BRAIN
ORIGINAL ARTICLE

Genome-wide association study of occupational attainment as a proxy for cognitive reserve

Hyunwoong Ko,[1,2,3,†] Soyeon Kim,[4,5,6,†] Kiwon Kim,[7] Sang-Hyuk Jung,[4] Injeong Shim,[4] Soojin Cha,[4] Hyewon Lee,[8] Beomsu Kim,[4] Joohyun Yoon,[9] Tae Hyon Ha,[9] Seyul Kwak,[2,10] Jae Myeong Kang,[11] Jun-Young Lee,[2] Jinho Kim,[12] Woong-Yang Park,[13] Kwangsik Nho,[14] Doh Kwan Kim,[15] Woojae Myung[9,16,‡] and Hong-Hee Won[4,13,‡]

[†,‡]These authors contributed equally to this work.

Occupational attainment, which represents middle-age cognitive activities, is a known proxy marker of cognitive reserve for Alzheimer's disease. Previous genome-wide association studies have identified numerous genetic variants and revealed the genetic architecture of educational attainment, another marker of cognitive reserve. However, the genetic architecture and heritability for occupational attainment remain elusive.

We performed a large-scale genome-wide association study of occupational attainment with 248 847 European individuals from the UK Biobank using the proportional odds logistic mixed model method. In this analysis, we defined occupational attainment using the classified job levels formulated in the UK Standard Occupational Classification system considering the individual professional skill and academic level.

We identified 30 significant loci ($P < 5 \times 10^{-8}$); 12 were novel variants, not associated with other traits. Among them, four lead variants were associated with genes expressed in brain tissues by expression quantitative trait loci mapping from 10 brain regions: rs13002946, rs3741368, rs11654986 and rs1627527. The single nucleotide polymorphism-based heritability was estimated to be 8.5% (standard error of the mean = 0.004) and partitioned heritability was enriched in the CNS and brain tissues. Genetic correlation analysis showed shared genetic backgrounds between occupational attainment and multiple traits, including education, intelligence, leisure activities, life satisfaction and neuropsychiatric disorders. In two-sample Mendelian randomization analysis, we demonstrated that high occupation levels were associated with reduced risk for Alzheimer's disease [odds ratio (OR) = 0.78, 95% confidence interval (CI) = 0.65–0.92 in inverse variance weighted method; OR = 0.73, 95% CI = 0.57–0.92 in the weighted median method]. This causal relationship between occupational attainment and Alzheimer's disease was robust in additional sensitivity analysis that excluded potentially pleiotropic single nucleotide polymorphisms (OR = 0.72, 95% CI = 0.57–0.91 in the inverse variance weighted method; OR = 0.72, 95% CI = 0.53–0.97 in the weighted median method). Multivariable Mendelian randomization confirmed that occupational attainment had an independent effect on the risk for Alzheimer's disease even after taking educational attainment into account (OR = 0.72, 95% CI = 0.54–0.95 in the inverse variance weighted method; OR = 0.68, 95% CI = 0.48–0.97 in the weighted median method).

Overall, our analyses provide insights into the genetic architecture of occupational attainment and demonstrate that occupational attainment is a potential causal protective factor for Alzheimer's disease as a proxy marker of cognitive reserve.

Received January 12, 2021. Revised July 22, 2021. Accepted August 16, 2021. Advance access publication October 6, 2021

1 Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, South Korea
2 Department of Psychiatry, SMG-SNU Boramae Medical Center, Seoul National University College of Medicine, Seoul, South Korea
3 Dental Research Institute, Seoul National University School of Dentistry, Seoul, South Korea
4 Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, South Korea
5 Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
6 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
7 Department of Psychiatry, Kangdong Sacred Heart Hospital, Hallym University College of Medicine, Seoul, South Korea
8 Department of Health Administration and Management, College of Medical Sciences, Soonchunhyang University, Asan, South Korea
9 Department of Neuropsychiatry, Seoul National University Bundang Hospital, Seongnam, South Korea
10 Department of Psychology, Pusan National University, Busan, South Korea
11 Department of Psychiatry, Gil Medical Center, Gachon University College of Medicine, Incheon, South Korea
12 Precision Medicine Center, Future Innovation Research Division, Seoul National University Bundang Hospital, Seongnam, South Korea
13 Samsung Genome Institute, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea
14 Department of Radiology & Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA
15 Department of Psychiatry, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea
16 Department of Psychiatry, Seoul National University College of Medicine, Seoul, South of Korea

Correspondence to: Hong-Hee Won
Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, Seoul 06351, South Korea
E-mail: wonhh@skku.edu

Correspondence may also be addressed to: Woojae Myung
Department of Neuropsychiatry, Seoul National University Bundang Hospital 29, Gumi-ro 173 beon-gil Bundang-gu, Seongnam-si, Gyeonggi-do 13619, South Korea
E-mail: wmyung@snu.ac.kr

Keywords: occupational attainment; cognitive reserve; genome-wide association study; Alzheimer's disease; Mendelian randomization

Abbreviations: eQTL = expression quantitative trait loci; GWAS = genome-wide association study; IVW = inverse variance weighted; LD = linkage disequilibrium; LDSC = LD score regression; MAF = minor allele frequency; MR = Mendelian randomization; POLMM = proportional odds logistic mixed model; SNP = single nucleotide polymorphism; SOC = Standard Occupational Classification; TSMR = two-sample Mendelian randomization

Introduction

Occupational attainment has been found to be associated not only with physical health, such as cardiovascular or digestive diseases, but also with mental health. The hypothesis of cognitive reserve has been suggested to account for the disjunction between the degree of brain pathology and its clinical manifestations of Alzheimer's disease.[1] According to the hypothesis, the brain actively attempts to cope with the pathology by pre-existing compensatory mechanisms acquired by life experiences such as educational and occupational exposures. As a proxy marker of cognitive reserve, occupational attainment is associated with reduced risk of developing dementia and with slower rates of memory decline in the normal ageing process.[1–3] The role of occupational attainment as a cognitive proxy has been demonstrated in multiple studies including a twin registry study,[4] an imaging study[5] and a systemic review.[6]

Educational attainment is the most well studied proxy marker for cognitive reserve. High educational attainment levels are reportedly associated with a reduced risk of dementia incidence in the general population[7,8] Additionally, previous genetic studies have not only found that the genetic architecture of dementia was related to educational attainment,[9,10] but have also suggested that education had a causal effect on Alzheimer's disease, via Mendelian randomization (MR) analysis.[11] Similarly, occupational attainment was found to be associated with dementia risk and progression.[6,12] Moreover, the heritability of occupational attainment was reportedly comparable to that of educational attainment observed in a twin study (up to 0.43).[13] In addition, occupational attainment represents middle-age cognitive activities and has an independent role as a cognitive reserve that is in contrast with educational attainment that reflects early-life cognitive enrichment. However, our current knowledge of the genetic architecture of occupational attainment is limited.

There have been very few studies regarding the discovery of genetic loci associated with occupational attainment. Although some genetic studies have investigated on the relationship between occupational characteristics and cognitive performance in

older adults, they did not focus on the genetic architecture of occupational attainment.[4,14] A previous twin study demonstrated that the heritability of the occupational status was moderately high (43%),[13] and such occupational characteristics might lower the dementia risk.[4] A recent genetic study in the Estonian population estimated the single nucleotide polymorphism (SNP) heritability of occupational status to be 15%.[15] However, no previous study has identified the genetic variants associated with occupational attainment.

Therefore, we performed a genome-wide association study (GWAS) to identify genetic variants that could clarify the genetic architecture for occupational attainment, using the UK Biobank. In addition, we conducted a two-sample Mendelian randomization (TSMR) analysis to examine the causal association between occupational attainment and Alzheimer's disease risk. To the best of our knowledge, this study is the first GWAS of occupational attainment.

Materials and methods

UK Biobank

The UK Biobank is a nation-wide prospective study-based database involving >500 000 individuals between ages of 40–69 years recruited in 2006–10 in multiple centres throughout the UK. The UK Biobank has collected electronic medical records, computer-assisted interview data, touchscreen-based self-reported questionnaire data, physical and functional measures, and biological samples, including genotype data. The UK Biobank was approved by the National Research Ethics Committee (REC reference /NW/0382). All UK Biobank participants provided informed consent. A detailed description of the study can be found at https://www.ukbiobank.ac.uk/about-biobank-uk (accessed 8 February 2022).

Occupational attainment measures

Occupational attainment was operationally defined as a job code based on a hierarchy, starting from the highest level of managers and senior officials to the lowest level of elementary occupations. It was derived from the Standard Occupational Classification (SOC) system, developed by the UK Office of National Statistics.[16] The SOC was developed to classify all occupations in the UK into standard codes, according to the following four hierarchical structures: (i) nine major groups (one-digit); (ii) 25 submajor groups (two-digit); (iii) 81 minor groups (three-digit); and (iv) 353-unit groups (four-digit). The SOC was originally designed for statistical purposes, in which a higher number of digits would represent a greater level of detail in the job definition. Participants in the UK Biobank cohort were questioned about their current or most recent job during visits. Of the four hierarchical structures, we defined the structure composed of nine major groups as the occupational attainment phenotype for our genome-wide association analysis. Although the SOC code defined the levels 'one' and 'nine' as the highest and lowest skill levels, respectively, occupational attainment was recoded in our study as numeric ordinal variables ranging from one to nine with higher values reflecting greater and more complex occupational attainment. The nine major groups in the SOC are not only a result of categorized job titles but are also a hierarchical coding system considering skill level and specialization. Here, the skill level is defined as in terms of the task complexity and performed duties and skill specialization reflects knowledge and expertise of specific field and required job competencies. In this regard, the nine major groups reflect occupational complexity as well as the required academic skills and training, which suggests that the SOC code takes intellectual enrichment into consideration. Thus, we can further infer that the greater the major code digit, the greater the level of cognitive activity.

Genotyping and quality control

A total of 487 409 UK Biobank samples (v.3, March 2018) were genotyped using either the Affymetrix UK BiLEVE Axiom or Affymetrix UK Biobank Axiom arrays (Santa Clara, CA, USA), which have over 95% coverage and include >800 000 variants. Imputation was carried out centrally by the UK Biobank from a combined 1000 Genomes Project and UK 10K panel; phasing was performed using SHAPEIT3[17] and imputation was carried out using IMPUTE2.[18] The variant-level quality control (QC) exclusion metrics were applied to imputed data for GWAS as the followings: call rate <95%, Hardy–Weinberg equilibrium $P < 1 \times 10^{-6}$, and minor allele frequency (MAF) $<1 \times 10^{-4}$. Additionally, we excluded the variants if MAF values were <0.005 or imputation quality scores (INFO) were <0.4; we considered genotypes with a posterior call probability of <0.90 to be missing. A total of 9 575 249 SNPs met the QC criteria. The sample-level QC exclusion metrics applied to the imputed GWAS data were as follows: non-Europeans, samples with mismatched sex, putative sex chromosome aneuploidy or no sex information, and participants who withdrew from the UK Biobank. Finally, out of 310 527 who had a measure of occupational attainment, 248 847 participants of European ancestry were included in the analysis.

Genome-wide association analysis

For ordinal categorical phenotypes, genome-wide association analysis was performed using a proportional odds logistic mixed model (POLMM).[19] POLMM uses a sparse genetic relationship matrix to adjust for relatedness within samples as a random effect, and uses a saddle point approximation to control inflated type I error rates due to unbalanced distribution in ordinal categorical data. The simulations and real data analyses in the original paper[19] have shown that POLMM could reduce type I error rates and increase the power of GWAS when analysing ordinal traits, compared to commonly used mixed models; SAIGE[20], BOLT-LMM[21] and fastGWA.[22] Age, sex and 10 principal components of genetic ancestry were adjusted for association analysis. A genome-wide significance threshold of $P < 5 \times 10^{-8}$ was used to identify variants associated with occupational attainment. Regional association plots were generated using LocusZoom (http://locuszoom.sph.umich.edu/locuszoom/, accessed 8 February 2022).[23]

Identification of significant loci by GWAS and functional annotation

Independent significant SNPs with $P < 5 \times 10^{-8}$ and $r^2 < 0.2$ were identified from GWAS using FUMA.[24] The most significant SNPs per locus were selected as lead SNPs. The maximum distance for linkage disequilibrium (LD) blocks to merge into a genomic locus was 3000 kb. The genetic data of European population in the 1000 Genomes Project v.3 were considered as reference data for LD analyses. ANNOVAR[25] implemented in FUMA was used to annotate SNPs.

Gene mapping and functional annotation

For the 30 independent genomic risk loci identified via LD clumping, evidence of expression quantitative trait loci (eQTL) and functional annotation was examined by using FUMA platform.[26] Mapping SNPs to genes significantly associated with eQTL was analysed by eQTL analysis using the Genotype-Tissue Expression (GTEx) (https://www.gtexportal.org/home/datasets, accessed 8

February 2022) database v.8.[27] From reported cis-eQTL SNP-gene pairs in FUMA, we regarded as significant eQTL associations at a false discovery rate (FDR) <0.05.

Pathway analysis

Biological pathway analysis was performed on results from gene-based analysis using MAGMA implemented in FUMA. The enriched gene set was examined based on the Gene Ontology (GO) Consortium.[28]

GWAS catalogue lookup

We examined whether the genomic risk loci identified in our genome-wide association analysis were overlapped in loci reported associations in published GWAS listed in the NHGRI-EBI catalog[29] using FUMA.

SNP-based heritability and cell type-specific analyses

LD score regression (LDSC) estimated the SNP-based heritability for occupational attainment using GWAS summary statistics.[30] We obtained the pre-computed European LD scores of the 1000 Genomes Project v.3 from GitHub (https://github.com/bulik/ldsc, accessed 8 February 2022). We included common autosomal variants with a MAF >1% in the EUR population and excluded variants at the MHC region in this evaluation.

We conducted cell type-specific analyses to prioritize phenotype-associated tissues or cell types and detected significant tissue-specific enrichment (FDR < 5%) using gene expression data with GWAS summary statistics.[31] We used several gene sets previously described by Finucane *et al.*[31] and Cahoy *et al.*,[32] as well as multi-tissue gene expression (includes both GTEx[33] data and Franke laboratory[34,35] data), multi-tissue chromatin (includes both Roadmap Epigenomics[36] and ENCODE[37] data) and ImmGen data.[38]

Genetic correlation

To examine the underlying shared genetic background and obtain etiological insights, we estimated the level of cross-trait genetic correlation (r_g) between occupational attainment and 89 other phenotypes using LDSC.[30] We downloaded the European GWAS summary-level data of 89 phenotypes from publicly available sources (Supplementary Table 11). All data used in this analysis were controlled for quality; their imputation quality score was >0.8 and MAF was >0.5%. The FDR correction was used for multiple test correction (89 traits).

Brain annotation

Brain regions were visualized using BrainNet Viewer v.1.7[39] and default interpolations and perceptually uniform colour scales were normalized to the MNI-ICBM-152 template[40] in MATLAB R2020a (Mathworks, Inc., Natick, MA, USA). We mapped volumes of the region of interest in the brain and performed diffusion tensor imaging (DTI) using the brain region of interest atlas[41] and brain DTI atlas,[42] respectively. Each brain region was coloured blue and red, depending on the corresponding value of r_g, which ranged from –1 to 1. A detailed description of the brain volume and DTI measure can be found at UK Biobank Brain Imaging Documentation (https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf, accessed 8 February 2022).

Mendelian randomization

We performed TSMR to investigate the credible association between occupational attainment and Alzheimer's disease. We used European summary statistics for Alzheimer's disease, obtained after performing the International Genomics of Alzheimer's Project (IGAP) meta-analyses.[43] Variants in GWAS summary statistics for the exposure and outcome were filtered if the imputation scores were <0.8 or MAF values were <0.5%. With regard to instrumental variables for the exposure of interest (occupational attainment levels), we clumped SNPs exhibiting genome-wide significance (5×10^{-8}) with r^2 values $\leqslant$ 0.001 using the European population reference from the 1000 Genomes Project v.3. We excluded instrumental variables that were non-existent in the Alzheimer's disease data or non-inferable palindromic SNPs and harmonized the effects (beta coefficients) of genetic variants on occupational attainment and Alzheimer's disease for the same allele. All analyses were conducted using the R package 'TwoSampleMR' v.0.4.25. The inverse variance weighted (IVW) analysis results provide an estimate of causal effect under the assumption that all genetic variants are valid as instrumental variables.[44] The weighted median method provides a more reliable estimate because the genetic variation contributing more significantly to the causal effect has more weight.[45] We used the intercept test for MR-Egger regression to test for pleiotropic effects. The null test hypothesis is that the intercept value indicating pleiotropy is zero. We applied the MR-PRESSO global test to the primary MR result to identify pleiotropic outliers and obtain estimates regarding the causal effect, after excluding any outliers.[46]

Multivariable MR analyses were conducted using the multivariable MR functions[47] built in the MendelianRandomization R package (v.0.5.0). GWAS summary statistics for educational attainment as an additional exposure were obtained from Okbay *et al.*[10]. Although there were partially overlapping samples due to the use of the same UK Biobank data in both GWASs for occupational attainment and educational attainment, the bias derived from sample overlap would be negligible as we included only strong instruments ($P < 5 \times 10^{-8}$) in multivariable MR.[48,49] The GWAS summary statistics for occupational attainment and educational attainment had no samples overlapping with that of Alzheimer's disease. To select instruments for a multivariable MR model, we selected independent variants with r^2 values $\leqslant$0.001 from each exposure GWAS and harmonized the effects of each variant and corresponding effect allele across exposure and outcome GWASs. After removing invalid SNPs, a total of 69 SNPs were used as instruments (16 for occupational attainment and 53 for educational attainment).

Data availability

The data are available from the UK Biobank (https://www.ukbiobank.ac.uk, accessed 8 February 2022) on application. The GWAS summary statistics on occupational attainment can be obtained from the GWAS catalogue (https://www.ebi.ac.uk/gwas/). The GWAS summary statistics for Alzheimer's disease are available from the IGAP website (http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php, accessed 1 January 2021).

Results

Distribution of occupational attainment status in the UK Biobank

To create a harmonized measure of occupational attainment, we used the SOC system, as defined by the UK Office for National

Statistics.[16] Descriptions and frequencies of SOC-coded occupation traits are presented in Supplementary Table 1. Because the nine major groups in the SOC system have a hierarchical structure that reflects the training or academic skill levels for specific occupations, we considered SOC-coded groups as a quantitative trait and conducted association analysis. This quantitative definition of the occupational attainment level is consistent with that used in previous studies using SOC or other job descriptive measures.[50–52] Among the 487 409 individuals in the UK Biobank, GWAS was conducted for 248 847 individuals after excluding those without an SOC code and after applying sample QC for the genotype data. Participant characteristics by SOC-coded groups are presented in Supplementary Table 1.

Genome-wide significant association signals

The GWAS of occupational attainment was performed at the cohort level for individuals of European descent, using the POLMM method with age, sex and principal components of genetic ancestry as covariates. Using a clumping method, we identified 30 independent genomic loci with a genome-wide significance threshold ($P < 5 \times 10^{-8}$) and considered the most significant SNPs in each locus as lead SNPs (Fig. 1 and Supplementary Fig. 1). The traits and regional plots of lead SNPs are described in detail in the Supplementary material. The quantile-quantile plot of the GWAS results (Supplementary Fig. 2) demonstrates genomic inflation ($\lambda = 1.29$), which is attributable to their expected polygenicity ([LDSC intercept of 1.056, standard error of the mean (SEM) = 0.009]. No GWAS for occupational attainment has been reported; hence, we investigated the related genes or traits for the 30 loci. Of these loci, 18 contained lead SNPs previously associated with cognition-related traits (i.e. educational attainment, cognitive ability, math ability and intelligence), whereas 12 were not previously reported and thus represented novel variants without any association with other cognitive traits (Table 1, Fig. 1 and Supplementary Table 2).

Functional annotation of the identified loci and biological pathways

To link the associated variants with relevant genes, we employed the GTEx database implemented in the FUMA platform, to perform the functional annotation of GWAS results. Using eQTL analysis, we found the GWAS SNPs mapping to 63 cis-eQTL genes in 13 brain tissue types (Supplementary Tables 3 and 4). Four novel lead SNPs were identified as eQTLs for 10 genes within 10 specific brain tissues (Table 1 and Supplementary Table 4); rs13002946, rs3741368, rs11654986 and rs1627527. *AFF3* and *RAD51C* were mapped for cortex, *AFF3* for frontal cortex, *AFF3* and *ZDHHC24* for anterior cingulate cortex, *AFF3*, *C19orf71*, *FZR1*, *MFSD12*, *MTMR4*, *RP11-867G23.8* and *TEX14* for cerebellum and cerebellar hemisphere, *TEX14* and *SKA2* for hippocampus, *AFF3*, *C19orf71*, *TEX14* and *ZDHHC24* for nucleus accumbens, caudate and putamen, and *TEX14* for hypothalamus, respectively.

To investigate potential biological pathways, we conducted pathway analysis using MAGMA,[53] which was implemented in FUMA. The results highlighted 11 genes and a significant GO pathway [nucleotide excision repair (NER) complex] (Bonferroni-corrected $P < 0.05$, see Supplementary Table 5 for further details).

SNP heritability and partitioned heritability analysis

We applied partitioned LDSC[54] to evaluate how the GWAS results of occupational attainment were enriched in 53 genomic annotations using the full baseline model. The SNP heritability of occupational attainment was estimated to be 8.5% in this study, which was lower than the estimated heritability from a twin study for post-war Norwegian cohorts (43%)[13] and comparable to the educational attainment level (SNP heritability = 12%).[55] The functional enrichment test showed that only one of the 53 annotations, known as the 'Conserved Lindblad Toh', passed the FDR criterion of 0.05 for occupational attainment (Fig. 2A and Supplementary Table 6).[56] The proportion of SNPs for the conserved region was 2.6% and the estimated enrichment value was ~18 (coefficient $P = 1.97 \times 10^{-13}$). The conserved region is

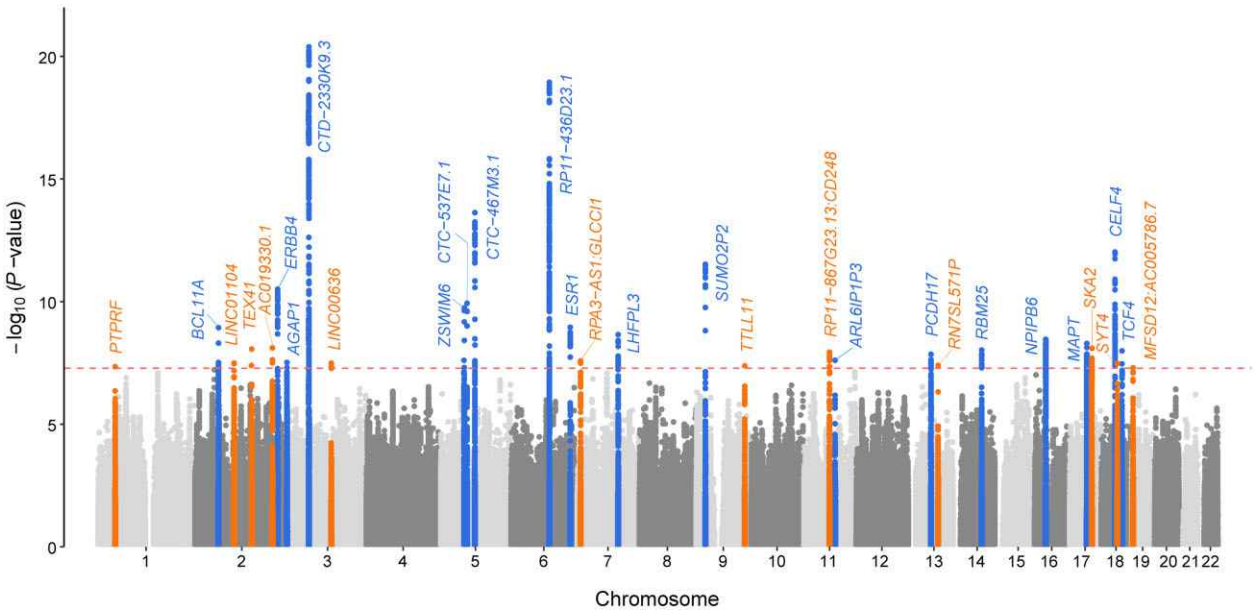

Figure 1 A Manhattan plot for a GWAS of occupational attainment. The *x*-axis shows genomic positions and *y*-axis shows statistical significance as $-\log_{10}$ (*P*) values. The threshold for significance, which accounts for values obtained after multiple tests, is shown by the red horizontal line ($P = 5 \times 10^{-8}$). The blue and yellow lines indicate the mapped genes from known and novel loci, respectively.

Table 1 Summary of the lead SNPs in the 30 loci associated with occupational attainment

	CHR	BP	A1/A2	EAF	Beta	SEM	P	Nearest genes	eQTL genes
l loci									
r 8 588	1	44012923	T/C	0.311	–0.043	0.008	4.31×10^{-8}	*PTPRF*	–
r 3 2946	2	100801959	A/T	0.276	0.044	0.008	3.07×10^{-8}	*LINC01104*	*AFF3*
r 1 94958	2	145940794	T/C	0.01	–0.209	0.036	8.19×10^{-9}	*TEX41*	–
r 4 351	2	199492201	G/A	0.431	–0.041	0.007	7.38×10^{-9}	*AC019330.1*	–
r 4 81	3	107691697	G/A	0.149	0.054	0.010	3.08×10^{-8}	*LINC00636*	–
r 2 857	7	8010634	G/T	0.139	–0.056	0.010	2.49×10^{-8}	*RPA3-AS1*, *GLCCI1*	–
r 8 9843	9	124604538	A/T	0.419	–0.039	0.007	3.98×10^{-8}	*TTLL11*	–
r 7 368	11	66083782	A/G	0.446	0.040	0.007	1.1×10^{-8}	*RP11-867G23.13*, *CD248*	*RP11-867G23.8*, *ZDHHC24*
r 5 530	13	76809010	G/C	0.150	–0.050	0.009	3.85×10^{-8}	*RN7SL571P*	–
r 1 4986	17	57206124	A/G	0.384	–0.042	0.007	7.68×10^{-9}	*SKA2*	*MTMR4*, *RAD51C*, *SKA2*, *TEX14*
r 5 4947	18	40940932	C/T	0.126	–0.056	0.010	3.23×10^{-8}	*SYT4*	–
r 6 527	19	3555498	A/G	0.475	0.039	0.007	4.77×10^{-8}	*MFSD12*, *AC005786.7*	*C19orf71*, *FZR1*, *MFSD12*
Known loci associated with other cognitive traits									
r 3 9832	2	60710571	A/G	0.425	0.043	0.007	1.09×10^{-9}	*BCL11A*	–
r 5 9536	2	212634084	T/A	0.261	–0.051	0.008	2.93×10^{-11}	*ERBB4*	–
r 5 7393	2	236832758	T/C	0.364	–0.041	0.007	2.90×10^{-8}	*AGAP1*	–
r 6 084	3	49949834	A/T	0.496	0.066	0.007	3.59×10^{-21}	*CTD-2330K9.3*	*FAM212A*, *GMPPB*, *MST1R*, *RBM6*, *RNF123*
rs7719676	5	60736949	A/G	0.31	0.048	0.007	1.71×10^{-10}	*ZSWIM6*	–
rs10515086	5	67781021	T/C	0.175	–0.059	0.009	1.11×10^{-10}	*CTC-537E7.1*	–
rs448809	5	88005828	G/T	0.44	0.055	0.007	2.18×10^{-14}	*CTC-467M3.1*	–
rs9375188	6	98555272	T/C	0.492	0.064	0.007	1.01×10^{-19}	*RP11-436D23.1*	–
rs12210020	6	152220889	A/G	0.248	–0.049	0.008	1.06×10^{-9}	*ESR1*	–
rs6944796	7	104505787	T/C	0.191	0.052	0.009	2.11×10^{-9}	*LHFPL3*	–
rs12553324	9	23347865	G/C	0.430	0.049	0.007	2.82×10^{-12}	*SUMO2P2*	–
rs1880692	11	80338069	G/A	0.455	–0.039	0.007	2.36×10^{-8}	*ARL6IP1P3*	–
rs4886031	13	58372213	T/C	0.235	–0.044	0.008	1.34×10^{-8}	*PCDH17*	–
rs2806047[a]	14	73532676	G/A	0.345	0.042	0.007	8.98×10^{-9}	*RBM25*	–
rs2726036	16	28347140	C/A	0.334	–0.042	0.007	3.28×10^{-9}	*NPIPB6*	*EIF3C*, *LAT*, *NPIPB6*, *NPIPB7*, *NPIPB9*, *NUPR1*, *SH2B1*, *SULT1A2*, *TUFM*
rs77875796	17	44051612	G/A	0.241	–0.049	0.008	4.89×10^{-9}	*MAPT*	*ARHGAP27*, *ARL17A*, *CRHR1*, *FMNL1*, *KANSL1*, *LRRC37A*, *LRRC37A2*, *MAPT*, *NMT1*, *NSF*, *PLEKHM1*, *SPPL2C*
rs9964724	18	35159124	C/T	0.332	–0.054	0.007	8.89×10^{-13}	*CELF4*	–
rs619466	18	53198836	A/G	0.097	0.069	0.012	9.74×10^{-9}	*TCF4*	–

A1 = effect allele; A2 = non-effect allele; Beta = regression coefficient; BP = genomic position in human genome assembly GRCh37 (hg19); CHR = chromosome; EAF = effect allele frequency.

[a] The high LD ($r^2 > 0.8$) of the variants within ±500 kb from the variants related to cognition in previous studies.

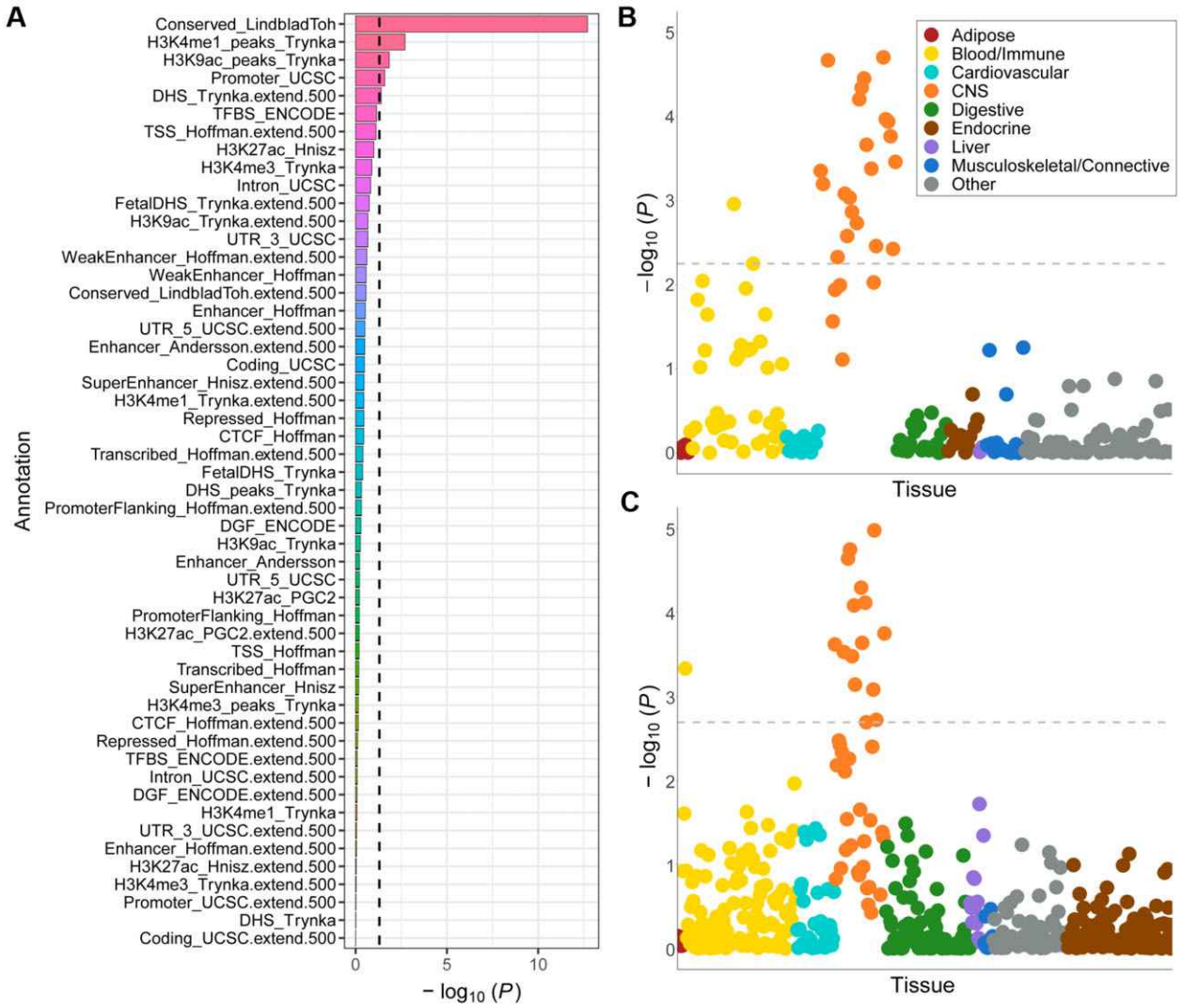

Figure 2 Partitioned heritability analyses using LDSC. (**A**) Enrichment estimates for 53 functional annotations. Annotations are ordered by their *P*-values. The dashed line indicates the significance at $P < 0.05$. (**B**) Results of multiple-tissue analysis using gene expression data. Each circle represents a tissue or cell type from either the GTEx dataset or Franke laboratory dataset. The dashed line indicates the cut-off of FDR, which is <5% at $-\log_{10}(P) = 2.25$. (**C**) Results of multiple-tissue analysis using chromatin data. Each circle represents peaks for DNase I hypersensitivity (DHS) or histone marks in a tissue or cell type. The dashed line indicates the cut-off of FDR, which is <5% at $-\log_{10}(P) = 2.69$.

reportedly associated with cognitive function and intelligence.[57,58] No other annotations were significantly enriched.

formed LDSC for specifically expressed genes (LDSC-SEG; Methods) identified after the occupational attainment GWAS, using gene sets obtained from the LDSC site (https://github.com/bulik/ldsc, accessed 8 Feb 2022).[31] The analysis of occupational attainment across multiple tissues revealed that areas in the CNS, such as the hippocampus, limbic system or frontal cortex, were strongly enriched at an FDR of <5% (Supplementary Table 7). The immune cells and neurons were also found to be enriched (Supplementary Tables 7–9).

The hippocampus and limbic system showed the highest enrichment in the subcortical structure with regard to multi-tissue gene expression; the cortex area, including the frontal, middle and temporal areas, also demonstrated high enrichment levels (Fig. 2B and Supplementary Table 7). Multi-tissue chromatin results showed that the highest enrichment was observed in the foetal brain and germinal matrix regions (Fig. 2C and Supplementary Table 10). These results were in accordance with those of a previous report, suggesting that genomic loci associated with educational attainment, another proxy of cognitive reserve, were biologically linked to brain development-related phenotypes.[59]

Genetic correlation between occupational attainment and other traits

LDSC was conducted to test whether genetic variants associated with health-related traits shared genetic bases with genetic variants for occupational attainment (Fig. 3 and Supplementary Table 11). A significant positive correlation was observed between occupational attainment and miscarriage ($r_g = 0.33$), satisfaction-related traits [r_g range = (0.34, 0.45)], bipolar disorder ($r_g = 0.34$), autism spectrum disorder ($r_g = 0.09$), leisure-related traits [r_g range = (0.49, 0.65)], high-density lipoprotein (HDL) cholesterol ($r_g = 0.23$), testosterone ($r_g = 0.09$) and cognitive function-related traits [r_g range = (0.66, 0.90)]. A significant negative correlation was

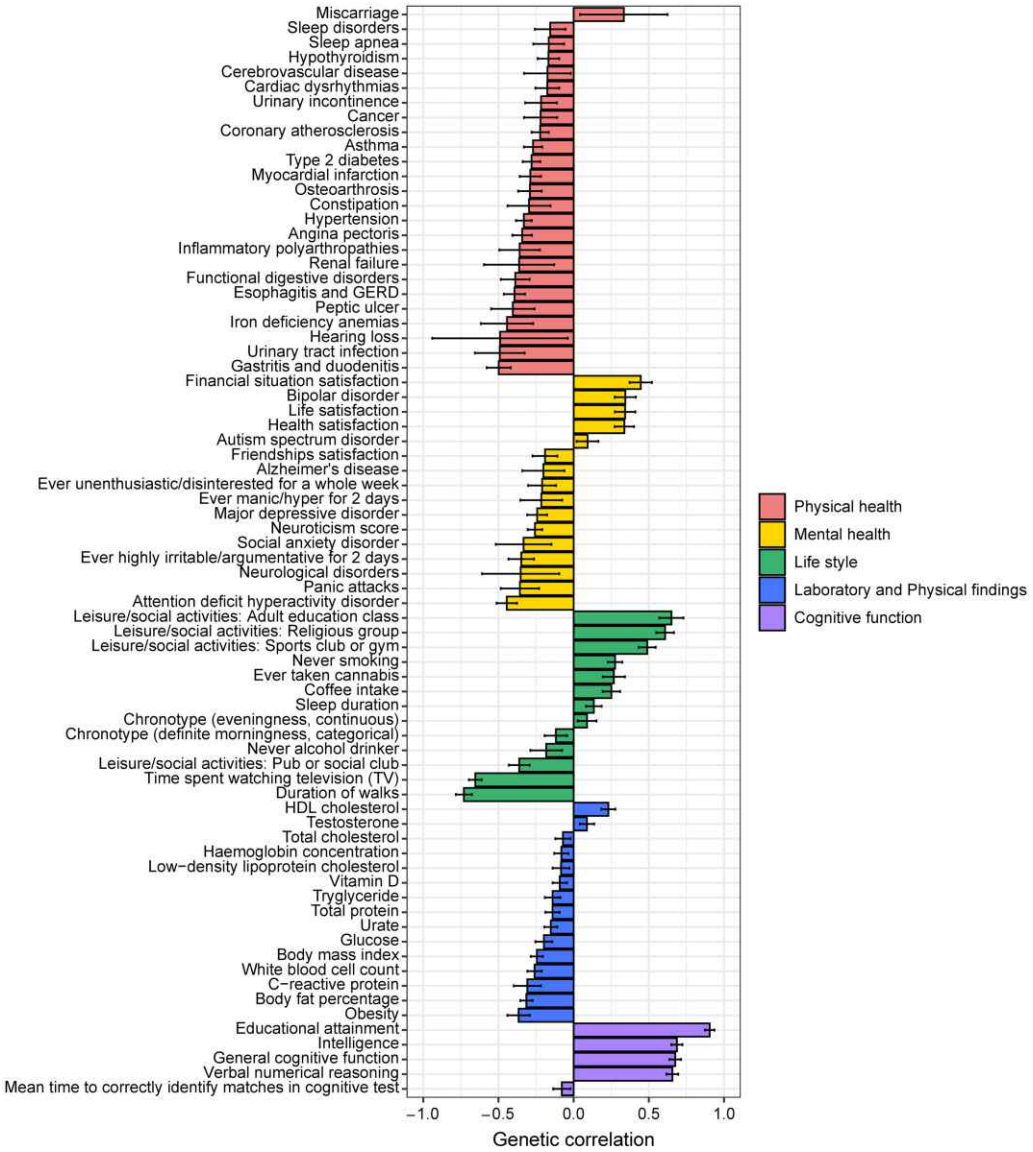

Figure 3 Genetic correlation estimates between occupational attainment and other phenotypes using LDSC. This figure only includes significant genetic correlations, where FDR values are <5% (see Supplementary Table 11 for all results). GERD = gastroesophageal reflux disease.

observed with overall physical/mental health [r_g range = (–0.05, –0.50)], morningness chronotype (r_g = –0.12), walk duration (r_g = –0.73), overall laboratory/physical findings [r_g range = (–0.07, –0.37)] and mean time for correctly identifying matches in a cognitive test (r_g = –0.08). These genetic correlations suggest that there are common genetic associations between occupational attainment and multiple mental/physical health-related traits, cognitive functions, lifestyle factors and laboratory findings.

LDSC analysis was also conducted to estimate the genetic correlation between occupational attainment and the brain regional volume and connectivity traits, via structural brain imaging or DTI (Fig. 4 and Supplementary Table 12). A positive correlation was observed between occupational attainment and total brain volume (r_g = 0.24), and the left inferior temporal (r_g = 0.21), left inferior parietal (r_g = 0.17)s and left and right insular regions (r_g = 0.14; r_g = 0.13), with regard to the brain regional volume trait. A negative correlation was observed with the left pericalcarine (r_g = –0.18). With respect to DTI traits, no significant results were observed after adjusting for multiple comparisons.

Mendelian randomization analysis

We conducted a TSMR analysis to estimate the causal effect of occupational attainment as a proxy marker of cognitive reserve, on the Alzheimer's disease risk. The summary statistics for Alzheimer's disease was obtained from the IGAP meta-analyses.[43] Among all the independent genome-wide significant SNPs associated with occupational attainment, 18 variants were used as instrumental variables after harmonizing the data for the two GWAS results (see 'Materials and methods' section and Supplementary Table 13). The TSMR analysis results revealed a negative association between occupational attainment and Alzheimer's disease (OR = 0.78, 95% CI = 0.65–0.92, $P = 4.26 \times 10^{-3}$ with the IVW method; OR = 0.73, 95% CI = 0.57–0.92, $P = 9.10 \times 10^{-3}$ in the weighted median method; Table 2). The intercept from MR-Egger regression analysis showed the absence of horizontal pleiotropy (P for the MR-Egger intercept test[60] was 0.90), and MR-PRESSO[46] analysis did not detect any outliers from instrumental variables.

To ensure that the significant causal effect was not attributable to pleiotropic bias, we performed sensitivity analysis after excluding

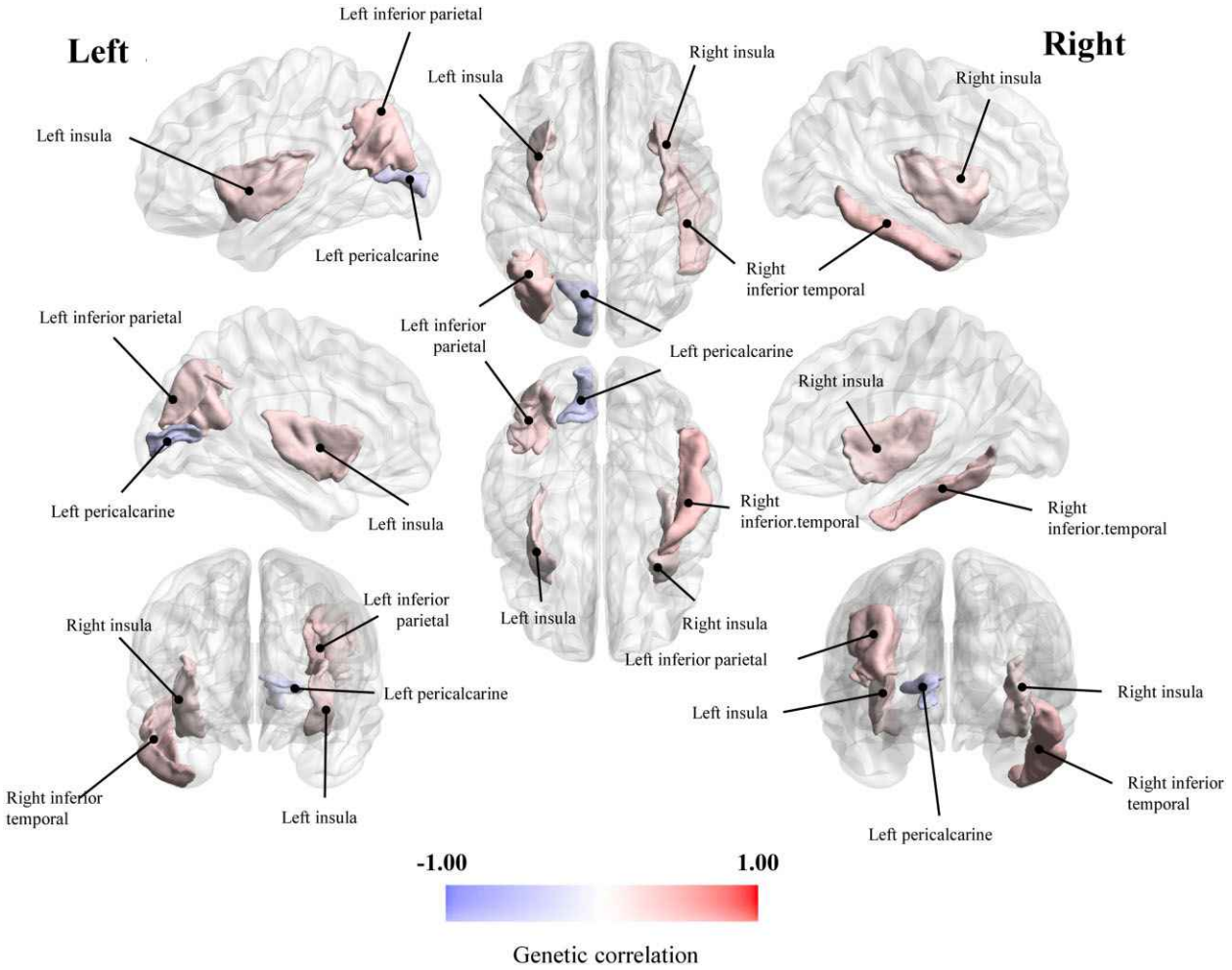

Figure 4 Brain regions representing significant genetic correlations with occupational attainment in genetic correlation analysis using LDSC. Brain image showing genetic correlations between occupational attainment and volumes of region of interest of the brain at FDR values <5%. The blue and red colours indicate a negative and positive correlation, respectively. The darker the colour, the stronger the correlation.

seven potentially pleiotropic SNPs (Supplementary Table 13). Even after excluding these SNPs, the estimate of the causal effect of occupational attainment for Alzheimer's disease patients remained significant (OR = 0.72, 95% CI = 0.57–0.91, $P = 5.54 \times 10^{-3}$ in the IVW method; OR = 0.72, 95% CI = 0.53–0.97, $P = 0.03$ in the weighted median method; Table 2). The results indicate that occupational attainment might have a protective effect on Alzheimer's disease risk.

Subsequently, we sought to evaluate whether the observed causal relationship between occupational attainment and Alzheimer's disease is derived from an independent effect of occupational attainment or is partially driven by educational attainment, which is a well-known protective factor for Alzheimer's disease.[11,61,62] We performed multivariable MR with educational attainment as an additional exposure. When we included both occupational attainment and educational attainment (years of schooling[10]) in a multivariable MR model, we found strong evidence of independent effects of occupational attainment on the risk for Alzheimer's disease (OR = 0.72, 95% CI = 0.54–0.95, $P = 0.02$ in the IVW method; OR = 0.68, 95% CI = 0.48–0.97, $P = 0.04$ in the weighted median method; Table 2 and Supplementary Table 14). However, we found no evidence of independent effects of educational attainment on the risk for Alzheimer's disease (OR = 1.08, 95% CI = 0.62–1.91, $P = 0.78$ in the IVW method; OR = 1.10, 95% CI = 0.53–2.30, $P = 0.79$ in the weighted median method).

Discussion

In this study, we identified 30 genetic variants that elucidate the genetic architecture of occupational attainment. Although many of these SNPs were associated with cognitive traits and mental/physical health, we also discovered 12 novel signals, which were not reported in any published GWAS for cognitive traits. The proportion of variance attributable to the additive effects of all SNPs was estimated to be 8.5% in SNP heritability analysis. In eQTL, genetic correlation and LDSC-SEG analysis, we found that occupational attainment was associated with the CNS. Moreover, our TSMR analysis results indicated that occupational attainment could be a protective factor for Alzheimer's disease, as a proxy marker of cognitive reserve.

We identified 10 genes expressed in brain tissues using the eQTL mapping of 12 newly identified genetic loci: *MTMR4*, *AFF3*, *SKA2*, *TEX14*, *FZR1*, *RAD51C*, *C19orf71*, *MFSD12*, *ZDHHC24* and *RP11-867G23.8*. Several genes have functions related to cognition and neuropsychiatric diseases. *MTMR4*, which regulates macrophage phagocytes,[63] is related to general cognitive ability[64] and cognitive aspects of educational attainment.[65] *AFF3*, an autosomal homolog of X-linked *AFF2*, which encodes members of the *ALF* family (*AFF1/AF4FF2/FMR2* and *AFF4*, *MCEF*),[66] is associated with verbal-numerical reasoning, which was representative of general cognitive

Table 2. Univariable a multivariable MR estimates for occupational attainment on Alzheimer's disease

Method	*n* SNPs	OR (95% CI)	*P*
Primary MR for occupational attainment on risk of Alzheimer's disease			
Inverse variance weighted	18	0.78 (0.65 to 0.92)	4.26×10^{-3}
Weighted median		0.73 (0.57 to 0.92)	9.10×10^{-3}
MR-Egger (*P* for pleiotropy = 0.90)		0.73 (0.27 to 1.95)	0.54
Sensitivity analysis for occupational attainment on risk of Alzheimer's disease after the exclusion of pleiotropic SNPs			
Inverse variance weighted	11	0.72 (0.57 to 0.91)	5.54×10^{-3}
Weighted median		0.72 (0.53 to 0.97)	0.03
MR-Egger (*P* for pleiotropy = 0.97)		0.70 (0.12 to 4.00)	0.70
Sensitivity analysis for independent effect of occupational attainment on risk of Alzheimer's disease by multivariable MR controlling for educational attainment			
Exposure: Occupational attainment			
variance weighted	69	0.72 (0.54 to 0.95)	0.02
Median based		0.68 (0.48 to 0.97)	0.04
MR-Egger (*P* for pleiotropy[a] = 0.21)		0.63 (0.45 to 0.89)	8.27×10^{-3}
Exposure: Educational attainment			
variance weighted	69	1.08 (0.62 to 1.91)	0.78
Median based		1.10 (0.53 to 2.30)	0.79
MR-Egger (*P* for pleiotropy[a] = 0.21)		0.63 (0.23 to 1.74)	0.38

CI = confidence interval; IV = instrumental variable; OR = odds ratio.
[a]Note that there is only one *P*-value for the MR-Egger intercept in the multivariable MR model.

ability.[67] *AFF3* is also involved in the pathophysiology of intellectual disability.[59,66] *SKA2*, which is considered to be involved in the cortisol stress system related to development of post-traumatic stress disorder,[68] is associated with general cognitive function.[64] In addition, *TEX14*, which is involved in midbody function, is not only related to cognitive performance and educational attainment[55] but also biomarkers such as ferritin level,[69] white blood cell count[70] and neutrophil count.[70] *FZR1* and *RAD51C* are also related to blood markers, red cell distribution width[70–72] and blood protein levels and eosinophil counts[72,73], whereas *C19orf71* and *MFSD12* are associated with skin status and disease, including skin pigmentation[74] and cutaneous malignant melanoma.[75] *ZDHHC24*, which is expressed at high levels, mainly in the brain, pancreas, prostate and stomach,[76] is associated with mean corpuscular haemoglobin,[67] apolipoprotein B[77] and low-density lipoprotein cholesterol levels.[78] The functions and associated traits of *RP11-867G23.8* have not been reported yet. Our findings from pathway analysis are consistent with those of previous reports, demonstrating that the NER complex pathway deficiencies are involved in accelerated cognitive decline and neurodegeneration in a mouse model.[79]

Using TSMR analysis, we demonstrated evidence implicating occupational attainment as one of the causal factors that might reduce Alzheimer's disease risk, as it is a marker of cognitive reserve. Although previous studies have shown that the measurement of cognitive reserve using occupational attainment was associated with reduced Alzheimer's disease risk, most of these studies were limited to observational evidence that was insufficient to establish the causal relationship between occupational attainment and Alzheimer's disease risk due to the inherent limitations of an observational study, such as uncontrolled potential environmental confounders.[80,81] Because allele segregation occurs randomly and is unaffected by environmental confounding factors, TSMR analysis has emerged as a promising approach for examining causal associations.[82] Therefore, our TSMR-related findings could act as robust evidence illustrating the protective effects of occupational attainment against Alzheimer's disease. Moreover, in view of high genetic correlations between cognition-related traits and occupational attainment (Fig. 3), we conducted sensitivity analysis to investigate the causal effects on Alzheimer's disease after removing pleiotropic SNPs associated with cognition-related traits and confirmed that the results remained significant. In addition, the results of the multivariable MR to estimate independent causal effects of occupational and educational attainments on the risk for Alzheimer's disease suggest that occupational attainment indicating middle-age cognitive activities may be a more influential independent protective factor for Alzheimer's disease than educational attainment showing early-life cognitive enrichment. This may suggest that educational attainment improves occupational attainment, and the effect of educational attainment on Alzheimer's disease is mediated through occupational attainment. Because educational attainment and occupational attainment can be confounders, colliders or mediators with each other, caution is needed when interpreting the multivariable MR results of occupational attainment as having a significant mediating effect between educational attainment and Alzheimer's disease.[83] Therefore, to further understand how individual factors for cognitive reserve such as educational attainment and occupational attainment have an effect on Alzheimer's disease, additional MR analyses with more instrumental variables and appropriate analytic methods for causal pathways may be warranted.

Our findings showed that the CNS is involved with the majority of the biological implications of occupational attainment. The enrichment analysis results illustrated a broad involvement of brain tissues, including the cortex and subcortical area, and highlighted the significant enrichment of GWAS variants in CNS tissues, as compared to that of other types of tissues (Fig. 2B and C and Supplementary Table 7); we found that neurons were implicated in our genetic findings, not microglia or astrocytes (Supplementary Table 9). EQTL mapping prioritized brain tissue genes (Table 1). In accordance with the findings of previous studies,[82,84] our genetic correlation analysis not only demonstrated the relationship between occupational attainment and various health-related traits, but also identified the genetic link between educational attainment and intelligence, cognitive function, and brain phenotypes via neuroimaging (Figs 3 and 4). These genetic findings suggested that our GWAS results were strongly associated with brain function and indicated that patterns were consistent with those of previous GWAS findings regarding intelligence and educational attainment.[10,58]

Several limitations were associated with this study. The current GWAS data used for examining the effects of rare variants or environment factors was limited; this should be taken into account during future studies, including whole-exome or whole-genome sequencing studies. Our findings provide insights into the genetic architecture of occupational attainment, but there was no single genetic determinant for occupational attainment. Because human genetic architectures and behavioural phenotypes are highly complex and are affected by environmental factors, our results should not be used to predict an individual's occupational attainment level. Instead, our findings might be used as novel biomarkers for cognitive reserve and neurodegenerative diseases. For several SOC-coded groups, we observed an imbalanced sex ratio that may reflect gender differences in occupational distributions in the UK Biobank. Therefore, in our analysis, we adjusted for sex as a covariate to reduce sex biases due to this imbalance. Moreover, our study was conducted in the UK Biobank cohort alone; thus, the replication of our findings in independent cohorts is necessary. We applied a recently developed POLMM to analyse occupational attainment as an ordinal categorical phenotype. Although ordinal phenotypes are widely investigated in GWAS, most studies on categorical phenotypes used conventional methods designed for binary or quantitative traits. The POLMM was introduced as a scalable and accurate mixed model approach for ordinal categorical data analysis in large-scale GWAS and tested in multiple ordinal categorical phenotypes using the UK Biobank data. Via this method, we avoided inflated type I error rates and reduced statistical power that could result from treating the categorical phenotypes as a quantitative trait. Owing to the scarcity of GWASs using POLMM, further studies for the validation and replication of this tool in diverse cohort datasets are warranted. An additional GWAS and subsequent meta-analysis across diverse populations are essential for identifying more novel genetic factors that would further explain the genetic architecture of occupational attainment.

Funding

This research has been conducted using the UK Biobank Resource under Application Number 33002. It was supported by the National Research Foundation of Korea of Korea Grant funded by the Ministry of Science and Information and Communication Technologies, South Korea (grant numbers NRF-2018R1C1B6001708 and NRF-2021R1A2C4001779 to W.M. and NRF-2019R1A2C4070496 to H.H.W.), and a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (HI19C1132 to H.H.W. and HI19C1328000020 to S.K.). This research was also supported by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program (IITP-2021–2017-0–01630 and IITP-2021–2018-0–01833 to H.K.) supervised by the Institute for Information & Communications Technology Planning & Evaluation (IITP).

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Stern Y. Cognitive reserve. *Neuropsychologia*. 2009;47(10):2015–2028.
2. Stern Y, Gurland B, Tatemichi TK, Tang MX, Wilder D, Mayeux R. Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA*. 1994;271(13):1004–1010.
3. Stern Y, Albert S, Tang MX, Tsai WY. Rate of memory decline in AD is related to education and occupation: Cognitive reserve? *Neurology*. 1999;53(9):1942–1947.
4. Potter GG, Helms MJ, Burke JR, Steffens DC, Plassman BL. Job demands and dementia risk among male twin pairs. *Alzheimers Dement*. 2007;3(3):192–199.
5. Chapko D, McCormack R, Black C, Staff R, Murray A. Life-course determinants of cognitive reserve (CR) in cognitive aging and dementia–a systematic literature review. *Aging Ment Health*. 2018;22(8):921–932.
6. Valenzuela MJ, Sachdev P. Brain reserve and dementia: A systematic review. *Psychol Med*. 2006;36(4):441–454.
7. Stern Y, Arenaza-Urquijo EM, Bartres-Faz D, et al.; the Reserve, Resilience and Protective Factors PIA Empirical Definitions and Conceptual Frameworks Workgroup. Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimers Dement*. 2020;16(9):1305–1311.
8. Xu W, Tan L, Wang HF, et al. Education and risk of dementia: Dose-response meta-analysis of prospective cohort studies. *Mol Neurobiol*. 2016;53(5):3113–3123.
9. Rietveld CA, Medland SE, Derringer J, et al.; LifeLines Cohort Study. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*. 2013;340(6139):1467–1471.
10. Okbay A, Beauchamp JP, Fontana MA, et al.; LifeLines Cohort Study. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. 2016;533(7604):539–542.
11. Larsson SC, Traylor M, Malik R, et al. Modifiable pathways in Alzheimer's disease: Mendelian randomisation analysis. *BMJ*. 2017;359:j5375.
12. Myung W, Lee C, Park JH, et al. Occupational attainment as risk factor for progression from mild cognitive impairment to Alzheimer's disease: A CREDOS study. *J Alzheimers Dis*. 2017;55(1):283–292.
13. Tambs K, Sundet JM, Magnus P, Berg K. Genetic and environmental contributions to the covariance between occupational status, educational attainment, and IQ: A study of twins. *Behav Genet*. 1989;19(2):209–222.
14. Carlson MC, Helms MJ, Steffens DC, Burke JR, Potter GG, Plassman BL. Midlife activity predicts risk of dementia in older male twin pairs. *Alzheimers Dement*. 2008;4(5):324–331.
15. Rimfeld K, Krapohl E, Trzaskowski M, et al. Genetic influence on social outcomes during and after the Soviet era in Estonia. *Nat Hum Behav*. 2018;2(4):269–275.
16. Office for National Statistics. *Standard occupational classification 2000, volume 1: Structure and descriptions of unit groups*. The Stationery Office; 2000.
17. O'Connell J, Sharp K, Shrine N, et al. Haplotype estimation for biobank-scale data sets. *Nat Genet*. 2016;48(7):817–820.
18. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5(6):e1000529.
19. Bi W, Zhou W, Dey R, Mukherjee B, Sampson JN, Lee S. Efficient mixed model approach for large-scale genome-wide association studies of ordinal categorical phenotypes. *Am J Hum Genet*. 2021;108(5):825–839.

20. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50(9):1335–1341.
21. Loh PR, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47(3):284–290.
22. Jiang L, Zheng Z, Qi T, et al. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet*. 2019; 51(12):1749–1755.
23. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336–2337.
24. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826.
25. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
26. Watanabe K, Umicevic Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. *Nat Commun*. 2019;10(1):3222.
27. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509): 1318–1330.
28. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res*. 2017; 45(D1):D331–D338.
29. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1): D1005–D1012.
30. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236–1241.
31. Finucane HK, Reshef YA, Anttila V, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet*. 2018;50(4):621–629.
32. Cahoy JD, Emery B, Kaushal A, et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *J Neurosci*. 2008;28(1):264–278.
33. GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science*. 2015;348(6235):648–660.
34. Fehrmann RS, Karjalainen JM, Krajewska M, et al. Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet*. 2015;47(2):115–125.
35. Pers TH, Karjalainen JM, Chan Y, et al.; Genetic Investigation of ANthropometric Traits (GIANT) Consortium. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun*. 2015;6(1):5890-
36. Kundaje A, Meuleman W, Ernst J, et al.; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518(7539):317–330.
37. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414): 57–74.
38. Heng TS, Painter MW; Immunological Genome Project Consortium. The Immunological Genome Project: Networks of gene expression in immune cells. *Nat Immunol*. 2008;9(10): 1091–1094.
39. Xia M, Wang J, He Y. BrainNet Viewer: A network visualization tool for human brain connectomics. *PLoS ONE*. 2013;8(7):e68910-
40. Mazziotta J, Toga A, Evans A, et al. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philos Trans R Soc Lond B Biol Sci*. 2001; 356(1412):1293–1322.
41. Desikan RS, Segonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006;31(3): 968–980.
42. Mori S, Oishi K, Jiang H, et al. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *Neuroimage*. 2008;40(2):570–582.
43. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013;45(12):1452–1458.
44. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. 2013;37(7):658–665.
45. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304–314.
46. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50(5):693–698.
47. Burgess S, Thompson SG. Multivariable Mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol*. 2015;181(4):251–260.
48. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol*. 2016;40(7):597–608.
49. Anderson EL, Howe LD, Wade KH, et al. Education, intelligence and Alzheimer's disease: Evidence from a multivariable two-sample Mendelian randomization study. *Int J Epidemiol*. 2020; 49(4):1163–1172.
50. Cumberland PM, Rahi JS; UK Biobank Eye and Vision Consortium. Visual function, social position, and health and life chances: The UK Biobank Study. *JAMA Ophthalmol*. 2016; 134(9):959–966.
51. Howe LD, Kanayalal R, Harrison S, et al. Effects of body mass index on relationship status, social contact and socio-economic position: Mendelian randomization and within-sibling study in UK Biobank. *Int J Epidemiol*. 2020;49(4):1173–1184.
52. Kivimaki M, Batty GD, Pentti J, et al. Association between socioeconomic status and the development of mental and physical health conditions in adulthood: A multi-cohort study. *Lancet Public Health*. 2020;5(3):e140–e149.
53. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219-
54. Finucane HK, Bulik-Sullivan B, Gusev A, et al.; RACI Consortium. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228–1235.
55. Lee JJ, Wedow R, Okbay A, et al.; Social Science Genetic Association Consortium. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50(8): 1112–1121.
56. Lindblad-Toh K, Garber M, Zuk O, et al.; Genome Institute at Washington University. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. 2011;478(7370): 476–482.
57. Hill WD, Davies G, Harris SE, et al.; neuroCHARGE Cognitive Working group. Molecular genetic aetiology of general cognitive function is enriched in evolutionarily conserved regions. *Transl Psychiatry*. 2016;6(12):e980-

58. Savage JE, Jansen PR, Stringer S, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet*. 2018;50(7):912–919.
59. Lam M, Trampush JW, Yu J, et al. Large-scale cognitive GWAS meta-analysis reveals tissue-specific neural expression and potential nootropic drug targets. *Cell Rep*. 2017;21(9):2597–2613.
60. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512–525.
61. Meng X, D'Arcy C. Education and dementia in the context of the cognitive reserve hypothesis: A systematic review with meta-analyses and qualitative analyses. *PLoS ONE*. 2012;7(6):e38268.
62. Sharp ES, Gatz M. Relationship between education and dementia: An updated systematic review. *Alzheimer Dis Assoc Disord*. 2011;25(4):289–304.
63. Sheffield DA, Jepsen MR, Feeney SJ, et al. The myotubularin MTMR4 regulates phagosomal phosphatidylinositol 3-phosphate turnover and phagocytosis. *J Biol Chem*. 2019;294(45): 16684–16697.
64. Davies G, Lam M, Harris SE, et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun*. 2018;9(1):2098.
65. Demange PA, Malanchini M, Mallard TT, et al. Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *NatGenet*. 2021;53(1):35–44.
66. Metsu S, Rooms L, Rainger J, et al. FRA2A is a CGG repeat expansion associated with silencing of AFF3. *PLoS Genet*. 2014;10(4): e1004242-
67. Davies G, Marioni RE, Liewald DC, et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol Psychiatry*. 2016;21(6):758–767.
68. Boks MP, Rutten BP, Geuze E, et al. SKA2 methylation is involved in cortisol stress reactivity and predicts the development of posttraumatic stress disorder (PTSD) after military deployment. *Neuropsychopharmacology*. 2016;41(5):1350–1356.
69. Benyamin B, Esko T, Ried JS, et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nat Commun*. 2014;5(1):1–11.
70. Vuckovic D, Bao EL, Akbari P, et al. The polygenic and monogenic basis of blood traits and diseases. *Cell*. 2020;182(5): 1214–1231.e1211.
71. Kichaev G, Bhatia G, Loh PR, et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am J Hum Genet*. 2019;104(1):65–75.
72. Chen M-H, Raffield LM, Mousas A, et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell*. 2020;182(5):1198–1213.e1114.
73. Ahsan M, Ek WE, Rask-Andersen M, et al. The relative contribution of DNA methylation and genetic variants on protein biomarkers for human diseases. *PLoS Genetics*. 2017;13(9):e1007005.
74. Adhikari K, Mendoza-Revilla J, Sohail A, et al. A GWAS in Latin Americans highlights the convergent evolution of lighter skin pigmentation in Eurasia. *Nat Commun*. 2019;10(1):358.
75. Landi MT, Bishop DT, MacGregor S, et al.; MelaNostrum Consortium. Genome-wide association meta-analyses combining multiple risk phenotypes provide insights into the genetic architecture of cutaneous melanoma susceptibility. *Nat Genet*. 2020;52(5):494–504.
76. Korycka J, Łach A, Heger E, et al. Human DHHC proteins: A spotlight on the hidden player of palmitoylation. *Eur J Cell Biol*. 2012; 91(2):107–117.
77. Richardson TG, Sanderson E, Palmer TM, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: A multivariable Mendelian randomisation analysis. *PLoS Med*. 2020;17(3): e1003062-
78. Tachmazidou I, Suveges D, Min JL, et al. Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. *Am J Hum Genet*. 2017;100(6):865–884.
79. Borgesius NZ, de Waard MC, van der Pluijm I, et al. Accelerated age-related cognitive decline and neurodegeneration, caused by deficient DNA repair. *J Neurosci*. 2011;31(35):12543–12553.
80. Andel R, Crowe M, Pedersen NL, et al. Complexity of work and risk of Alzheimer's disease: A population-based study of Swedish twins. *J Gerontol B Psychol Sci Soc Sci*. 2005;60(5): P251–258.
81. Boots EA, Schultz SA, Almeida RP, et al. Occupational complexity and cognitive reserve in a middle-aged cohort at risk for Alzheimer's disease. *Arch Clin Neuropsychol*. 2015;30(7):634–642.
82. Byrne EM, Yang J, Wray NR. Inference in psychiatry via 2-sample Mendelian randomization-from association to causal pathway? *JAMA Psychiatry*. 2017;74(12):1191–1192.
83. Sanderson E. Multivariable Mendelian randomization and mediation. *Cold Spring Harb Perspect Med*. 2021;11(2):a038984.
84. Chetty R, Stepner M, Abraham S, et al. The association between income and life expectancy in the United States, 2001-2014. *JAMA*. 2016;315(16):1750–1766.