C	0.28	0.025 (0.004)	$1x10^{-10}$	
HTT	rs362273	4	3227419	G	0.32	0.024 (0.004)	$4x10^{-10}$	
ANAPC4	rs34811474	4	25408838	A	0.23	0.025 (0.004)	$3x10^{-9}$	
4q13.2	rs10434248	4	67842921	A	0.57	0.021 (0.004)	$8x10^{-9}$	
MAML3	rs56172573	4	140919381	C	0.38	0.021 (0.004)	$7x10^{-9}$	
C6orf47/GPANK1	rs805262	6	31628733	C	0.52	0.027 (0.004)	$1x10^{-12}$	
HLA-DRB1/DQA1	rs660895	6	32577380	A	0.79	0.043 (0.004)	$1x10^{-22}$	*
SLC22A1/A2	rs9456508	6	160598596	T	0.98	0.082 (0.012)	$4x10^{-11}$	**
LPA	rs118039278	6	160985526	G	0.92	0.045 (0.007)	$1x10^{-11}$	***
MAD1L1/SNORA114	rs11764780	7	2020904	C	0.19	0.029 (0.005)	$3x10^{-10}$	
FOXP2	rs12705966	7	114248851	G	0.34	0.023 (0.004)	$7x10^{-10}$	
CSMD3	rs560719	8	113032374	T	0.50	0.021 (0.004)	$5x10^{-9}$	
CDKN2B/-AS1	rs9632885	9	22072638	G	0.52	0.025 (0.004)	$5x10^{-12}$	**
CTSF	rs2924807	11	66341005	G	0.50	0.021 (0.004)	$5x10^{-9}$	
CCDC90B/DLG2	rs2512690	11	83145469	C	0.34	0.023 (0.004)	$7x10^{-10}$	
TTC12/ANKK1	rs2186800	11	113242860	A	0.54	0.022 (0.004)	$2x10^{-9}$	

USP28/HTR3B	rs61907878	11	113751052	C	0.90	0.037 (0.006)	$1x10^{-9}$	
MIR6074	rs7306710	12	66376091	T	0.48	0.023 (0.004)	$4x10^{-10}$	
12q21.31	rs6539846	12	84811217	A	0.49	0.022 (0.004)	$8x10^{-10}$	
LINC01065	rs8002970	13	53924489	C	0.45	0.024 (0.004)	$6x10^{-11}$	
ZNF652/PHB	rs28394864	17	47450775	G	0.54	0.021 (0.004)	$6x10^{-9}$	
APOE	rs429358	19	45411941	T	0.84	0.066 (0.005)	$2x10^{-40}$	***
ZFP64	rs67442863	20	51031131	T	0.18	0.029 (0.005)	$6x10^{-10}$	

Table 1: Twenty-seven independent genomic loci are associated with the first genetic principal component of ageing-related trait GWAS (Ageing-GIP1). Loci were defined as 500 kb regions centred on a lead genome-wide significant SNP ($P < 5x10^{-8}/6$) in linkage equilibrium ($r^2_{EUR} < 0.1$) 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. Chr—Chromosome. Position—Base-pair position (GRCh37). A1—Effect allele, associated with higher GIP1. Freq1—Allele frequency of the effect allele in UK Biobank. β1—Effect estimate (and standard error) of the A1 allele on Ageing-GIP1 in standard deviation units. P—Two-sided nominal P value from Wald test. Het—Evidence of heterogeneity: asterisks indicate Ageing-GIP1 effect size changes significantly when leaving out one of the core ageing traits from GIP1 calculation (* all but one leave-one-out effects are the same, ** all but two effects are the same, *** three or fewer effects are the same).

Exposure	β_{MR} (SE)	P	FDR	$P_{steiger}$	$P_{reverse}$	Coloc PP	LD check
LPA	**-0.035 (0.005)**	**$2x10^{-11}$**	**$2x10^{-8}$**	**$6x10^{-274}$**	**0.556**	**100%**	**Pass**
VCAM1	**-0.095 (0.019)**	**$7x10^{-7}$**	**$2x10^{-4}$**	**$9x10^{-12}$**	**0.063**	**99%**	**Pass**
OLFM1	-0.111 (0.024)	$3x10^{-6}$	$6x10^{-4}$	$1x10^{-7}$	-	-	Pass
LRP12	-0.078 (0.021)	$2x10^{-4}$	0.033	$4x10^{-10}$	-	-	Pass

Table 2: Mendelian randomisation of genetically predicted blood levels of four proteins suggests they have a causal detrimental effect on Ageing-GIP1. In bold are exposures passing all quality checks, including reverse MR and coloc which required access to genome-wide summary statistics. β_{MR}—MR effect with standard error in parentheses. P—Two-sided nominal Wald test P value for the MR effect. FDR—FDR-adjusted P value taking into account the 857 proteins tested. $P_{steiger}$—Two-sided nominal Wald test P value for the MR Steiger test assessing if the exposure-outcome pair has the correct causal direction. $P_{reverse}$—Two-sided nominal Wald test P value for the MR effect of Ageing-GIP1 on the exposure, i.e. evidence of reverse causality. Coloc PP—Posterior probability of colocalisation estimated by coloc. LD check—A secondary check for colocalisation, which requires at least one of the 30 strongest Ageing-GIP1 SNPs within 500 kb of the pQTL to be in strong LD ($r^2_{EUR} \geq 0.8$) with the pQTL itself. LPA—apolipoprotein(a). VCAM1—vascular cell adhesion molecule 1. OLFM1—olfactomedin 1. LRP12— LDL receptor related protein 12.

Figure Legends

Figure 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 ageing-related phenotype and off-diagonal circles show

the genetic correlation among traits, with both values calculated using High-Definition Likelihood[20]. 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 ageing-related phenotypes used in follow-up analyses. EAA—Epigenetic Age Acceleration. mLOY—mosaic Loss of Y chromosome.

Figure 2: Principal component loadings used to create the genetically independent phenotypes for human ageing. GIP—Genetically independent phenotype. Resilience—Inverse of frailty index. Error bars represent 95% confidence intervals of the loading estimate, inferred using 1000 rounds of Monte Carlo simulations (see **Supplementary Note** for details).

Figure 3: Phenome-wide genetic correlations show Ageing-GIP1 captures both physical and mental wellbeing. a) Table showing the estimated sample size (N_{est}) and SNP heritability (h^2_{SNP}) of each Ageing-GIP GWAS, with standard errors in parentheses. b) Genetic correlations between Ageing-GIP GWAS and the six ageing-related trait GWAS used to construct them. Blank values failed to pass nominal significance ($P \geq 0.05$). c) GWAS of 402 phenotypes measured in European-ancestry individuals from the GWAS-MAP platform have strong and significant associations with GIP1 ($P < 1.5x10^{-5}$ and $|r_g| > 0.25$). These 402 traits are summarised 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 (i.e. the trait with the highest total Z score across Ageing-GIPs). Values failing to pass nominal significance ($P \geq 0.05$) are greyed out. The dendrogram displayed here shows the hierarchical relationship between the most informative trait in each cluster. See **Supplementary Data 3** for the full list of correlations. Genetic correlation P values are derived from a two-sided Wald test with one degree of freedom.

Figure 4: Twenty-seven independent genomic loci are associated with the shared genetic component of ageing-related GWAS. a) Manhattan plot showing associations of SNPs across the genome with Ageing-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 Bonferroni-adjusted genome-wide significance threshold of $5x10^{-8}/6$. 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 $P = 1x10^{-20}$. Loci which exceed this cap and are represented as triangles: *APOE* ($P = 1.5x10^{-40}$) and *HLA-DRB1* ($P = 1.4x10^{-22}$). b) Significance of the association of each lead SNP with the ageing-related trait GWAS used to construct Ageing-GIP1. Bonf-GWS: Bonferroni-adjusted genome-wide significant ($P < 5x10^{-8}/6$). GWS: Genome-wide significant ($P < 5x10^{-8}$). Bonferroni: $P < 0.05$ adjusted for 6 traits and 27 SNPs ($P < 3x10^{-4}$). Nominal: $P < 0.05$. NS: Not significant. P values are derived from a two-sided Wald test with one degree of freedom.

Figure 5: MR of blood protein expression levels on ageing-related GWAS. Exposures which showed causal evidence of an effect on Ageing-GIP1 were tested for association with each of the ageing-related traits used to construct Ageing-GIP1. The x-axis shows the MR effect estimate, with lines representing 95% confidence intervals of the mean, as estimated by TwoSampleMR[3]. The sample sizes of exposure GWAS were 6,861 for LPA; 3,301 for VCAM1; and 3,200 for OLFM1 and LRP12. LPA—apolipoprotein(a). VCAM1—vascular cell adhesion molecule 1. OLFM1—olfactomedin 1. LRP12— LDL receptor related protein 12.

References

1. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194 (2013).
2. Zenin, A. *et al.* Identification of 12 genetic loci associated with human healthspan. *Commun. Biol.* **2**, 41 (2019).
3. 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).
4. 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**, (2019).
5. 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).
6. 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).
7. 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).
8. Atkins, J. *et al.* A Genome-Wide Association Study of the Frailty Index Highlights Synaptic Pathways in Aging. *A Genome-Wide Assoc. Study Frailty Index Highlights Synaptic Pathways Aging* 19007559 (2019) doi:10.1101/19007559.
9. Mitnitski, A. B., Mogilner, A. J. & Rockwood, K. Accumulation of deficits as a proxy measure of aging. *ScientificWorldJournal.* **1**, 323–336 (2001).
10. Ruby, J. G. *et al.* Estimates of the Heritability of Human Longevity Are Substantially Inflated due to Assortative Mating. *Genetics* **210**, 1109–1124 (2018).
11. 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**, 1–10 (2020).
12. Melzer, D., Pilling, L. C. & Ferrucci, L. The genetics of human ageing. *Nature Reviews Genetics* vol. 21 88–101 (2020).
13. Shen, X. *et al.* Multivariate discovery and replication of five novel loci associated with Immunoglobulin G N-glycosylation. *Nat. Commun.* **8**, 447 (2017).
14. Ning, Z. *et al.* Beyond power: Multivariate discovery, replication, and interpretation of pleiotropic loci using summary association statistics. *bioRxiv* (2015) doi:10.1101/022269.
15. Tsepilov, Y. A. *et al.* Analysis of genetically independent phenotypes identifies shared genetic factors associated with chronic musculoskeletal pain conditions. *Commun. Biol.* **3**, 1–13 (2020).
16. Harris, S. E. *et al.* Molecular genetic contributions to self-rated health. *Int. J. Epidemiol.* **46**, 994–1009 (2017).
17. Deelen, J. *et al.* A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat. Commun.* **10**, 3669 (2019).
18. 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).
19. Gibson, J. *et al.* A meta-analysis of genome-wide association studies of epigenetic age acceleration. *PLoS Genet.* **15**, e1008104 (2019).
20. Ning, Z., Pawitan, Y. & Shen, X. High-definition likelihood inference of genetic correlations across human complex traits. *Nat. Genet.* **52**, 859–864 (2020).
21. Gorev, D. D. *et al.* GWAS-MAP: a platform for storage and analysis of the results of thousands of genome-wide association scans. *11-Ая Международная Конференция По Биоинформатике Регуляции И Структуры Геномов И Системной Биологии* 43–43 (2018) doi:10.18699/bgrssb-2018-020.
22. Hill, W. D. *et al.* Molecular Genetic Contributions to Social Deprivation and Household Income in UK Biobank. *Curr. Biol.* **26**, 3083–3089 (2016).
23. Abdellaoui, A., Verweij, K. J. H. & Nivard, M. G. Geographic Confounding in Genome-Wide Association Studies. *bioRxiv* 2021.03.18.435971 (2021) doi:10.1101/2021.03.18.435971.
24. Abdellaoui, A. *et al.* Genetic correlates of social stratification in Great Britain. *Nat.*

Hum. Behav. **3**, 1332–1342 (2019).
25. 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. Ser. B Stat. Methodol.* **82**, 1273–1300 (2020).
26. Kamat, M. A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).
27. 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).
28. 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).
29. Hemani, G. *et al.* The MR-base platform supports systematic causal inference across the human phenome. *Elife* **7**, e34408 (2018).
30. Zheng, J. *et al.* Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).
31. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**, e1004383 (2014).
32. Boffa, M. B., Marcovina, S. M. & Koschinsky, M. L. Lipoprotein(a) as a risk factor for atherosclerosis and thrombosis: Mechanistic insights from animal models. *Clinical Biochemistry* vol. 37 333–343 (2004).
33. 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).
34. 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).
35. Nordestgaard, B. G. *et al.* Lipoprotein(a) as a cardiovascular risk factor: Current status. *Eur. Heart J.* **31**, 2844–2853 (2010).
36. Allen, S. *et al.* Expression of adhesion molecules by Lp(a): a potential novel mechanism for its atherogenicity. *FASEB J.* **12**, 1765–1776 (1998).
37. 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).
38. 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).
39. 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).
40. 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).
41. Grote, A. *et al.* LRP12 silencing during brain development results in cortical dyslamination and seizure sensitization. *Neurobiol. Dis.* **86**, 170–176 (2016).
42. Bethge, N. *et al.* A gene panel, including LRP12, is frequently hypermethylated in major types of B-cell lymphoma. *PLoS One* **9**, e104249 (2014).
43. 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).
44. 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).
45. Shi, W. *et al.* Olfactomedin 1 negatively regulates NF-κB signalling and suppresses the growth and metastasis of colorectal cancer cells. *J. Pathol.* **240**, 352–365 (2016).
46. 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).
47. Cruchaga, C. *et al.* Genomic and multi-tissue proteomic integration for understanding the biology of disease and other complex traits. *medRxiv* 2020.06.25.20140277 (2020) doi:10.21203/rs.3.rs-70284/v1.
48. Rebo, J. *et al.* A single heterochronic blood exchange reveals rapid inhibition of multiple tissues by old blood. *Nat. Commun.* **7**, 1–11 (2016).
49. Katsimpardi, L. *et al.* Vascular and neurogenic rejuvenation of the aging mouse brain by young systemic factors. *Science (80-.).* **344**, 630–634 (2014).
50. Balla, J. *et al.* Haem, haem oxygenase and ferritin in vascular endothelial cell injury. *Nephrol. Dial. Transplant.* **18**, v8–v12 (2003).
51. Higashi, Y. & Yoshizumi, M. Endothelial function. *Nippon rinsho. Japanese J. Clin. Med.* **61**, 1138–1144 (2003).
52. Engelhardt, B. Development of the blood-brain barrier. *Cell Tissue Res.* **314**, 119–129 (2003).
53. 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).
54. Montagne, A. *et al.* APOE4 leads to blood–brain barrier dysfunction predicting cognitive decline. *Nature* **581**, 71–76 (2020).
55. Popov, N. *et al.* The ubiquitin-specific protease USP28 is required for MYC stability. *Nat. Cell Biol.* **9**, 765–774 (2007).
56. Serrano, M. The Tumor Suppressor Protein p16INK4a. *Exp. Cell Res.* **237**, 7–13 (1997).
57. Li, F. *et al.* USP28 regulates deubiquitination of histone H2A and cell proliferation. *Exp. Cell Res.* **379**, 11–18 (2019).
58. Baker, D. J. *et al.* Naturally occurring p16Ink4a-positive cells shorten healthy lifespan. *Nature* **530**, 184–189 (2016).
59. Timmers, P. R. H. J. *et al.* Genetically independent phenotype analysis identifies LPA and VCAM1 as drug targets for human ageing. (2021) doi:https://doi.org/10.7488/ds/2972.
60. Zenin, A. *et al.* Genome-wide association summary statistics for human healthspan (Version 1) [dataset]. (2018) doi:http://doi.org/10.5281/zenodo.1302861.
61. Neale, B. UK Biobank — Neale lab. http://www.nealelab.is/uk-biobank/%0Ahttp://www.nealelab.is/uk-biobank (2018).
62. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
63. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
64. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).
65. Shashkova, T. I. *et al.* PheLiGe: an interactive database of billions of human genotype–phenotype associations. *Nucleic Acids Res.* **49**, D1347–D1350 (2021).
66. Ying, K. *et al.* Genetic and Phenotypic Evidence for the Causal Relationship Between Aging and COVID-19. *medRxiv* 2020.08.06.20169854 (2020) doi:10.1101/2020.08.06.20169854.

67. Initiative, T. C.-19 H. G. 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).
68. 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).
69. Schwarzer, G., Carpenter, J. R. & Rücker, G. *Meta-Analysis with R*. (Springer International Publishing, 2015). doi:10.1007/978-3-319-21416-0.
70. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
71. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
72. 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. *bioRxiv* 2020.10.31.358234 (2020) doi:10.1101/2020.10.31.358234.
73. Hirata, M. *et al.* Overview of BioBank Japan follow-up data in 32 diseases. *J. Epidemiol.* **27**, S22–S28 (2017).
74. 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).
75. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425 (2015).
76. Hemani, G., Tilling, K. & Smith, G. D. Orienting the causal relationship between imprecisely measured traits using genetic instruments. *bioRxiv* **13**, e1007081 (2017).

Biobank Japan Project
Yoichiro Kamatani[9,18]

9) Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
18) Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School
A full list of Biobank Japan Project members and their affiliations appears in the Supplementary Information.

FinnGen
Tuomo T. J. Kiiskinen[8,11], *Wei Zhou[12,8,13],* Andrea Ganna[8,11,12]

8) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
11) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
12) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA
13) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
A full list of FinnGen members and their affiliations appears in the Supplementary Information.