A genome-wide association study identified 10 novel genomic loci associated with intrinsic capacity

Melkamu Bedimo Beyene, Renuka Visvanathan, Robel Alemu, Beben Benyamin, Rudrarup Bhattacharjee, Habtamu B. Beyene, Olga Theou, Matteo Cesari, John R. Beard, Azmeraw T. Amare*

Abstract

In 2015, the World Health Organization introduced the concept of intrinsic capacity (IC), a composite of all the physical and mental attributes that contribute to healthy aging. While substantial evidence supports the biological basis of IC and its subdomains, the extent of genetic influence on IC remains largely unexplored, with no studies currently available. Understanding the genetic basis of IC is crucial to advancing our knowledge and identifying interventions that promote healthy aging. To investigate the genetic basis of IC, we used data from the UK Biobank (UKB; N=44,631) and the Canadian longitudinal study on aging (CLSA; N=13,085). We estimated SNP-based heritability (h^2_{SNP}) at 25.2% in UKB and 19.5% in CLSA. A Genome-Wide Association Study (GWAS) identified 38 independent SNPs for IC across 10 genomic loci and 4,289 candidate SNPs, mapped to 197 genes. Post-GWAS analysis revealed the role of these genes on cellular processes such as cell proliferation, immune function, metabolism, and neurodegeneration, with high expressions in muscle, heart, brain, adipose, and tibial nerve tissues. Of the 52 traits tested, 23 showed significant genetic correlations with IC, and a higher genetic loading for IC was associated with higher IC scores. This study is the first to identify genetic variants and pathways associated with IC, providing a foundation for future research on healthy aging.

Main

Aging is a complex biological process characterized by a gradual decline in biological functions, ultimately impacting health and well-being. These changes increase the risk of chronic conditions, but declines in functioning occur even in the absence of disease. This process is driven by the interplay between genetic, environmental, and lifestyle factors and their interactions, resulting in considerable variability in healthy aging trajectories among individuals. As the global population is aging rapidly, a better understanding of the determinants of healthy aging is becoming increasingly important and could lead to novel and personalized health promotion interventions [1, 2].

In 2015, the World Health Organization proposed a framework for healthy aging focused on building and maintaining the functional ability of older adults. Central to this framework was the concept of intrinsic capacity (IC), which refers to the composite of all the physical and mental capacities an individual can draw upon across the lifespan [3]. Unlike disease and disability, which are relatively late and dichotomous measures of poor health in older adults, IC provides a holistic and continuous framing of health that mirrors recent work on the complex and dynamic biological changes that drive aging. IC is operationalized through five key domains: cognitive, psychological, sensory, locomotor, and vitality capacities [3-8]. IC has been identified as a powerful predictor of many subsequent outcomes, including mortality, care dependence, and a range of chronic diseases, even after accounting for personal characteristics and multimorbidity.

While research on IC has expanded significantly since it was conceptualized, most studies have primarily focused on defining its operational framework and examining associations with different health outcomes, sociodemographic, behavioral, and lifestyle choices, and selected biological markers [9]. To date, no study has specifically explored the genetic basis of IC across the genome, with only one candidate gene study documenting an association with the *APOE4* genotype [10]. This represents a paucity of evidence about the genetics of IC, particularly given the potential for biological factors to inform targeted interventions for maintaining IC and promoting healthy aging. The identification of genetic variants associated with IC has the potential to provide valuable insight into the biological pathways underlying aging and functional ability and for planning personalized interventions targeted to promote healthy aging.

In this study, we address the critical gap in understanding the genetic basis of IC by leveraging data from the United Kingdom Biobank (UKB) and the Canadian Longitudinal Study on Aging (CLSA). We performed a comprehensive analysis that included estimating h^2_{SNP} and conducting GWAS and post-GWAS functional characterization to identify genetic variants associated with IC. The other analyses included pathway analysis, gene-based and gene set analysis, genetic correlation, and polygenic score analyses.

Results

Sample characteristics

After applying quality control (QC) filtering, data from 57,716 individuals were analyzed, comprising 44,631 participants from the UKB with a mean (SD) age of 56 (7.6) years and 13,085 participants from the CLSA with a mean (SD) age of 61 (9.5) years. Just over half of the participants from the UKB (N=24,142; 54.1%) and CLSA (N=6,645; 50.8%) were female. Most participants (94.6% of UKB and 92.1% of CLSA) were of European ancestry. The IC scores in both cohorts have an approximate normal distribution with a mean (SD) of 0.000014 (0.64) in the UK biobank and 0.0005 (0.60) in CLSA. A total of ~8.9 million SNPs, including 5.8 million common SNPs across both cohorts, were included in the final analysis.

Heritability (h^2_{SNP})

The h^2_{SNP} of IC was estimated at 25.2% (95% CI: 23.2 - 27.2%) in the UKB and 19.5% (95% CI: 14.2 - 24.8%) in the CLSA. These estimates reflect the proportion of total phenotypic variance in IC attributable to additive genetic effects. The breakdown of the total phenotypic variance (V_{IC}) into genetic (V_G) and environmental components (Ve) is provided in Supplementary Table 1.

Genome-wide Wide Association Study (GWAS)

Initially conducting GWAS in both cohorts using a mixed linear model method (fastGWA_mlm), we performed an inverse variance weighted fixed effects GWAS Meta-analysis. The Meta GWAS has identified 38 independent genome-wide significant signals ($P < 5x10^{-8}$) across 10 genomic loci on chromosomes 1, 4, 10, 13, 17, and 20, near or within the *PTP4A2, PRPF3, LCORL, RN7SL89P, ANAPC10, HK1, DLEU1, MAPT, SCN4A,* and *STAU1* genes, respectively (Figure 1). The loci with the strongest association, rs9891103, $P=6.50x10^{-14}$, located near the *MAPT* gene, contains 18 of the 38 independent significant SNPs. Details of the 10 genomic loci, lead SNPs, and nearest genes are presented in Table 1 and Figure 1. Regional plots of the 10 loci, illustrating the genomic context and patterns of association, are provided in Supplementary Figure 1.

The LD-based annotation of independent significant SNPs identified 4,289 candidate SNPs mapped to 197 genes. Among the candidate SNPs, 813 SNPs have been reported for their association with phenotypes related to IC domains and ageing. These include certain

biomarkers, cognitive and psychological traits, heart and lung function measures, grip strength, mobility and physical activity, body composition traits, neurodegenerative diseases (e.g., Alzheimer's and Parkinson's disease), and cancer. These findings suggest potential pleiotropy, highlighting shared genetic pathways underlying IC and related traits. The full details of the candidate SNPs associated traits are provided in Supplementary Table 2.

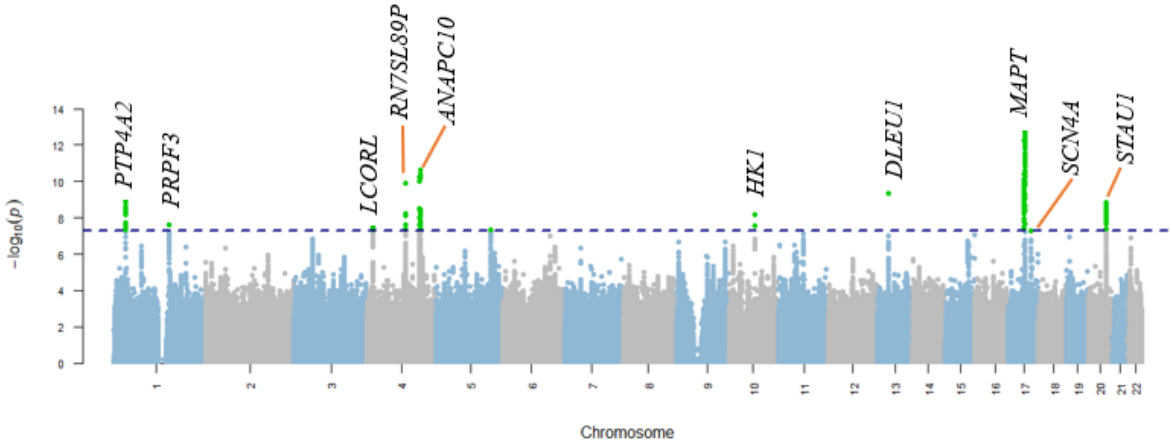


Figure 1: Manhattan plot for the meta-analysis of GWAS summary statistics for IC from the UKB and CLSA GWAS analyses on IC. Each data point represents a SNP, with their positions along the chromosomes (x-axis) plotted against their respective -log10 transformed P-values (y-axis). The plot highlights significant genomic loci associated with IC, with notable peaks surpassing the genome-wide significance threshold cutoff ($P < 5x10^{-8}$) (horizontal blue line).

Table 1: Ten genomic loci associated with IC, lead SNPs, and nearest genes.

Locus	Lead SNPs	Chr	Start	End	A1	A2	MAF (%)	P-value	Genes near to lead SNPs
1	rs72666714	1	32327685	32437775	C	G	8.5	$7.9x10^{-10}$	*PTP4A2*
2	rs11581793	1	150292561	151047648	C	T	0.3	$8.7 x10^{-9}$	*PRPF3*
3	rs2707450	4	17792869	18025484	C	T	28.6	$2.9 x10^{-8}$	*LCORL*
4	rs200768290	4	106073735	106217588	C	T	44.4	$1.3 x10^{-10}$	*RN7SL89P*
5	rs200457388	4	145523071	146177583	C	CA	28.9	$2.4 x10^{-11}$	*ANAPC10*
6	rs72805692	10	71094504	71099109	A	G	3.6	$6.9 x10^{-9}$	*HK1*
7	rs182234180	13	50760125	50993083	C	T	0.8	$4.3 x10^{-10}$	*DLEU1*
8	rs9891103	17	43463493	44874453	C	T	17.7	$6.5 x10^{-14}$	*MAPT*
9	rs2532111	17	61781331	62017421	A	G	40.4	$4.9 x10^{-8}$	*SCN4A*
10	rs34841991	20	47432799	47930866	C	T	15.5	$1.4 x10^{-9}$	*STAU1*

IC = intrinsic capacity, GWAS = genome-wide association study, UKB = UK biobank, CLSA =Canadian longitudinal study on aging.

Gene-based tests and gene set analysis

Using MAGMA gene annotation, the 8.9 million GWAS SNPs were mapped to 19,327 protein-coding genes. Of these, 18 genes were significantly associated with IC after Bonferroni correction (P-value cut off $\leq 2.6 \times 10^{-6}$). The gene-based Manhattan plot (Supplementary Figure 2) highlights the 18 genes significantly associated with IC, and the full list of the gene-based associations can be found in Supplementary Table 3.

A gene expression analysis of each of the 197 prioritized genes identified varying expression patterns across 54 tissue types in the GTEx v8 dataset, with seven genes *(MCL1, JTB, COX6B1, DDX5, PSMC5, PSMD4,* and *PTP4A2)* showing consistently high expression in all 54 tissues (Figure 2). Notably, in the heatmap, some of the genes have high expression in tissues relevant to IC domains, including *MAPT, SEMA6C,* and *SUCLA2* in muscle, *MAPT, PREX1,* and *CRHR1* in the brain, *RAB34, PFKM,* and *SLC25A4* in the heart, and *ICAM2* and *CTSK* in adipose and nerve tissues (Figure 2). The tissue specificity analysis showed a significant over-representation of our prioritized genes set in DEGs for pancreas, skeletal muscle, left ventricle, liver, whole blood, and two brain regions—the Putamen basal ganglia and Hippocampus (Figure 3) and the gene set enrichment analysis (GSEA) in GWAS catalog gene sets found a statistically significant over-representation of our prioritized genes in 38 curated gene sets from previous GWAS studies for different phenotypes and diseases (Figure 4).

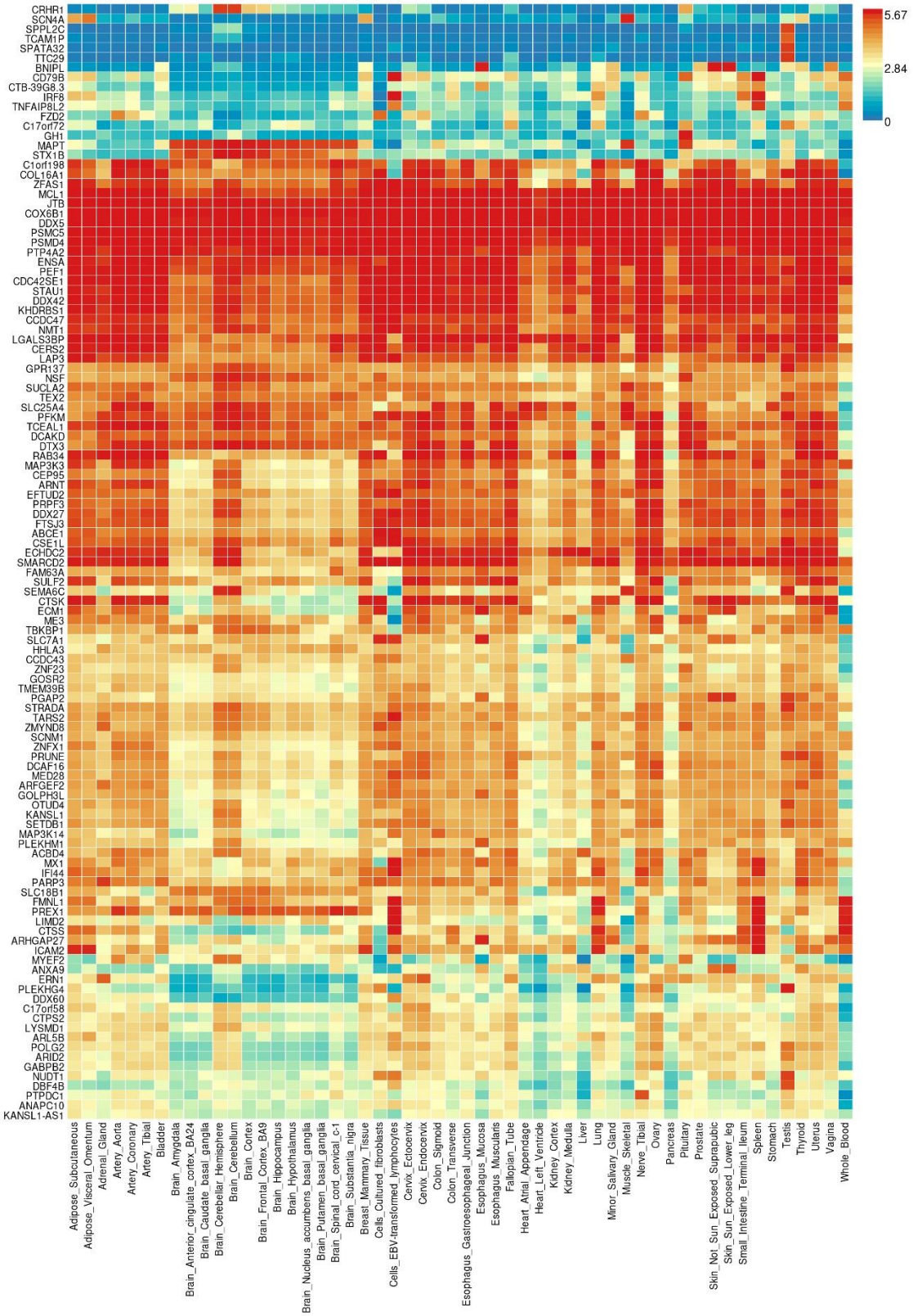


Figure 2: Heatmap showing gene expression pattern – average log 2(RPKM) per tissue for prioritized genes. The color intensity represents the level of gene expression (red shows the highest expression and blue are least expression). *RPKM = Read Per Kilobase per Million*

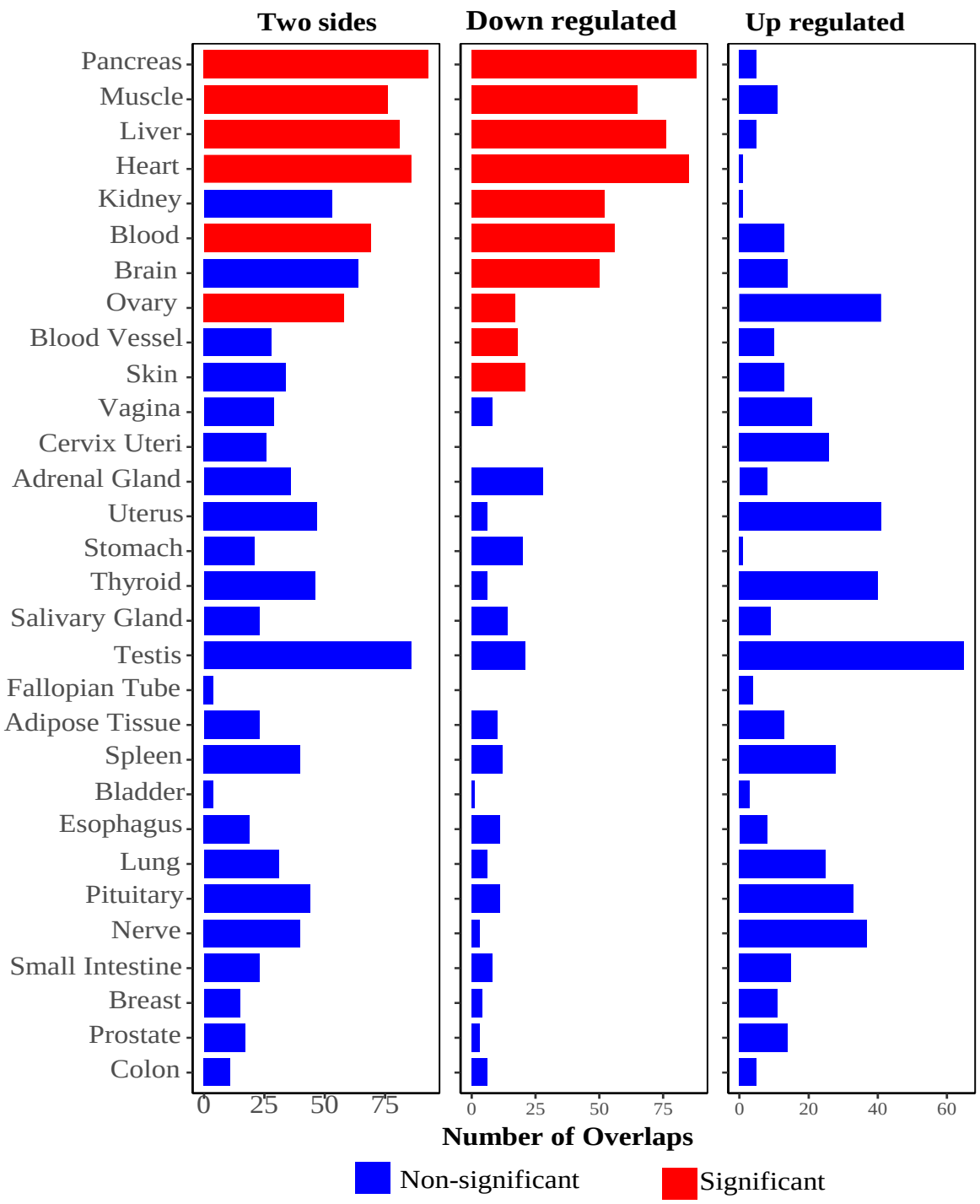


Figure 3: Bar diagram of tissue specificity analysis for our prioritized genes set in differentially expressed gene sets (DEGs) for 30 general tissue types in the GTEx v8 database. Each bar shows the DEGs for the tissues, and the height of the bars shows the number of overlaps. The DEG sets in which our prioritized gene set has a significant over-representation (Bonferroni-corrected P-value < 0.05) are highlighted in red.

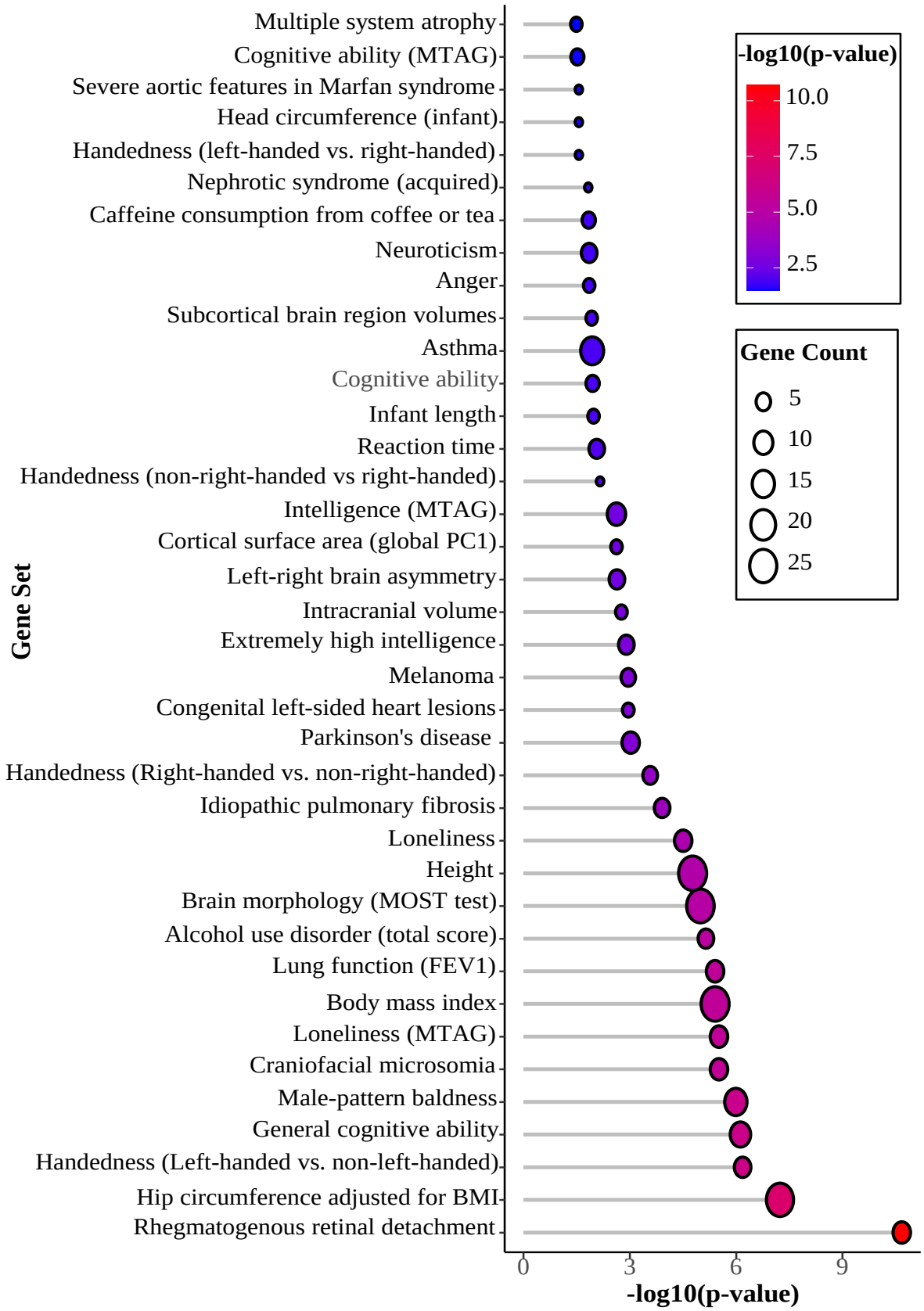


Figure 4: Lollipop plot for GSEA results shows the overlap of our 197 prioritized genes set with gene sets associated with different traits or diseases from the GWAS catalog. On the x-

axis are the -log10 adjusted P-values, while the y-axis represents the gene sets ranked by level of significance, with stronger associations at the bottom. The length of the grey horizontal line indicates the -log10 adjusted P-value, with longer lines signifying greater enrichment significance. Circle sizes in the legend reflect the number of overlapping genes between our prioritized genes and the GWAS Catalog gene sets.

Pathways Analysis

Pathway enrichment analysis performed on 197 prioritized genes identified 88 suggestive pathways (6 from Panther pathways and 82 from reactome pathways) at a nominal significance threshold ($P < 0.05$), although they were not significant after correction for multiple testing. Among these nominally significant pathways include the cell cycle, synaptic vesicle trafficking, mRNA splicing, Alzheimer's disease-presenilin pathway, coenzyme A biosynthesis, ubiquitin-proteasome pathway, transcriptional regulation by RUNX1, programmed cell death/apoptosis, class B/2 (secretin family receptors), TCF-dependent signaling in response to WNT, C-type lectin receptors (CLRs), DC42 GTPase cycle, and RAC1 GTPase cycle (Figure 5, supplementary Table 4).

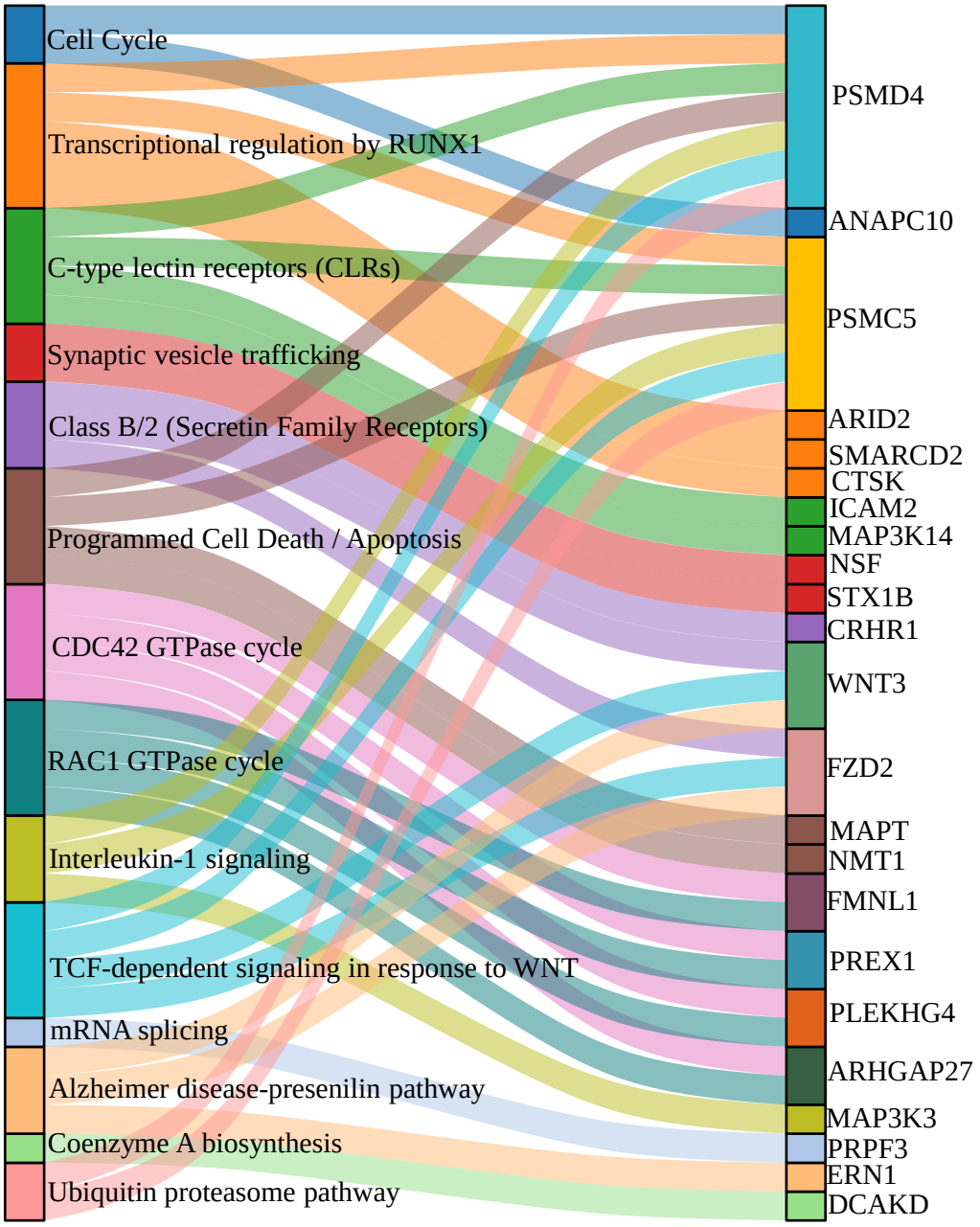


Figure 5: Sankey diagram for selected pathways (left) related to IC-associated genes (right) that showed significant overrepresentation at nominal P-value threshold ($P<0.05$).

Polygenic score analysis

The polygenic score (PGS) for IC was significantly associated with the IC score ($P < 9.75 \times 10^{-16}$). A decile-based comparison revealed a linear trend of association, indicating that individuals with greater genetic loading for IC variants are more likely to have a higher IC score. For example, compared to individuals in the first decile (with a low genetic loading for IC variants),

those in the tenth decile (with a higher genetic loading) had on average 0.19 increased IC score (95%CI: 0.13-0.25) (Supplementary figure 3).

Genetic correlation analysis

A genetic correlation analysis was performed for 52 selected phenotypic traits related to physical, cognitive, and metabolic functions and selected for their potential relevance to IC. Of these, 23 traits demonstrated statistically significant genetic correlations with IC, indicating shared genetic architectures (Figure 6). Notably, the strongest genetic correlations were observed with traits associated with vitality, including forced vital capacity (FVC), forced expiratory volume in 1 second (FEV1), and hand grip strength.

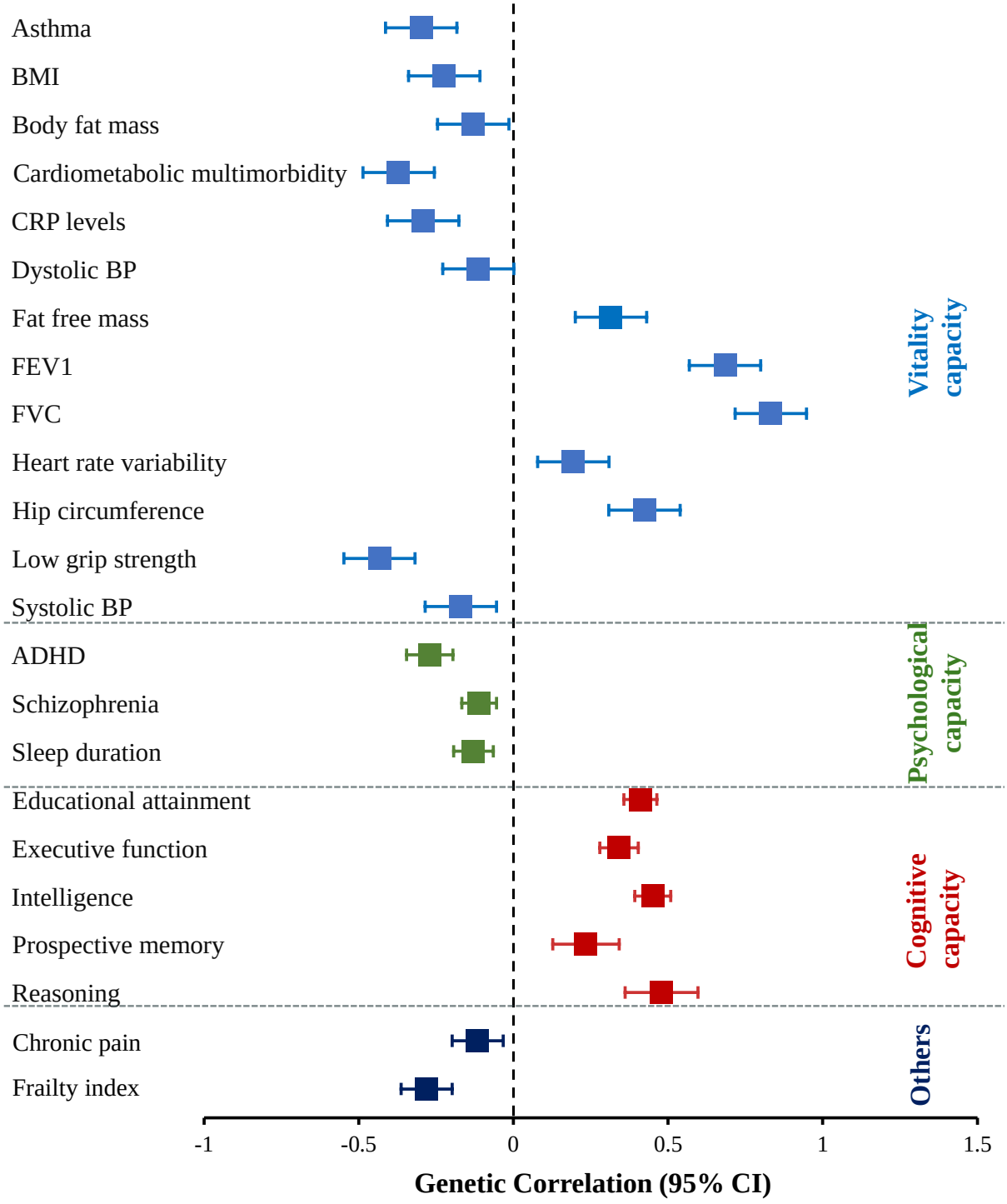


Figure 6: Box plots showing the results of the genetic correlation analyses and 95% confidence intervals between IC and 23 traits with statistically significant associations. The horizontal axis represents genetic correlations (r) ranging from zero (at the centre) to negative correlations on the left, and positive correlations on the right. The vertical axis lists the traits that significantly correlate with IC, organized by categories based on the domains they define. *BP: blood pressure, FVC: Forced vital capacity, FEV1: Forced expiratory volume in one second, CRP: C-reactive protein, BMI: Body mass index, ADHD: Attention deficit/hyperreactivity disorder.*

Discussion

This study presents the first GWAS for intrinsic capacity, providing novel insights into the genetic architecture of this complex trait, with key findings that genetic factors accounted for about a quarter of the total phenotypic variability in IC. We identified 38 independent SNPs across 10 genomic loci, which map to 197 genes expressed in tissues relevant to IC and are significantly overrepresented in gene sets and biological pathways related to IC. The PGS analysis revealed a significant association between PGS for IC and IC scores, implying the polygenic nature of this trait, and genetic correlation analyses found a strong genetic overlap between IC and 23 traits related to IC subdomains and associated phenotypes, highlighting potential pleiotropy.

Our findings indicated that common genetic factors accounted for about a quarter of the total phenotypic variability in IC (h^2_{SNP} estimates ranging from 19.6% to 25.2%), with the remaining (three-quarters) variability related to environmental factors, gene-to-environment interaction effects, rare genetic variants' effects, and stochastic variations. These estimates are comparable to previously reported heritability estimates for traits related to IC domains, including 11%-31% for cognitive function measures, 23.9% for grip strength [11], 9%-16% for functional decline [12], 22%-37% for depression [13], 25%-29% for unilateral hearing difficulty [14], and 32%-36% for FEV1 [15].

Our analysis of genes near the lead variants of IC loci and significant genes from the gene-based test revealed a plausible list of genes involved in biological processes with potential implications for IC, aligning with previous studies. For instance, several genes have been implicated in physiological processes and metabolism, such as *PTP4A2* (near rs72666714), and other Protein tyrosine phosphatases (*PTP*s) play roles in a wide range of fundamental physiological processes [16], while *HK1*(near rs72805692) is central to glucose metabolism and energy production [17]. Similarly, *LCORL* and *GPX1* are metabolic regulators [18, 19] and *MAPT* is linked to glucose metabolism and mitochondrial function [20]. Other genes, such as *RNF123* and *SCARF2*, are associated with lung function (e.g., FEV1 and FVC) and inflammatory markers (e.g., Glycan), with *SCARF2* emerging as a potential therapeutic target for chronic obstructive pulmonary disease [21, 22]. Our findings, together with these supporting evidence, highlight the relevance of the identified genes for vitality capacity-related processes, which in turn suggests a central role that vitality capacity may have for IC - supporting a working definition of vitality as an underlying physiological determinant of IC [23].

Intriguingly, some of the genes are also implicated in psychological and cognitive functioning and neurodegenerative disease phenotypes, such as gene *STAU1* - critical for synaptic function and neuronal plasticity [24], and *MAPT, ARFGEF2, SCN4A,* and *GOSR2* are involved in vesicle trafficking, neuronal functions, and neurological conditions such as epilepsy and seizures [25-28]. Specifically, *MAPT* (which encodes tau protein), the nearest gene to the top hit intron variant rs9891103 in this study, has been implicated in more than 26 tauopathies, neurodegenerative diseases, psychiatric traits, cognitive and brain structure phenotypes [25, 29-32] and *ARHGAP27* has been implicated in various neuropsychiatric conditions [33]. Studies also showed that mutations in *PRPF3* have been discovered to cause visual impairment during old age [34] and Tauopathies (related to the *MAPT* gene) have been studied to have associations with hearing impairment [35] which implies that some of the genes associated with IC in this study may be because of their role in the sensory function. Furthermore, genes such as *MAPT* [29, 36] and *SCN4A* [37, 38] have roles in muscle, bone (osteoblast and osteoclast), and motor functions. As such, their association with IC, supported by previous findings, may imply the role of these genes in locomotive capacity. Taken together, our findings, consistent with previous studies, suggest that the genes identified in this study and implicated in neuropsychological and cognitive functions may have critical roles in the biology of IC.

Additionally, several genes are involved in programmed cell death (apoptosis) and cellular senescence – hallmarks of biological aging. *PTP4A2* (generally *PTP*s) have been reported as emerging regulators of apoptosis [39], *HK1* encodes a protein that inhibits TNF-induced apoptosis at the mitochondria [40] and *STAU1* is involved in amplifying pro-apoptotic activation during the unfolded protein response [41]. Genes such as *AL136115.1* and *CDHR4* are directly implicated in cellular senescence [42, 43]. These findings suggest a potential common biological mechanism underlying IC and the aging process, further supporting the role of these pathways in maintaining functional capacity across the lifespan.

The gene expression and tissue specificity findings also support the top-hit results discussed above. This is implied by the expression of the prioritized genes in tissues related to IC domains and their involvement in aging-related processes such as apoptosis, cellular senescence, telomere shortening, metabolism, and motor and cognitive functions. Seven of the prioritized genes (*MCL1*, *JTB, COX6B1*, *DDX5*, *PSMC5*, *PSMD4*, and *PTP4A2*) showed consistently high expression across all 54 tissues. Of these, *MCL1, JTB, PSMC5,* and *PSMD4* are involved in apoptosis [44-46]; *DDX5* slows telomere shortening and has anti-aging roles [47, 48]; *JTB* and *COX6B1* support metabolism and energy production [45, 49] with potential implications for

vitality capacity. The *NSF* and *STX1B* are key for vesicular transport and neurotransmitter release [50, 51], *PSMC5* is related to cognitive deficits and motor impairments [52], and *SEMA6C* supports neuronal development and synaptic plasticity [53] - all of which are critical for cognitive functioning.

Specifically, some of the prioritized genes were expressed in tissues relevant to IC domains such as muscle (*MAPT*, *SEMA6C*, and *SUCLA2),* brain *(MAPT, PREX1, SEMA6C, and CRHR1),* heart *(RAB34, PFKM, and SLC25A4),* adipose tissue *(ICAM2, CTSK), and* nerve tissue *(IFI44, PTPDC1, and CTSK)*. *MAPT* and *SEMA6C* have been discussed above for having roles in neuronal development and synaptic plasticity. Their expression in muscle and brain tissues emphasizes their involvement in motor and cognitive function aspects of IC. The *CTSK*, expressed in the nerve tissue, has been linked to neuronal plasticity and cognitive function [54].

Similarly, *SUCLA2* has roles in muscle function and energy metabolism – and hence expression in muscle here [55] and *PREX1* is involved in processes such as axon guidance, synaptic plasticity, and neuronal migration [56], which aligns with our finding that it is expressed in brain tissues. *CRHR1* has been reported for its associations with cognitive functioning and mental health [57].

PFKM enhances glycolysis and oxidative phosphorylation [58], and *SLC25A4* maintains mitochondrial energy balance [59] with mutations linked to cardiomyopathy and exercise intolerance [60]. The *ICAM2* and *CTSK* genes, expressed in adipose tissues, have in previous studies been linked to inflammation, adipocyte differentiation, and obesity [61-63]. The expression of these genes in heart and adipose tissues in our study is consistent with these findings, and it may imply their potential roles in vitality capacity.

Tissue specificity analyses also supported the findings from tissue expression analysis that a significant over-representation of our prioritized genes was found in tissues potentially relevant to IC – pancreas, muscle, heart, blood, liver, and brain, and the GSEA of our prioritized genes set in the GWAS catalog gene sets showed significant enrichment in various curated gene sets for traits related to IC domains [5, 9] both highlighting the relevance of the genetic variants we identified for IC.

The pathways identified in this study, with suggestive enrichment, have been implicated in biological aging and likely provide important insight into the biological processes underlying

IC. For example, enriched pathways such as cell cycle (*PSMD4, ANAPC10*), TCF-dependent signaling in response to the WNT pathway (*PSMC5, WNT3, PSMD4, FZD2*), ubiquitin-proteasome pathway (*PSMC5, PSMD4*), programmed cell death/apoptosis (*PSMC5, PSMD4, MAPT, NMT1*) modulate cell proliferation, cellular turnover, and immune function, all of which play key roles in the biology of aging [64-67]. Other enriched pathways, such as coenzyme A biosynthesis (*DCAKD*), class B/2 secretin family receptors (*CRHR1, WNT3, FZD2*), and C-type lectin receptors pathway (*ICAM2, PSMC5, MAP3K14, PSMD4*) are involved in metabolism and immune response functions [68-70] with potential implications for IC domains, mainly the vitality capacity. Likewise, the synaptic vesicle trafficking pathway (*NSF, STX1B*), the cornerstone of learning and memory function [71], and Alzheimer's disease-presenilin pathway (*ERN1, WNT3, FZD2*), a key part of the molecular mechanism of Alzheimer's disease [72] – have roles for cognitive function, and neuronal health, and are linked to neurodegeneration [73, 74].

The PGS analysis findings implied polygenicity of IC, revealing a significant association between IC and PGS for IC, with PGS explaining a small (0.3%) but statistically significant variability in IC. This is not special for IC as existing evidence suggests that for complex traits, the PGS explains only a small amount of variability in the phenotype [75]. It is consistent with findings for some complex traits – such as 0.7-2.1% for major depression and anxiety[76], and 0.4% - 2.1% for frailty in the ELSA cohort [77] but lower than that for traits like educational attainment (up to 12%) or cognitive performance (7-10%) [78].

The results of the genetic correlation analysis indicated how IC was interrelated with multiple traits at the genetic level, suggesting that genetic variants influencing IC also affect these other traits. The strongest genetic correlations were found between IC and traits such as FVC, FEV1, and hand-grip strength, and more than 50% of the significant correlations were with vitality-related phenotypes. These findings imply the central position of the vitality domain for IC and align with the recent definition of vitality as an underlying physiological determinant of IC [23]. Positive correlations with cognitive traits (e.g., intelligence and educational attainment) underscore that genetic factors contributing to cognitive capacity may also contribute positively to maintaining IC, whereas negative correlations with frailty and cardiometabolic multimorbidity, however, may imply that genetic factors that increase the risk of frailty and cardiometabolic multimorbidity will decrease IC, supporting evidence that physical decline accelerates aging-related deterioration [9].

The main strength of this study lies in the use of data from two big international cohort studies with detailed phenotypic and genetic data and the breadth of genetic analyses implemented, ranging from heritability testing and GWAS analyses up to various post-GWAS analyses. However, the study has certain limitations that should be considered when interpreting the findings. First, the UKB and CLSA study cohorts are of predominantly European genetic ancestry, limiting the generalizability of results to other populations. Second, IC is a complex trait encompassing multiple domains, and the genetic relationships within these domains are not well understood, complicating biological interpretation. Third, the lack of standardized IC measurement across studies may pose challenges for replication and external validity. This may be the case for our study between UKB and CLSA cohorts, but we minimized the possible difference between cohorts by using similar or closely related variables, confirming no significant differences in IC scores using appropriate tests and performing a meta-GWAS analysis. Lastly, biological interpretation from gene expression, gene set, and pathway analyses are inherently constrained by the specific databases utilized. Despite these limitations, our study provides foundational insights into the genetic architecture of IC.

In conclusion, this study provides the first comprehensive genetic analysis of IC, demonstrating its heritability and identifying novel genetic variants, genes, and pathways associated with this complex trait. GWAS revealed several distinct genomic loci, many overlapping with IC domains and related traits, suggesting the validity of our findings. Gene expression analyses highlighted the involvement of GWAS-prioritized genes in tissues critical to IC, such as muscle, brain, heart, lung, and adipose, while gene set and pathway analyses implicated significant biological pathways relevant to IC. Overall, these results enhance our understanding of the genetic underpinnings of IC and offer a foundation for exploring biological mechanisms that may serve as a baseline for future research aimed at developing personalized interventions to maintain or improve IC in aging populations. Future research should validate these findings in ancestrally diverse populations and conduct functional studies to elucidate the roles of these genes and pathways. Investigating gene-environment interactions, including the influence of lifestyle, socioeconomic status, and behavioral factors, is essential to understanding how genetic variants affect IC. Additionally, Mendelian randomization studies could establish causal links between genetic variants and IC, helping to identify modifiable factors and inform strategies to mitigate aging-related declines in IC.

Methods

Study Sample

We analysed data from two large international cohort studies, the UKB (N~500,000) and the CLSA (N~50,000). The UKB recruited adults aged 40-69 years from 22 centers across the United Kingdom (UK) between 2006 and 2010, with 94.6% of participants self-identifying as having European genetic ancestry [79, 80]. The CLSA recruited adults aged 45-85 years from all 10 provinces in Canada between 2011 and 2015, with 92.09% self-identified as being of European genetic ancestry [81-83]. Both studies collected comprehensive phenotypic data, including questionnaires, physical assessments, and biomarkers, along with genetic data. The CLSA data has genetic information for the comprehensive cohort only, and we included only the comprehensive cohort participants in this study. For the comprehensive cohort, the participants were recruited from within 25-50 kms of 11 data centres located in 7 provinces.

Development of IC score

The IC score was developed and validated separately for each of the two cohorts. For UKB participants (N=44,631), it was previously reported [5] and using the same methodology, IC scores were computed for the CLSA participants (N = 13,112) – which were participants with full responses to 14 question items used to measure IC. A detailed description of the approach used to generate the IC score in CLSA is presented in Supplementary File 1. Briefly, these IC scores were derived using factor analysis, specifically through a bifactor model within a structural equation model framework. This approach was chosen due to its superior goodness-of-fit statistics compared to conventional methods [9].

Genetic data quality control

Quality control was performed separately on the imputed genetic dataset of the UKB (released in July 2017) [79] and the CLSA (released in September 2018) [81] using consistent SNP and individual-based filtering criteria. First, SNPs with an imputation information score below 0.3, missing call rates exceeding 5%, minor allele frequencies below 1%, and Hardy-Weinberg equilibrium $P < 1x10^{-7}$ were removed. Then, individuals with SNP missing rates exceeding 5%, sex mismatches between reported and genetically inferred sex, chromosome aneuploidy (ambiguous genetically imputed sex), and heterozygous/homozygous rate deviating significantly from the mean (>3SD) were excluded.

After applying these QC filters, the UKB dataset retained 6.4 million SNPs and 44,631 individuals, while the CLSA dataset retained 8.3 million SNPs and 13,085 individuals, all of whom had IC scores and were included in the genetic analysis. Full details of the QC parameters and pipeline are available in Supplementary File 2.

Genetic data analyses

We performed a comprehensive suite of genetic data analyses to explore the genetic architecture of IC, encompassing SNP-based heritability estimation (h^2_{SNP}), GWAS, post-GWAS analyses such as functional mapping and annotation, gene-based and gene set analyses, pathway analysis, polygenic score (PGS) calculation, and genetic correlation analyses (Figure 6). Each analytical method is detailed below:

SNP-based heritability (h^2SNP): The SNP-based heritability of IC was estimated for both cohorts using the restricted maximum likelihood (REML) method implemented in GCTA software [84]. This approach partitions the total phenotypic variance into genetic and environmental components, specifically quantifying the proportion attributable to additive genetic effects of SNPs across the genome [84]. The REML framework leverages genome-wide SNP data to provide robust heritability estimates [84, 85]. The statistical significance of h^2SNP was evaluated using a likelihood ratio test (LRT), contrasting the null hypothesis (h^2SNP =0) with the alternative hypothesis (h^2SNP $\neq$0).

Genome-Wide Association Study (GWAS): We conducted GWAS using a linear mixed model (fastGWA-mlm) implemented in the GCTA software to identify genetic variants associated with IC. This approach adjusts for fixed effects (age, sex, batch effect, and the first ten principal components) and incorporates random effects to account for relatedness among individuals [86]. Population structure was controlled using PCs, while a sparse genetic relationship matrix (GRM), derived from an initial dense GRM with a cutoff of 0.05, was used to model relatedness [84]. The GWAS analysis was performed separately for each cohort, applying a stringent genome-wide significance threshold of $P < 5x10^{-8}$ to correct for multiple testing. Subsequently, a genome-wide meta-analysis (Meta-GWAS) was performed using the summary statistics from the cohort-specific GWAS results. Meta-analysis was conducted using an inverse variance-weighted fixed effects model in Metal Software [84]. Meta-GWAS results provided a comprehensive set of summary statistics, which are used for downstream functional mapping and pathway analyses, described in subsequent sections

Post-GWAS analyses

Functional mapping and annotation: Functional mapping and annotation of candidate SNP were performed using FUMA (Functional Mapping and Annotation) software (Version 1.6.1) [87], leveraging 1000 Genomes Phase 3 reference panel. Independent SNPs were first identified - defined as significant SNPs with $r^2 < 0.6$ in linkage disequilibrium (LD) with one another. These independent SNPs were then used to annotate candidate SNPs (those with LD with independent significant SNPs; $r^2 \geq 0.6$) and delineate genomic loci (by first assigning independent significant SNPs with r>0.1 into the same locus and then assigning independent significant SNPs which are closer than 250kb into same genomic locus). Gene mapping for the candidate SNPs was subsequently conducted using three complementary strategies within FUMA: (1) Positional mapping, which links SNPs to genes based on physical proximity; (2) Expression Quantitative Trait Loci (eQTL), which connects SNPs to genes whose expression levels they influence, using tissue-specific eQTL data; and (3) Chromatin interaction mappings, which identifies SNP-gene connections through long-range chromatin interactions, using data from chromatin conformation studies. These approaches generated a comprehensive list of prioritized genes associated with all candidate SNPs, which were then used as inputs for subsequent gene expression, tissue specificity, and gene set and pathway enrichment analyses.

Gene-based and gene-set analyses: The gene-based test was performed using Multi-marker Analysis of Genomic Annotation (MAGMA, version 1.08), integrated within the FUMA framework [88]. SNPs from the GWAS were mapped to genes based on a 10kb window on either side of a gene, resulting in the identification of 19,327 genes [87]. A genome-wide significance threshold for the gene-based test was set at $p = 0.05/19327 = 2.6\text{x}10^{-6}$. All the other gene-based analyses (i.e., gene expression, tissue specificity, gene set enrichment, and pathway analyses) were conducted using the prioritized genes through FUMA's SNP2GENE annotation.

The gene expression analysis examined the normalized gene expression values measured in reads per kilobase per million (RPKM) for each of the prioritized genes individually across 54 tissue types in the Genotype-Tissue Expression database version 8 (GTEx.v8).

A hypergeometric test with the Benjamini-Hochberg procedure for multiple test adjustment was used for tissue specificity analysis, gene set enrichment analysis (GSEA), and pathway analyses to test the overrepresentation (significant overlap) of our GWAS prioritized genes in differentially expressed gene sets (DEGs), GWAS catalog gene sets, and gene sets for

biological pathways respectively [87, 89, 90]. In tissue specificity analysis, this method tested whether the prioritized genes were significantly enriched in differentially expressed gene sets (DEGs) for various tissue types [87]. DEGs were pre-calculated in the GTEx database (including all genes in the database) using a two-sided t-test comparing the expression of each gene in every tissue to all other tissues. The genes were then considered differentially expressed in each tissue if they met a Bonferroni corrected P-value of ≤ 0.05 and showed a log fold change ≥ 0.58 (equivalent to a 1.5-fold change). In GSEA, it tested if our GWAS prioritized gene set overlapped significantly with GWAS Catalog gene sets (i.e., with gene sets constructed for different traits in previous GWAS studies) [89]. In pathway analysis implemented in the PANTHER platform (incorporating panther and reactome pathways), it was used to test for enrichment of our GWAS prioritized genes set in gene sets built for different biological pathways [90].

Polygenic score analysis (PGS): To assess the cumulative effect of multiple genetic variants on IC, a PGS was computed for IC in the CLSA cohort (as the target sample) using GWAS summary statistics from the UKB cohort (as the discovery sample) [91]. The association of this PGS score with the IC score was tested using a linear regression model and a decile-based stratified analysis was performed to examine patterns of IC across PGS deciles[92-94].

Genetic correlation analysis: To evaluate the genetic relationship of IC with other related traits, a genetic correlation analysis was performed using the LD score (LDSC) regression method [95] that utilized our GWAS summary statistics and LD scores derived from the 1000 Genomes project European genetic ancestry reference panel. A total of 52 traits were included in the analysis, selected based on a literature review of phenotypes likely associated with IC and its domains [9]. The summary statistics for these traits were obtained from the GWAS catalog [24] and details of the specific studies from which the GWAS summary statistics for each trait were obtained are provided in supplementary file 3.

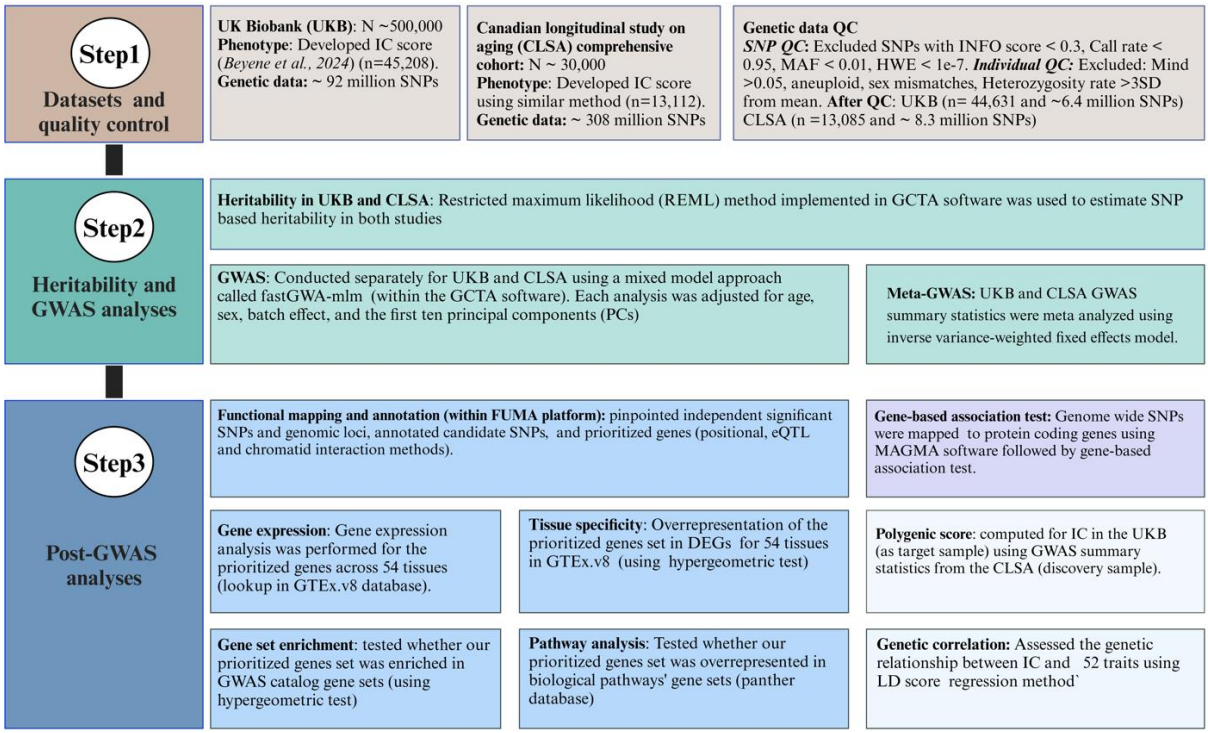


Figure 7: Data analysis flow chart. This chart visualizes an integrated pipeline from data preprocessing and quality control (step 1) through heritability and GWAS (step 2) and various post-GWAS analyses (step 3). *SNP = single nucleotide polymorphisms, IC = intrinsic capacity, QC = quality control, Mind = missing call rate, SD = standard deviation, INFO =imputation quality score, MAF = minor allele frequency, HWE = Hardy Weinberg equilibrium, GWAS = genome-wide association study, Meta-GWAS = Genome-wide association meta-analysis, FUMA = Functional mapping and annotation, eQTL = expression quantitative trait loci, GTEx.v8 = Genotype tissue expression project version 8, DEGs = differentially expressed gene sets, LD = linkage disequilibrium.*

Data Availability

Data, protocols, and other metadata of the UKB and CLSA are available to the scientific community upon request as per their respective data-sharing policies https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access and https://www.clsa-elcv.ca/data-access/. Data for the GWAS summary statistics used for genetic correlation analysis were obtained from the GWAS catalog https://www.ebi.ac.uk/gwas/, and access links for all the 52 summary statistics are available in Supplementary file 3.

Code Availability

The following codes and software we used for data analysis are freely available: Plink2.0 https://www.cog-genomics.org/plink/2.0/ GRM calculation https://yanglab.westlake.edu.cn/software/gcta/#MakingaGRM LDSC Heritability estimation and genetic correlation analysis https://github.com/bulik/ldsc GCTA

https://yanglab.westlake.edu.cn/software/gcta/#Overview fastGWA mixed linear model https://yanglab.westlake.edu.cn/software/gcta/#fastGWA Meta GWAS analysis https://genome.sph.umich.edu/wiki/METAL_Documentation FUMA https://fuma.ctglab.nl/ Bioconductor package for leftover between genome builds https://www.bioconductor.org/install/ polygenic score https://github.com/getian107/PRScsx Panther pathway analysis https://pantherdb.org/ R and Rstudio https://www.r-project.org/ and https://posit.co/download/rstudio-desktop/. Details of customized data analysis codes (in Linux format) for quality control, GRM calculation, heritability estimation, GWAS, Meta GWAS, polygenic score, and genetic correlation analyses can be made available upon request and/or can be found in online supplementary file 2.

References

1. Kirkwood TB. Understanding the odd science of aging. Cell. 2005;120(4):437-47.
2. López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The Hallmarks of Aging. Cell. 2013;153(6):1194-217. doi: 10.1016/j.cell.2013.05.039.
3. Organization WH. WHO clinical consortium on healthy ageing: topic focus: frailty and intrinsic capacity: report of consortium meeting, 1–2 December 2016 in Geneva, Switzerland. World Health Organization; 2017.
4. Belloni G, Cesari M. Frailty and Intrinsic Capacity: Two Distinct but Related Constructs. Front Med (Lausanne). 2019;6:133. doi: 10.3389/fmed.2019.00133.
5. Beyene MB, Visvanathan R, Ahmed M, Benyamin B, Beard JR, Amare AT. Development and validation of an intrinsic capacity score in the UK Biobank study. Maturitas. 2024;185:107976. doi: https://doi.org/10.1016/j.maturitas.2024.107976.
6. Beard JR, Jotheeswaran AT, Cesari M, Araujo de Carvalho I. The structure and predictive value of intrinsic capacity in a longitudinal study of ageing. BMJ Open. 2019;9(11):e026119. doi: 10.1136/bmjopen-2018-026119.
7. de Souto Barreto P, Gonzalez-Bautista E, Bischoff-Ferrari HA, Pelegrim de Oliveira V, Gorga Bandeira de Mello R, Andrieu S, et al. Real-life intrinsic capacity screening data from the ICOPE-Care program. Nature Aging. 2024;4(9):1279-89. doi: 10.1038/s43587-024-00684-2.
8. Beard JR, Hanewald K, Si Y, Amuthavalli Thiyagarajan J, Moreno-Agostino D. Cohort trends in intrinsic capacity in England and China. Nature Aging. 2025;5(1):87-98. doi: 10.1038/s43587-024-00741-w.
9. Beyene MB, Visvanathan R, Amare AT. Intrinsic Capacity and Its Biological Basis: A Scoping Review. The Journal of Frailty & Aging. 2024;13(3):193-202. doi: 10.14283/jfa.2024.30.
10. Meng L-C, Huang S-T, Peng L-N, Chen L-K, Hsiao F-Y. Biological features of the outcome-based intrinsic capacity composite scores from a population-based cohort study: pas de deux of biological and functional aging. Frontiers in Medicine. 2022;9:851882.
11. Willems SM, Wright DJ, Day FR, Trajanoska K, Joshi PK, Morris JA, et al. Large-scale GWAS identifies multiple loci for hand grip strength providing biological insights into muscular fitness. Nature Communications. 2017;8(1):16015. doi: 10.1038/ncomms16015.
12. Christensen K, Gaist D, Vaupel JW, McGue M. Genetic Contribution to Rate of Change in Functional Abilities among Danish Twins Aged 75 Years or More. American Journal of Epidemiology. 2002;155(2):132-9. doi: 10.1093/aje/155.2.132.
13. McGue M, Christensen K. The Heritability of Depression Symptoms in Elderly Danish Twins: Occasion-Specific Versus General Effects. Behavior Genetics. 2003;33(2):83-93. doi: 10.1023/A:1022545600034.
14. Maas IL, Brüggemann P, Requena T, Bulla J, Edvall NK, Hjelmborg JvB, et al. Genetic susceptibility to bilateral tinnitus in a Swedish twin cohort. Genetics in Medicine. 2017;19(9):1007-12. doi: 10.1038/gim.2017.4.
15. Hukkinen M, Kaprio J, Broms U, Viljanen A, Kotz D, Rantanen T, et al. Heritability of Lung Function: A Twin Study Among Never-Smoking Elderly Women. Twin Research and Human Genetics. 2011;14(5):401-7. doi: 10.1375/twin.14.5.401.
16. Tonks NK. Protein tyrosine phosphatases: from genes, to function, to disease. Nature Reviews Molecular Cell Biology. 2006;7(11):833-46. doi: 10.1038/nrm2039.
17. Bucci C, Bakke O, Progida C. Charcot–Marie–Tooth disease and intracellular traffic. Progress in Neurobiology. 2012;99(3):191-225. doi: https://doi.org/10.1016/j.pneurobio.2012.03.003.

18. Wyler SC, Gahlot S, Bideyan L, Yip C, Dushime J, Chen B, et al. LCoRL Regulates Growth and Metabolism. Endocrinology. 2024;165(12). doi: 10.1210/endocr/bqae146.
19. Lei XG, Cheng WH, McClung JP. Metabolic regulation and function of glutathione peroxidase-1. Annu Rev Nutr. 2007;27:41-61. doi: 10.1146/annurev.nutr.27.061406.093716.
20. Marciniak E, Leboucher A, Caron E, Ahmed T, Tailleux A, Dumont J, et al. Tau deletion promotes brain insulin resistance. J Exp Med. 2017;214(8):2257-69. doi: 10.1084/jem.20161731.
21. Guo Y, Liu Q, Zheng Z, Qing M, Yao T, Wang B, et al. Genetic association of inflammatory marker GlycA with lung function and respiratory diseases. Nature Communications. 2024;15(1):3751. doi: 10.1038/s41467-024-47845-w.
22. Wang S, Yue Y, Wang X, Tan Y, Zhang Q. SCARF2 is a target for chronic obstructive pulmonary disease: Evidence from multi-omics research and cohort validation. Aging Cell. 2024;23(10):e14266. doi: 10.1111/acel.14266.
23. Bautmans I, Knoop V, Amuthavalli Thiyagarajan J, Maier AB, Beard JR, Freiberger E, et al. WHO working definition of vitality capacity for healthy longevity monitoring. The Lancet Healthy Longevity. 2022;3(11):e789-e96. doi: 10.1016/S2666-7568(22)00200-8.
24. NHGRI-EBI. The NHGRI-EBI Catalog of human genome-wide association studies. 2024.
25. Leveille E, Ross OA, Gan-Or Z. Tau and MAPT genetics in tauopathies and synucleinopathies. Parkinsonism & Related Disorders. 2021;90:142-54. doi: https://doi.org/10.1016/j.parkreldis.2021.09.008.
26. Hanna MG. Genetic neurological channelopathies. Nature Clinical Practice Neurology. 2006;2(5):252-63. doi: 10.1038/ncpneuro0178.
27. de Wit MCY, de Coo IFM, Halley DJJ, Lequin MH, Mancini GMS. Movement disorder and neuronal migration disorder due to ARFGEF2 mutation. neurogenetics. 2009;10(4):333-6. doi: 10.1007/s10048-009-0192-2.
28. Fu C, Xu J, Cheng W, Rojas T, Chin AC, Snowman AM, et al. Neuronal migration is mediated by inositol hexakisphosphate kinase 1 via α-actinin and focal adhesion kinase. Proc Natl Acad Sci U S A. 2017;114(8):2036-41. doi: 10.1073/pnas.1700165114.
29. Strang KH, Golde TE, Giasson BI. MAPT mutations, tauopathy, and mechanisms of neurodegeneration. Laboratory Investigation. 2019;99(7):912-28. doi: https://doi.org/10.1038/s41374-019-0197-x.
30. Winder-Rhodes SE, Hampshire A, Rowe JB, Peelle JE, Robbins TW, Owen AM, et al. Association between MAPT haplotype and memory function in patients with Parkinson's disease and healthy aging individuals. Neurobiol Aging. 2015;36(3):1519-28. doi: 10.1016/j.neurobiolaging.2014.12.006.
31. Wingo TS, Liu Y, Gerasimov ES, Vattathil SM, Wynne ME, Liu J, et al. Shared mechanisms across the major psychiatric and neurodegenerative diseases. Nature Communications. 2022;13(1):4314. doi: 10.1038/s41467-022-31873-5.
32. Sexton C, Snyder H, Beher D, Boxer AL, Brannelly P, Brion JP, et al. Current directions in tau research: Highlights from Tau 2020. Alzheimers Dement. 2022;18(5):988-1007. doi: 10.1002/alz.12452.
33. Zhao B, Shan Y, Yang Y, Yu Z, Li T, Wang X, et al. Transcriptome-wide association analysis of brain structures yields insights into pleiotropy with complex neuropsychiatric traits. Nature Communications. 2021;12(1):2878. doi: 10.1038/s41467-021-23130-y.
34. Zhong Z, Yan M, Sun W, Wu Z, Han L, Zhou Z, et al. Two novel mutations in PRPF3 causing autosomal dominant retinitis pigmentosa. Scientific Reports. 2016;6(1):37840. doi: 10.1038/srep37840.
35. Wang H-F, Zhang W, Rolls ET, Li Y, Wang L, Ma Y-H, et al. Hearing impairment is associated with cognitive decline, brain atrophy and tau pathology. eBioMedicine. 2022;86:104336. doi: https://doi.org/10.1016/j.ebiom.2022.104336.

36. Biundo F, Del Prete D, Zhang H, Arancio O, D'Adamio L. A role for tau in learning, memory and synaptic plasticity. Scientific Reports. 2018;8(1):3184. doi: 10.1038/s41598-018-21596-3.
37. Stenbeck G, Coxon FP. Role of vesicular trafficking in skeletal dynamics. Current Opinion in Pharmacology. 2014;16:7-14. doi: https://doi.org/10.1016/j.coph.2014.01.003.
38. Vicart S, Sternberg D, Fontaine B, Meola G. Human skeletal muscle sodium channelopathies. Neurological Sciences. 2005;26(4):194-202. doi: 10.1007/s10072-005-0461-x.
39. Hallé M, Tremblay ML, Meng TC. Protein tyrosine phosphatases: emerging regulators of apoptosis. Cell Cycle. 2007;6(22):2773-81. doi: 10.4161/cc.6.22.4926.
40. Schindler A, Foley E. Hexokinase 1 blocks apoptotic signals at the mitochondria. Cellular Signalling. 2013;25(12):2685-92. doi: https://doi.org/10.1016/j.cellsig.2013.08.035.
41. Gandelman M, Dansithong W, Figueroa KP, Paul S, Scoles DR, Pulst SM. Staufen 1 amplifies proapoptotic activation of the unfolded protein response. Cell Death & Differentiation. 2020;27(10):2942-51. doi: 10.1038/s41418-020-0553-9.
42. Zeng C, Liu Y, He R, Lu X, Dai Y, Qi G, et al. Identification and validation of a novel cellular senescence-related lncRNA prognostic signature for predicting immunotherapy response in stomach adenocarcinoma. Front Genet. 2022;13:935056. doi: 10.3389/fgene.2022.935056.
43. Zheng S, Lin N, Wu Q, He H, Yang C. Prognostic model construction and validation of esophageal cancer cellular senescence-related genes and correlation with immune infiltration. Front Surg. 2023;10:1090700. doi: 10.3389/fsurg.2023.1090700.
44. Widden H, Placzek WJ. The multiple mechanisms of MCL1 in the regulation of cell fate. Communications Biology. 2021;4(1):1029. doi: 10.1038/s42003-021-02564-6.
45. Jayathirtha M, Whitham D, Alwine S, Donnelly M, Neagu AN, Darie CC. Investigating the Function of Human Jumping Translocation Breakpoint Protein (hJTB) and Its Interacting Partners through In-Solution Proteomics of MCF7 Cells. Molecules. 2022;27(23). doi: 10.3390/molecules27238301.
46. Larsson P, Pettersson D, Engqvist H, Werner Rönnerman E, Forssell-Aronsson E, Kovács A, et al. Pan-cancer analysis of genomic and transcriptomic data reveals the prognostic relevance of human proteasome genes in different cancer types. BMC Cancer. 2022;22(1):993. doi: 10.1186/s12885-022-10079-4.
47. Pérez-Martínez L, Wagner T, Luke B. Telomere Interacting Proteins and TERRA Regulation. Front Genet. 2022;13:872636. doi: 10.3389/fgene.2022.872636.
48. Yang S, Winstone L, Mondal S, Wu Y. Helicases in R-loop Formation and Resolution. J Biol Chem. 2023;299(11):105307. doi: 10.1016/j.jbc.2023.105307.
49. Kim SE, Mori R, Komatsu T, Chiba T, Hayashi H, Park S, et al. Upregulation of cytochrome c oxidase subunit 6b1 (Cox6b1) and formation of mitochondrial supercomplexes: implication of Cox6b1 in the effect of calorie restriction. Age (Dordr). 2015;37(3):9787. doi: 10.1007/s11357-015-9787-8.
50. Tolar LA, Pallanck L. NSF function in neurotransmitter release involves rearrangement of the SNARE complex downstream of synaptic vesicle docking. J Neurosci. 1998;18(24):10250-6. doi: 10.1523/jneurosci.18-24-10250.1998.
51. Schubert J, Siekierska A, Langlois M, May P, Huneau C, Becker F, et al. Mutations in STX1B, encoding a presynaptic protein, cause fever-associated epilepsy syndromes. Nature Genetics. 2014;46(12):1327-32. doi: 10.1038/ng.3130.
52. Bi W, Bao K, Zhou X, Deng Y, Li X, Zhang J, et al. PSMC5 regulates microglial polarization and activation in LPS-induced cognitive deficits and motor impairments by interacting with TLR4. Journal of Neuroinflammation. 2023;20(1):277. doi: 10.1186/s12974-023-02904-9.

53. Carulli D, de Winter F, Verhaagen J. Semaphorins in Adult Nervous System Plasticity and Disease. Front Synaptic Neurosci. 2021;13:672891. doi: 10.3389/fnsyn.2021.672891.
54. Tran AP, Silver J. Cathepsins in neuronal plasticity. Neural Regen Res. 2021;16(1):26-35. doi: 10.4103/1673-5374.286948.
55. El-Hattab AW, Scaglia F. SUCLA2-Related Mitochondrial DNA Depletion Syndrome, Encephalomyopathic Form with Methylmalonic Aciduria. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle

Copyright © 1993-2024, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.
56. Yoshizawa M, Kawauchi T, Sone M, Nishimura YV, Terao M, Chihama K, et al. Involvement of a Rac activator,P-Rex1, in neurotrophin-derived signaling and neuronal migration. J Neurosci. 2005;25(17):4406-19. doi: 10.1523/jneurosci.4955-04.2005.
57. Rogers J, Raveendran M, Fawcett GL, Fox AS, Shelton SE, Oler JA, et al. CRHR1 genotypes, neural circuits and the diathesis for anxiety and depression. Molecular Psychiatry. 2013;18(6):700-7. doi: 10.1038/mp.2012.152.
58. Zhou M, Sun X, Wang C, Wang F, Fang C, Hu Z. PFKM inhibits doxorubicin-induced cardiotoxicity by enhancing oxidative phosphorylation and glycolysis. Sci Rep. 2022;12(1):11684. doi: 10.1038/s41598-022-15743-0.
59. Cimadamore-Werthein C, Jaiquel Baron S, King MS, Springett R, Kunji ER. Human mitochondrial ADP/ATP carrier SLC25A4 operates with a ping-pong kinetic mechanism. EMBO Rep. 2023;24(8):e57127. doi: 10.15252/embr.202357127.
60. Tosserams A, Papadopoulos C, Jardel C, Lemière I, Romero NB, De Lonlay P, et al. Two new cases of mitochondrial myopathy with exercise intolerance, hyperlactatemia and cardiomyopathy, caused by recessive SLC25A4 mutations. Mitochondrion. 2018;39:26-9. doi: 10.1016/j.mito.2017.08.009.
61. Mulhem A, Moulla Y, Klöting N, Ebert T, Tönjes A, Fasshauer M, et al. Circulating cell adhesion molecules in metabolically healthy obesity. International Journal of Obesity. 2021;45(2):331-6. doi: 10.1038/s41366-020-00667-4.
62. Xiao Y, Junfeng H, Tianhong L, Lu W, Shulin C, Yu Z, et al. Cathepsin K in adipocyte differentiation and its potential role in the pathogenesis of obesity. J Clin Endocrinol Metab. 2006;91(11):4520-7. doi: 10.1210/jc.2005-2486.
63. Yadati T, Houben T, Bitorina A, Shiri-Sverdlov R. The Ins and Outs of Cathepsins: Physiological Function and Role in Disease Management. Cells. 2020;9(7). doi: 10.3390/cells9071679.
64. Huang W, Hickson LJ, Eirin A, Kirkland JL, Lerman LO. Cellular senescence: the good, the bad and the unknown. Nature Reviews Nephrology. 2022;18(10):611-27. doi: 10.1038/s41581-022-00601-z.
65. Elmore S. Apoptosis: a review of programmed cell death. Toxicol Pathol. 2007;35(4):495-516. doi: 10.1080/01926230701320337.
66. Cadigan KM, Waterman ML. TCF/LEFs and Wnt signaling in the nucleus. Cold Spring Harb Perspect Biol. 2012;4(11). doi: 10.1101/cshperspect.a007906.
67. Tai H-C, Schuman EM. Ubiquitin, the proteasome and protein degradation in neuronal function and dysfunction. Nature Reviews Neuroscience. 2008;9(11):826-38. doi: 10.1038/nrn2499.
68. Barritt SA, DuBois-Coyne SE, Dibble CC. Coenzyme A biosynthesis: mechanisms of regulation, function and disease. Nature Metabolism. 2024;6(6):1008-23. doi: 10.1038/s42255-024-01059-y.

69. Dong M, Deganutti G, Piper SJ, Liang Y-L, Khoshouei M, Belousoff MJ, et al. Structure and dynamics of the active Gs-coupled human secretin receptor. Nature Communications. 2020;11(1):4137. doi: 10.1038/s41467-020-17791-4.
70. Redelinghuys P, Brown GD. Inhibitory C-type lectin receptors in myeloid cells. Immunol Lett. 2011;136(1):1-12. doi: 10.1016/j.imlet.2010.10.005.
71. Kennedy MB. Synaptic Signaling in Learning and Memory. Cold Spring Harb Perspect Biol. 2013;8(2):a016824. doi: 10.1101/cshperspect.a016824.
72. Thinakaran G. The role of presenilins in Alzheimer's disease. J Clin Invest. 1999;104(10):1321-7. doi: 10.1172/jci8728.
73. Selkoe DJ, Hardy J. The amyloid hypothesis of Alzheimer's disease at 25 years. EMBO Mol Med. 2016;8(6):595-608. doi: 10.15252/emmm.201606210.
74. Südhof Thomas C. The Presynaptic Active Zone. Neuron. 2012;75(1):11-25. doi: https://doi.org/10.1016/j.neuron.2012.06.012.
75. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc. 2020;15(9):2759-72. doi: 10.1038/s41596-020-0353-1.
76. Demirkan A, Penninx BW, Hek K, Wray NR, Amin N, Aulchenko YS, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. Mol Psychiatry. 2011;16(7):773-83. doi: 10.1038/mp.2010.65.
77. Flint JP, Welstead M, Cox SR, Russ TC, Marshall A, Luciano M. Validation of a polygenic risk score for frailty in the Lothian Birth Cohort 1936 and English longitudinal study of ageing. Sci Rep. 2024;14(1):12586. doi: 10.1038/s41598-024-63229-y.
78. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics. 2018;50(8):1112-21. doi: 10.1038/s41588-018-0147-3.
79. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203-9. doi: 10.1038/s41586-018-0579-z.
80. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine. 2015;12(3):e1001779.
81. Forgetta V, Li R, Darmond-Zwaig C, Belisle A, Balion C, Roshandel D, et al. Cohort profile: genomic data for 26 622 individuals from the Canadian Longitudinal Study on Aging (CLSA). BMJ Open. 2022;12(3):e059021. doi: 10.1136/bmjopen-2021-059021.
82. Raina P, Wolfson C, Kirkland S, Griffith LE, Balion C, Cossette B, et al. Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). International Journal of Epidemiology. 2019;48(6):1752-3j. doi: 10.1093/ije/dyz173.
83. Raina PS, Wolfson C, Kirkland SA, Griffith LE, Oremus M, Patterson C, et al. The Canadian Longitudinal Study on Aging (CLSA). Canadian Journal on Aging / La Revue canadienne du vieillissement. 2009;28(3):221-9. doi: 10.1017/S0714980809990055.
84. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76-82. doi: 10.1016/j.ajhg.2010.11.011.
85. Zhu H, Zhou X. Statistical methods for SNP heritability estimation and partition: A review. Comput Struct Biotechnol J. 2020;18:1557-68. doi: 10.1016/j.csbj.2020.06.011.
86. Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, et al. A resource-efficient tool for mixed model association analysis of large-scale data. Nat Genet. 2019;51(12):1749-55. doi: 10.1038/s41588-019-0530-8.
87. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nature Communications. 2017;8(1):1826. doi: 10.1038/s41467-017-01261-5.

88. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11(4):e1004219. doi: 10.1371/journal.pcbi.1004219.
89. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50. doi: 10.1073/pnas.0506580102.
90. Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, et al. The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic acids research. 2005;33(suppl_1):D284-D8.
91. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nature Communications. 2019;10(1):1776. doi: 10.1038/s41467-019-09718-5.
92. Amare AT, Thalamuthu A, Schubert KO, Fullerton JM, Ahmed M, Hartmann S, et al. Association of polygenic score and the involvement of cholinergic and glutamatergic pathways with lithium treatment response in patients with bipolar disorder. Molecular Psychiatry. 2023;28(12):5251-61. doi: 10.1038/s41380-023-02149-1.
93. Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, Cearns M, et al. Association of polygenic score for major depression with response to lithium in patients with bipolar disorder. Molecular Psychiatry. 2021;26(6):2457-70. doi: 10.1038/s41380-020-0689-5.
94. Amare AT, Schubert KO, Hou L, Clark SR, Papiol S, Heilbronner U, et al. Association of Polygenic Score for Schizophrenia and HLA Antigen and Inflammation Genes With Response to Lithium in Bipolar Affective Disorder: A Genome-Wide Association Study. JAMA Psychiatry. 2018;75(1):65-74. doi: 10.1001/jamapsychiatry.2017.3433.
95. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47(11):1236-41. doi: 10.1038/ng.3406.

Acknowledgments

MB Beyene received postgraduate scholarship support from the University of Adelaide (The University Adelaide Research Scholarship) and AT Amare is currently supported by the National Health and Medical Research Council (NHMRC) Emerging Leadership (EL1) Investigator Grant (APP2008000). Access to the UK Biobank data was partially supported by funding from the Hospital Research Foundation to Professor Visvanathan.

The funding agencies had no role in the design or conduct of the study, data collection, data analysis, data interpretation, and preparing, reviewing, or approving the manuscript.

This research was made possible using the data/biospecimens collected by the CLSA and the UK biobank. Funding for the CLSA is provided by the Government of Canada through the Canadian Institutes of Health Research (CIHR) under grant reference: LSA 94473 and the

Canada Foundation for Innovation, as well as the following provinces, Newfoundland, Nova Scotia, Quebec, Ontario, Manitoba, Alberta, and British Columbia. This research has been conducted using the CLSA dataset [Comprehensive baseline v7.0 and GEN 3 data sets] under Application Number [2304009], and the UK biobank under application ID 70215. The CLSA is led by Drs. Parminder Raina, Christina Wolfson and Susan Kirkland.

The authors declare that authors alone are responsible for the views expressed in this article and do not necessarily represent the views, decisions, or policies of the institutions with which they are affiliated nor the views of the CLSA and UKB.

Author information

Authors and Affiliations

Adelaide Medical School, Faculty of Health and Medical Sciences, University of Adelaide, Adelaide, SA, Australia

Melkamu Bedimo Beyene, Renuka Visvanathan, Robel Alemu, Azmeraw T. Amare

Adelaide Geriatrics Training and Research with Aged Care Centre (GTRAC), Faculty of Health and Medical Sciences, University of Adelaide and the Basil Hetzel Institute, Central Adelaide Local Health Network Woodville, SA, 5011, Australia

Melkamu Bedimo Beyene, Renuka Visvanathan, Azmeraw T. Amare

Aged and Extended Care Services, The Queen Elizabeth Hospital, Central Adelaide Local Health Network, Woodville, SA, 5011, Australia

Renuka Visvanathan

University of California, Los Angeles (UCLA), CA, USA

Robel Alemu

Broad Institute of MIT and Harvard, Medical and Population Genetics Program, Cambridge, MA, USA

Robel Alemu

Australian Centre for Precision Health, Allied Health and Human Performance, University of South Australia, Adelaide, 5000, Australia

Beben Benyamin

South Australian Health and Medical Research Institute, Adelaide, 5000, Australia

Beben Benyamin

Adelaide Centre for Epigenetics, School of Biomedicine, The University of Adelaide, Adelaide, SA, 5005, Australia

Rudrarup Bhattacharjee

South Australian Immunogenomics Research Institute (SAiGENCI), The University of Adelaide, Adelaide, SA, 5005, Australia

Rudrarup Bhattacharjee

Metabolomics Laboratory, Baker Heart and Diabetes Institute, Melbourne

Habtamu B. Beyene

Department of Cardiovascular Research, Translation and Implementation, La Trobe University, Bundoora, Australia

Habtamu B. Beyene

Physiotherapy and Geriatric Medicine, Dalhousie University, Halifax, Nova Scotia, Canada

Olga Theou

Ageing and Health Unit, Department of Maternal, Newborn, Child & Adolescent Health and Ageing, World Health Organization, Geneva, Switzerland

Matteo Cesari

Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy

Matteo Cesari

Robert N. Butler Columbia Aging Centre, Columbia University, NY, USA

John R. Beard

Contributions

M.B. designed the study, data analysis, and methodology, wrote the initial draft, and contributed significantly to the subsequent revision of the article. A.A and R.V supervised the study, secured funding, contributed significantly to the design and analysis of the study, and drafting of the article as well as subsequent critical revision of the paper for its intellectual content. R.A., R.B, and H.B contributed significantly to the design of the study, drawing the graphs, drafting the article, and subsequent revision of the manuscript. J.B, M.C, B.B, O.T, Y.S, J.T contributed to the design of the study, contributed to the writeup of the paper, and critically revised the article for its intellectual content. All Authors critically reviewed and provided final approval of the manuscript submission.

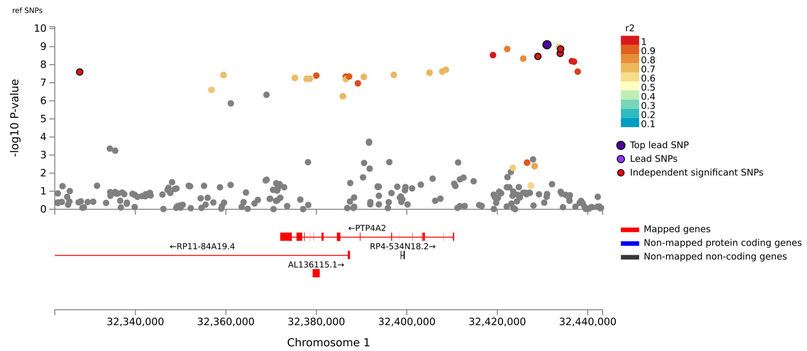
Corresponding author

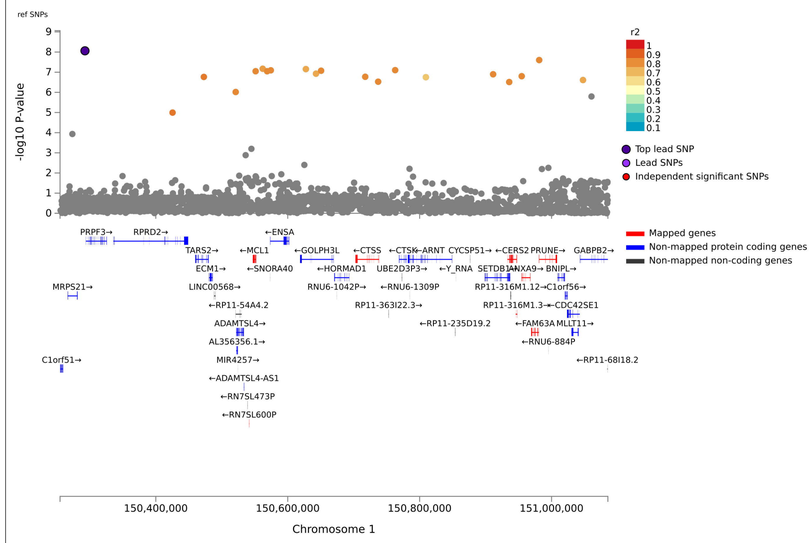
Correspondence to Azmeraw T. Amare

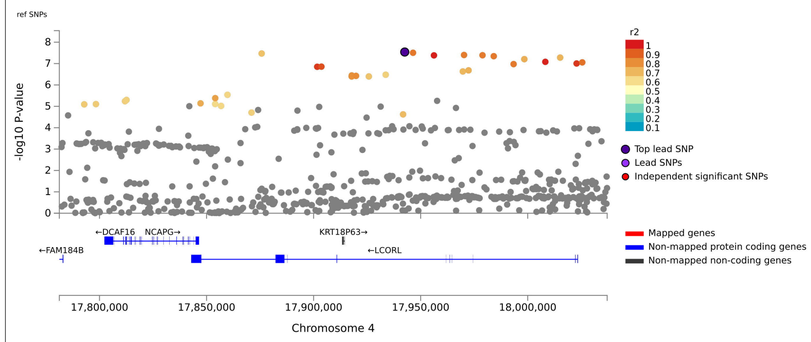
Ethics declarations

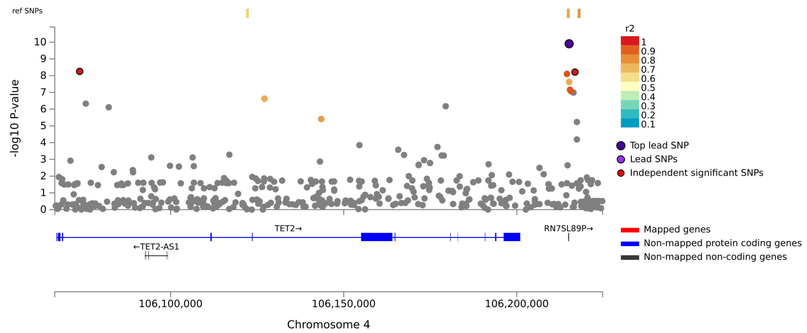
Competing interests

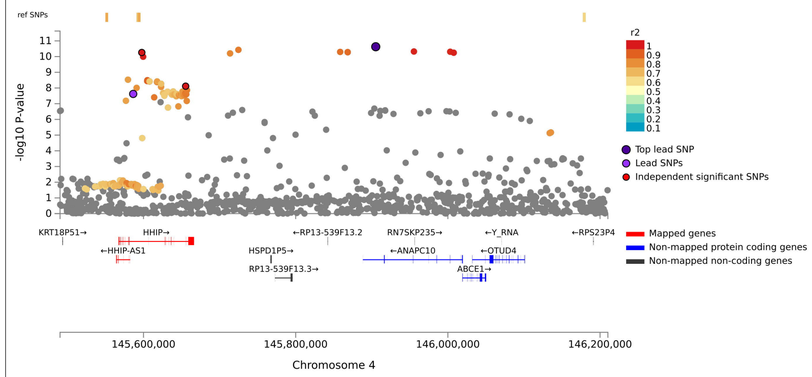
Professor Renuka Visvanathan and Professor John R. Beard are members of the World Health Organisation Clinical Consortium of Healthy Ageing.

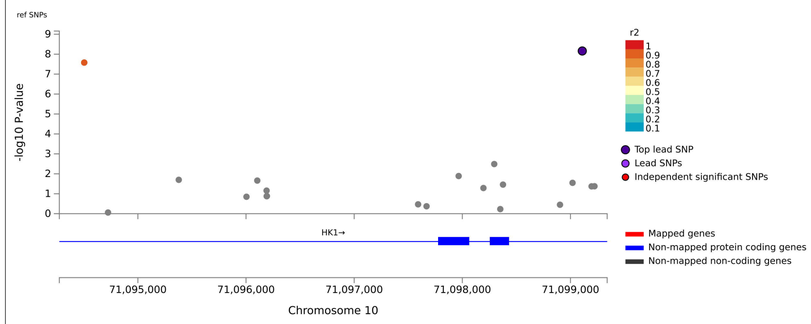

ref SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
←PTP4A2
←RP11-84A19.4
RP4-534N18.2→
AL136115.1→
32,340,000
32,360,000
32,380,000
32,400,000
32,420,000
32,440,000
Chromosome 1

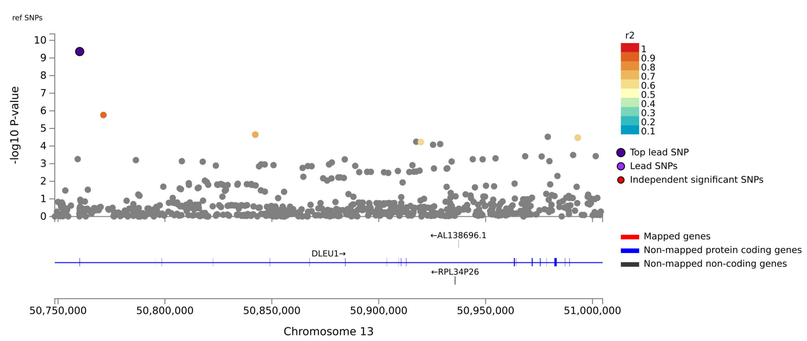

ref SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
150,400,000
150,600,000
150,800,000
151,000,000
Chromosome 1

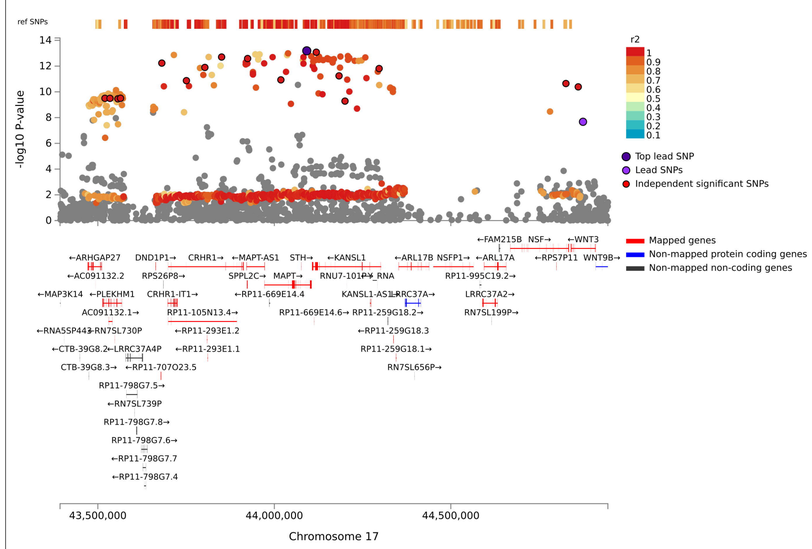

ref SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
←DCAF16
NCAPG→
KRT18P63→
←FAM184B
←LCORL
17,800,000
17,850,000
17,900,000
17,950,000
18,000,000
Chromosome 4

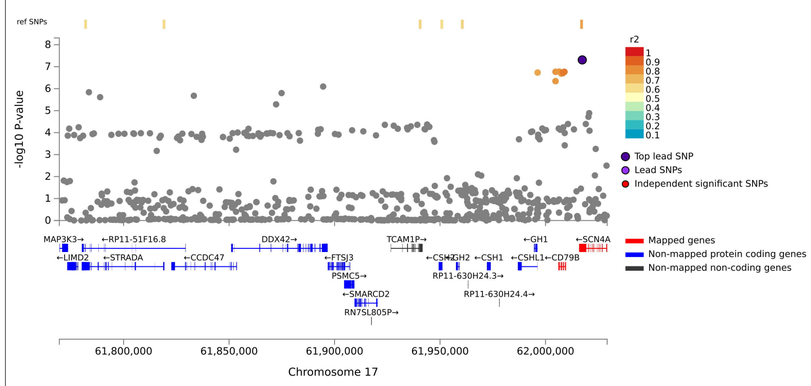

ref SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
TET2→
RN7SL89P→
←TET2-AS1
106,100,000
106,150,000
106,200,000
Chromosome 4

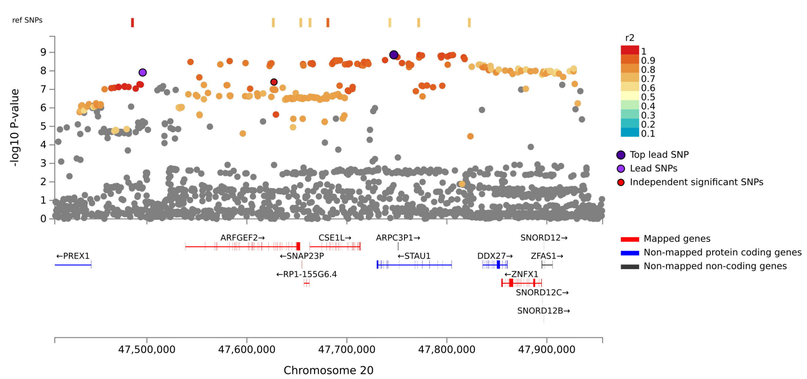

ref SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
KRT18P51→
HHIP→
←HHIP-AS1
←RP13-539F13.2
RN7SKP235→
←Y_RNA
←RPS23P4
HSPD1P5→
RP13-539F13.3→
←ANAPC10
←OTUD4
ABCE1→
145,600,000
145,800,000
146,000,000
146,200,000
Chromosome 4

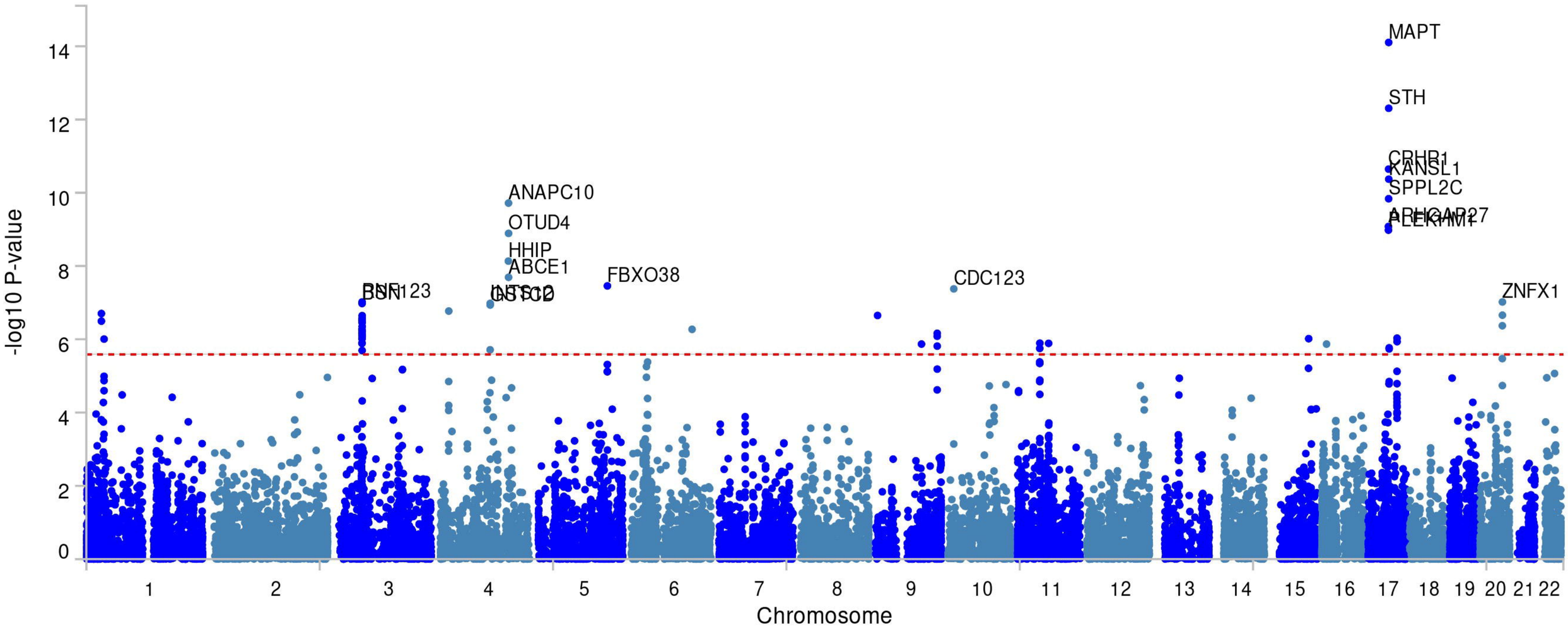

ref SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
HK1→
71,095,000
71,096,000
71,097,000
71,098,000
71,099,000
Chromosome 10

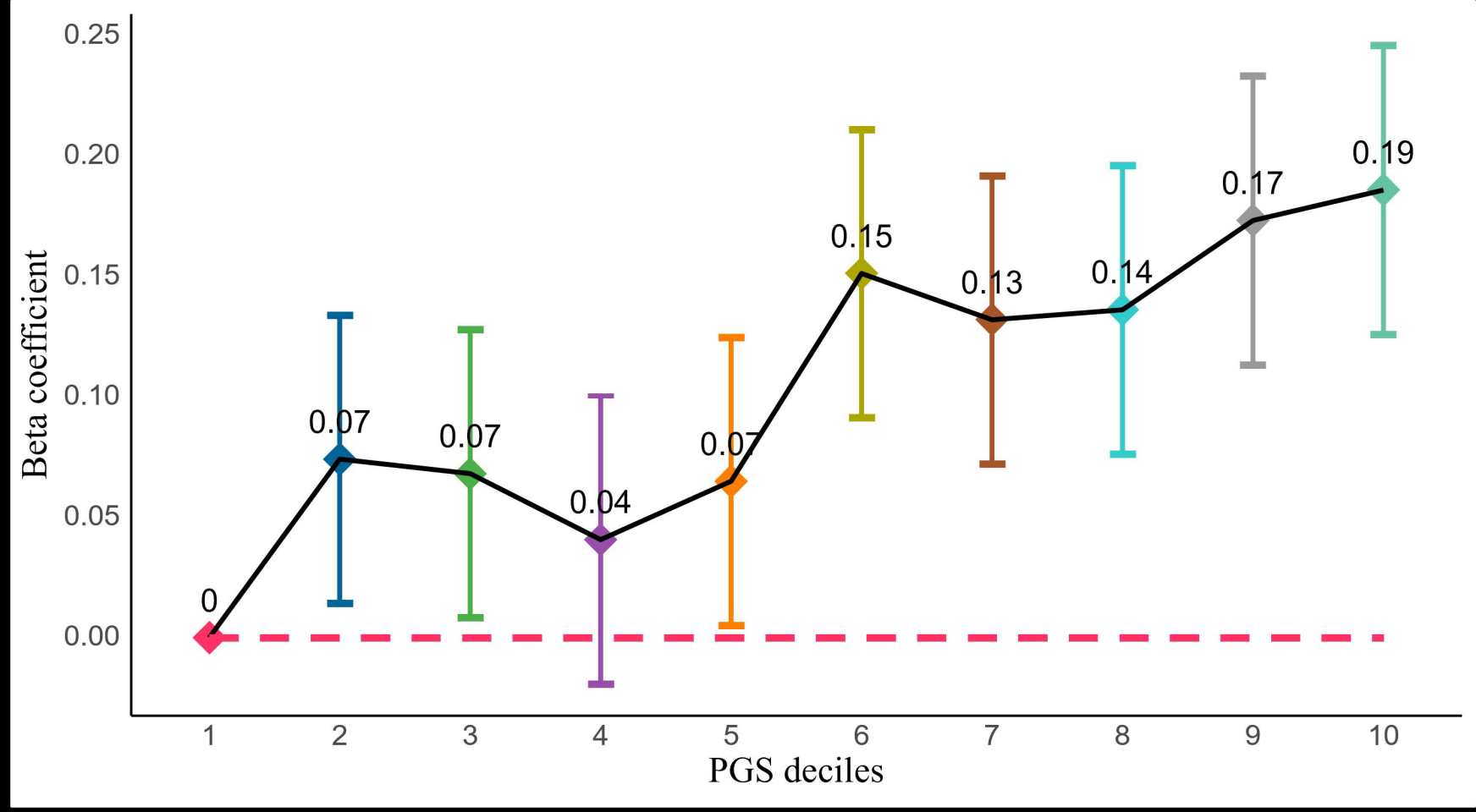

ref SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
←AL138696.1
DLEU1→
←RPL34P26
50,750,000
50,800,000
50,850,000
50,900,000
50,950,000
51,000,000
Chromosome 13


ref SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
43,500,000
44,000,000
44,500,000
Chromosome 17


ref SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
61,800,000
61,850,000
61,900,000
61,950,000
62,000,000
Chromosome 17


ref SNPs
-log10 P-value
r2
Top lead SNP
Lead SNPs
Independent significant SNPs
Mapped genes
Non-mapped protein coding genes
Non-mapped non-coding genes
←PREX1
ARFGEF2→
CSE1L→
ARPC3P1→
SNORD12→
←SNAP23P
←STAU1
DDX27→
ZFAS1→
←RP1-155G6.4
←ZNFX1
SNORD12C→
SNORD12B→
47,500,000
47,600,000
47,700,000
47,800,000
47,900,000
Chromosome 20


-log10 P-value
0
2
4
6
8
10
12
14
Chromosome
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
MAPT
STH
CRHR1
KANSL1
SPPL2C
ARHGAP27
PLEKHM1
ANAPC10
OTUD4
HHIP
ABCE1
FBXO38
CDC123
ZNFX1


0.25
0.20
0.15
0.10
0.05
0.00
Beta coefficient
0
0.07
0.07
0.04
0.07
0.15
0.13
0.14
0.17
0.19
1
2
3
4
5
6
7
8
9
10
PGS deciles