











BRAIN COMMUNICATIONS

Thirty novel sequence variants impacting human intracranial volume

Muhammad Sulaman Nawaz,^{1,2} Gudmundur Einarsson,¹ Mariana Bustamante,¹ Rosa S. Gisladdottir,^{1,3} G. Bragi Walters,^{1,2} Gudrun A. Jonsdottir,¹ Astros Th. Skuladottir,¹ Gyda Bjornsdottir,¹ Sigurdur H. Magnusson,¹ Bergrun Asbjornsdottir,¹ Unnur Unnsteinsdottir,¹ Engilbert Sigurdsson,^{2,4} Palmi V. Jonsson,^{2,5} Vala Kolbrun Palmadottir,⁶ Sigurjon A. Gudjonsson,¹ Gisli H. Halldorsson,^{1,7} Egil Ferkingstad,¹ Ingileif Jonsdottir,¹ Gudmar Thorleifsson,¹ Hilma Holm,¹ Unnur Thorsteinsdottir,¹ Patrick Sulem,¹ Daniel F. Gudbjartsson,¹ Hreinn Stefansson,¹ Thorgeir E. Thorgeirsson,¹ Magnus O. Ulfarsson^{1,8} and Kari Stefansson^{1,2}

Intracranial volume, measured through magnetic resonance imaging and/or estimated from head circumference, is heritable and correlates with cognitive traits and several neurological disorders. We performed a genome-wide association study meta-analysis of intracranial volume ($n = 79\ 174$) and found 64 associating sequence variants explaining 5.0% of its variance. We used coding variation, transcript and protein levels, to uncover 12 genes likely mediating the effect of these variants, including *GLI3* and *CDK6* that affect cranial synostosis and microcephaly, respectively. Intracranial volume correlates genetically with volumes of cortical and sub-cortical regions, cognition, learning, neonatal and neurological traits. Parkinson's disease cases have greater and attention deficit hyperactivity disorder cases smaller intracranial volume than controls. Our Mendelian randomization studies indicate that intracranial volume associated variants either increase the risk of Parkinson's disease and decrease the risk of attention deficit hyperactivity disorder and neuroticism or correlate closely with a confounder.

- 1 deCODE genetics/Amgen Inc., Sturlugata 8, 102 Reykjavik, Iceland
- 2 Faculty of Medicine, School of Health Sciences, University of Iceland, Vatnsmyrarvegur 16, 101 Reykjavik, Iceland
- 3 School of Humanities, University of Iceland, Saemundargata 2, 102 Reykjavik, Iceland
- 4 Department of Psychiatry, Landspítali-National University Hospital, Hringbraut 101, 101 Reykjavik, Iceland
- 5 Department of Geriatric Medicine, Landspítali University Hospital, Hringbraut 101, 101 Reykjavik, Iceland
- 6 Department of Internal Medicine, Landspítali University Hospital, Hringbraut 101, 101 Reykjavik, Iceland
- 7 School of Engineering and Natural Sciences, University of Iceland, Taeknigardur, Dunhagi 5, 107 Reykjavik, Iceland
- 8 Faculty of Electrical and Computer Engineering, University of Iceland, Taeknigardur, Dunhagi 5, 107 Reykjavik, Iceland

Correspondence to: Kari Stefansson
deCODE genetics/Amgen Inc., Sturlugata 8
102, Reykjavik, Iceland
E-mail: kstefans@decode.is

Keywords: intracranial volume; genome-wide association study; Mendelian randomization; genetic correlation; brain structure

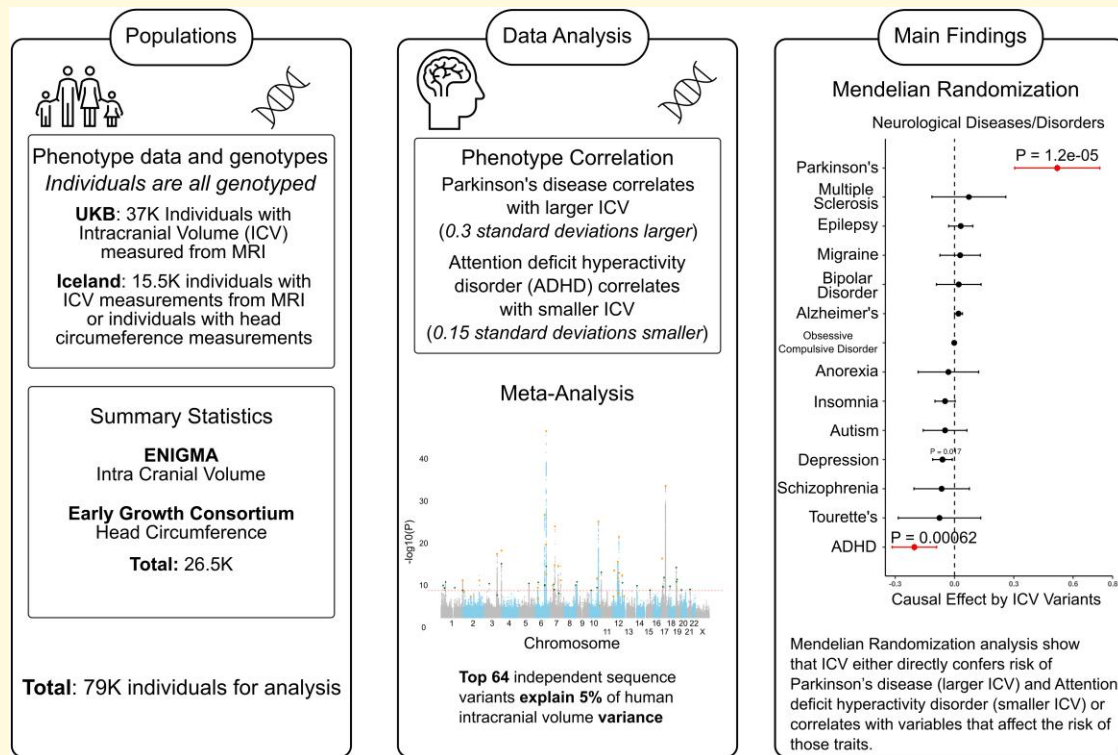
Abbreviations: ADHD = attention deficit hyperactivity disorder; BMI = body mass index; CNV = copy number variant; eQTL = expression quantitative trait locus; GWAS = genome-wide association study; HC = head circumference; ICV = intracranial volume; IV = instrumental variables; IVW = inverse-variance-weighted; LD = linkage disequilibrium; MR = Mendelian randomization; MRI = magnetic resonance imaging; pQTL = protein quantitative trait locus; SNP = single nucleotide polymorphism

Received April 13, 2022. Revised June 16, 2022. Accepted October 20, 2022. Advance access publication October 25, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of the Guarantors of Brain.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Graphical Abstract



Introduction

Overall intracranial volume (ICV) can be measured with CT or MRI and/or estimated through head circumference (HC) measurements. ICV and HC are heritable^{1,2} and highly correlated, both genetically ($rg = 0.91$)³ and phenotypically ($r = 0.73$).⁴ Studying HC, which is easily measured, allows for increased sample size for association studies. By finding sequence variants correlating with ICV, we can study the biological relationships between ICV and brain function and dysfunction. Variations in brain structure are associated with several neurological disorders.⁵⁻⁷ Genetically, attention deficit hyperactivity disorder (ADHD) is negatively correlated with ICV ($rg = -0.23$)⁸ while Parkinson's disease positively correlates with ICV ($rg = 0.35$).⁹ Nalls *et al.*⁹ reported via Mendelian randomization (MR) analysis that educational attainment increases risk of Parkinson's disease. However, they did not study the effect of ICV on Parkinson's disease.

Three meta-analyses of genome-wide association study (GWAS) results have previously uncovered sequence variants associating with ICV^{10,11} and cortical structures.¹² The largest reported GWAS for ICV ($n = 47\,316$)⁸ uncovered 18 variants, including an inversion polymorphism at 17q21.31, first described in the Icelandic population.¹³

This inversion polymorphism associates with many phenotypes, including personality traits,¹⁴ cognition,¹⁵ handedness,¹⁶ brain age¹⁷ and Parkinson's disease.⁹ Rare variants have also been associated with HC and many of these are single nucleotide polymorphisms (SNPs) conferring risk of neurodevelopmental disorders, intellectual disability, microcephaly and macrocephaly.¹⁸⁻²⁵ It has also been shown that variants in the sequence of the germline genome can impact the white⁷ or the grey matter²⁶ volumes. Furthermore, several rare recurrent copy number variants have been associated with ICV and/or regional volumes and for some the effects are dosage dependent.^{5,6,27} ICV and cortical structures are genetically correlated with cognitive functions,¹⁰ Parkinson's disease, insomnia, depression, neuroticism and ADHD.¹² Environmental factors, including infections during pregnancy, may impact ICV and are known causes of microcephaly.²⁸

Here, we present a GWAS meta-analysis of ICV ($n = 79\,174$ Europeans), where we study rare and common sequence variants and find 64 variants, of which 30 are novel. Ten of the associated variants are coding, five are *cis/trans*-pQTLs, and several are *cis*-eQTLs for a single or multiple genes. We conclude from our MR analysis that ICV either directly affects neurological traits or correlates with variables that affect the risk of those traits.

Methods

Phenotyping and cohorts included in the discovery meta-analysis

ICV data

The ICVs were either determined from HC or ICV data from the samples of the participants. These measurements were adjusted for known confounders (e.g. height, sex, age, age², sex × age²), and the residuals were rank transformed and inverse normalized to use for association studies.

Iceland ICV and HC

In Iceland, the ICV data of 1392 participants were extracted from MRI acquisitions as described earlier.^{5,6} These subjects participated in the various projects at deCODE genetics/Amgen. The ICV data were adjusted for known confounders,^{5,6} the residuals were rank transformed, and inverse normalized.

Additionally, we used manual HC measurements from 12 506 adults, and 1599 subjects measured in childhood who participated in various research projects at deCODE genetics, mostly as adults. At deCODE's recruitment centre, the HC measurements were performed as a part of a comprehensive phenotyping of a general population sample (the deCODE health study). For adults, the HC measurements were performed manually using a measuring tape, while the participant remained in a seated position, and each measurement was repeated three times, documenting only the largest value. Thus, the largest possible circumference was measured, from the most prominent part of the forehead above the ears to the most prominent part of the crown. While for children, HC measurements were performed during routine development assessment by Icelandic healthcare staff, using a measuring tape, with the child lying down, from the most prominent part of the forehead above the ears to the most prominent part of the crown.

The HC measures were also adjusted for known confounders (height, sex, age, age² and sex × age²) and the residuals were rank transformed, and inverse normalized. The Pearson correlation between the ICV and HC measurements is high ($N_{ICV + HC \text{ data}} = 1392$, $r = 0.69$, $P = 6.27 \times 10^{-92}$) as close to reported correlation ($r = 0.73$, $P < 0.01$).⁴ The residual of the inverse normalized, rank transformed and adjusted data of ICV and HC were combined (used ICV data where both ICV and HC were available) and used as a quantitative trait to run for association analysis. All of the participants (or their parents/guardian in case of minor) of the study gave written informed consent, in accordance with the declaration of Helsinki, and study was approved by the Icelandic Data Protection Authority and the National Bioethics committee (referral codes: VSN-15-241, VSN-09-098, and VSNb2015120006/03.01 with amendments, and VSN-16-093).

UKB ICV

The ICV processed data of 39 283 UK Biobank (UKB) participants subset of the 500 000 UKB study participants, was

received for those who underwent an MRI acquisition.²⁹ After the quality control checks, outliers' removal, European ancestry filtering, and additional filtering a final set of 37 100, subjects were retained for the final study. The ICV phenotype (volume of estimated total intra cranial, whole brain) was retrieved from UKB using field code '26 521' as described in Jansen *et al.*¹⁰ After the quality control criteria, the raw data were rank-transformed inverse normalized, and adjusted for known confounders (height, sex, age, age², sex × age and pc1-pc20). The residual of the inverse normalized adjusted data was used as a quantitative variable for association testing. This study was approved through UKB license number 24 898.

ENIGMA ICV + EGG HC (head circumference)

The GWAS meta-analysis of ENIGMA ICV + EGG HC published by Haworth *et al.*³ was accessed through web-portal (link in URLs) and subsequently meta-analysed together with ICV data from Iceland and UKB.

Genotyping and imputation

In the Icelandic data set, those with ICV measurements and with others phenotypes/traits participated in the number of projects at deCODE genetics. The preparation, chip-genotyping and or whole-genome sequencing of these samples was carried out at deCODE genetics.^{30,31} Using GraphTyper³² on WGS (using GAIIX, HiSeq, HiSeqX and NovaSeq Illumina technology to a mean depth of at least ×17.8), data of 61 205 Icelanders, 42.9 million high quality sequence variants were identified. Along with WGS set, deCODE genetics has also chip-genotyped 155 250 Icelanders using one of the Illumina genotyping arrays. The genotype calls (including SNP and insertions/deletions) based on WGS set were imputed into chip-genotyped subjects and long-range phased by using haplotype sharing and genealogical information.³³

In the UKB, the samples were genotyped on two Affymetrix arrays. Initial set of 50 000 samples was chip-genotyped using the Affymetrix UK BiLEVE Axiom array.³⁴ The additional set of 450 000 samples were chip-genotyped for 850 000 sequence variants using the Affymetrix UKB Axiom® array.³⁴ Both arrays target 95% of same variants.³⁴ The chip-genotypes were used to impute for additional markers by using 1000 Genomes phase 3,³⁵ UK10K³⁶ and HRC reference panels.

Association analysis

We used linear mixed model implemented by BOLT-LMM³⁷ on the normalized and adjusted ICV data assuming additive genetic model on Icelandic, and the UKB data set to test for association of 42.9 million sequence variants. For the quantitative traits, we assume that they follow a normal distribution. The LD score regression was used to account for inflation in the test statistics which may arise due to cryptic relatedness and population stratification.³⁸ To compute the

P-value, we used likelihood-ratio test applied through in-house software.³⁰

Meta-analysis

We performed GWAS meta-analysis of ICV using three GWAS summary statistics; ICV + HC from Iceland ($N_{\text{ICV} + \text{HC}} = 15,497$, male = 7,271, female = 8759), ICV from the UKB ($N_{\text{ICV}} = 37,100$, male = 19,381, female = 17779) and published GWAS summary data of ICV + HC from ENIGMA + EGG³ ($N_{\text{ICV} + \text{HC}} = 26577$). Before performing meta-analysis, the variants in three data sets were mapped to NCBI hg38 position, later the variants for UKB and ENIGMA + EGG data set were matched to the variants in the Icelandic data set based on the allele variation. We included variants that were properly imputed in all data sets, and which have a minor allele frequency > 0.1% in more than one cohort.

In total, we tested up to 42.9 million sequence variants for association with ICV. The GWAS results from the three data sets (Iceland, the UKB, the ENIGMA + EGG (the early growth genetics consortium)) were combined using a fixed effect inverse variance mode (based on the effect estimates and standard error) allowing different allele frequencies (of genotypes) in each population, where each data set was allowed to have different allele frequency of the tested genotypes but assumed to have same effect in each population. Moreover, to control for a heterogenetic effect of the markers tested in the populations, we used a likelihood ratio test (Cochran's *Q*) and so evaluated their test statistics. Total variation in the estimates, which is due to heterogeneity, was estimated using I^2 statistic.

We used weighted Bonferroni threshold to find lead associations. To claim a novel genome-wide association, the sequence variants used in the meta-analysis were split into five classes based on their genome annotation, and the weighted significance threshold for each class was used.

Polygenic risk score calculation

Polygenic risk score (PRS) for ICV were constructed in UKB, and in Iceland using genome-wide association meta-analysis by excluding the target population to avoid any bias in PRS estimates i.e. the ICV meta-analysis of Iceland + ENIGMA + EGG was used for PRS calculation in UKB. To construct the PRS, a set of high quality and well imputed 610 000 SNPs, spanning the whole genome, were used as described earlier.³⁹ These markers are also used for long-range-phasing analysis.

To construct the PRS, the LDpred was used to derive the allele-specific weights, per-locus, for each SNP from predictor GWAS. Using LDpred, we constructed the PRS for seven weight thresholds i.e. roughly corresponds to *P*-value threshold 1, 0.3, 0.1, 0.03, 0.01, 0.003 and 0.001. In Iceland, we constructed the PRS for 172 015 directly chip-typed and well imputed subjects. We used the ICV phenotype data of 15 497 Icelanders to estimate the phenotypic variance

explained by the ICV-PRS and best weight was used to perform PRS phenoscan in Iceland. We tested 5215 binary traits and 6290 quantitative traits in Iceland. Likewise, in UKB, we constructed PRS for 487 410 participants from UKB. The variance explained for ICV by the PRS in UKB is reported in [Supplementary Fig. 1](#) and [Supplementary Table 7](#). We tested 7378 binary traits and 8227 quantitative traits in UKB to test whether ICV PRS predicts phenotypic variance of any tested trait. We used Bonferroni threshold ($P_{\text{threshold}} < 0.05/27110 = 3.2 \times 10^{-6}$) to report significant associations.

ICV versus sMRI

The ICV measure summarizes the total/global size of the human brain. The local brain regions, though adjusted for total ICV, may or may not grow proportionally to the ICV. To test whether the genetic variants associated with ICV impact global and local volume differently, we used the 64 ICV associated sequence variants to test for their effect on 115 local ICV regions, defined through FreeSurfer,^{40,41} using the Aseg and Desikan-Killiany-Tourville atlases to label cortical and sub-cortical brain regions.⁴² The Bonferroni threshold ($P < 0.05/64/115 = 6.8 \times 10^{-6}$) was used to find significant associations.

Gene expression (eQTL analysis) study using deCODE and GTEx data sets

We assessed *cis*-eQTL effects of the variants associated with ICV. For the GTEx data of 49 tissues, we retrieved pre-computed eQTL estimates for all the genes (GTEx portal <https://www.gtexportal.org/>). Additionally, we used RNA sequence data of whole blood samples ($n = 13173$) from deCODE Genetics.⁴³ For both data sets, we included genes expressed more than one transcript per million (median value) that were defined in GENCODE 26 (GRCH38) within 1Mb of the ICV variants. Altogether, we tested 3310 moderately (median transcript per million > 1) or highly (median transcript per million > 10) expressed genes in 50 tissues and performed 75 728 (combination of variant \times gene \times tissue). Detailed description of eQTL analyses for deCODE data set is described here.⁴³ The associations were significant if ICV variants are in high LD ($r^2 > 0.8$) with top-eQTLs ($P_{\text{threshold}} < 0.05/75728 = 6.6 \times 10^{-7}$).

Proteomics data analysis

The proteomics data analysis (pQTL study) was performed using SOMAscan proteomics assays (SomaLogic, Inc.) of 4719 protein levels in plasma using proteomics and genotypes data of 35 559 Icelanders as described previously.⁴⁴ For that we used pre-computed associations for *cis/trans*-pQTL signals. For 64 ICV variants, we tested whether ICV variants are in high LD ($r^2 > 0.8$) with top *cis/trans*-pQTLs ($P_{\text{threshold}} < 0.05/4719/64 = 1.7 \times 10^{-7}$).

Phenome-wide association scan using GWAS catalogue

We looked up ICV variants using NHGRI GWAS Catalogue⁴⁵ (<https://www.ebi.ac.uk/gwas/>), updated on 8th June 2021, data of GWAS signals for 5004 traits. For phenome-wide association scan (pheWAS), we tested whether ICV variants are in high LD ($r^2 > 0.8$) with reported GWAS associations.

Genetic correlation analysis

To perform bivariate phenome-wide genetic correlation analysis between ICV meta-analysis and published studies, we used GWAS summary statistics data of 1483 traits^{9,12,10,46–60} ($P < 0.05/1483 = 3.4 \times 10^{-5}$). These studies are largely represented by Caucasian populations from different data-source each with effective sample size above 5000. To estimate the genetic correlations, we used LDSC^{61,62} where 1000 genome reference panel was used to estimate the LD structure. Major-histocompatibility complex region (6p22.1–6p21.3 which is about 30Mb size) was excluded due to its complex LD pattern. It is likely that most of the published GWAS traits contains a sub-sample from the UKB or other cohort that may/may not represent sample overlap; therefore, to eliminate the possibility that correlation estimates are inflated, we also performed genetic correlation of ICV meta-analysis using leave-one-sample out (ICV excluding UKB, or Iceland, or ENIGMA + EGG).

Bidirectional causal analysis

MR analysis was used to estimate the causal effect of exposure on outcome trait. The GWAS significant variants robustly and independently associated with exposure trait were used as instrumental variables. The effect estimates from exposure were tested for causal effect on outcome trait. We used inverse-variance-weighted (IVW) regression, as well as Egger regression methods (implemented in Mendelian Randomization package⁶³) to estimate the causal effect. To compute P -value, we used t -distribution ‘ t -dist’, and for standard error the ‘random model’ was used as implemented in MendelianRandomization package.⁶³ Further to test whether the effect estimates by IVW are biased, we used MR Egger method that includes intercept. The Egger method specifically test for pleiotropy (i.e. intercept is different from zero) in IVW estimates.

Data availability

The GWAS summary statistics for ICV meta-analysis will be made available at <https://www.decode.com/summarydata/>. Other data generated or analysed in this study are included in the article and its supplementary files. Additional GWAS data which was used to combine GWAS studies for meta-analysis can be accessed from respective resource as:

ENIGMA + EGG meta-analysis summary statistics: https://archive.mpi.nl/mpi/islandora/object/mpi%3A1839_ee7452e7_34c0_4d48_9598_5c14c0f89aae.

ENIGMA + EGG GWAS data: https://archive.mpi.nl/mpi/islandora/object/mpi:1839_3ab10f18_07d4_4171_8136_d0e0d100289f?asOfDateTime=2018-11-12T10:36:05.617Z.

Software/code availability (URLs): We used the following publicly available software to analyse the data:

GraphTyper (v2.0-beta, GNU GPLv3 license) at <https://github.com/DecodeGenetics/graph typer>.

Svimmer (v0.1, GNU GPLv3 license), the structural variant merging software at <https://github.com/DecodeGenetics/svimmer>.

GCTA (v1.93.3beta2) at <https://yanglab.westlake.edu.cn/software/gcta/#Overview>.

LDpred (v1.0.8) at <https://github.com/bvilhjal/ldpred>.

LD score regression at <https://github.com/bulik/ldsc>.

qqman package at <https://github.com/stephenturner/qqman>.

MandelianRandomization package at <https://github.com/cran/MendelianRandomization>.

The genotype-tissue expression: <https://www.gtexportal.org/>.

NHGRI GWAS Catalogue:⁴⁵ <https://www.ebi.ac.uk/gwas/>.

We used version 3.6.3 of R and version 1.2.5042 of RStudio for statistical analyses and for generating graphs and figures.

Results

GWAS meta-analysis for ICV

We meta-analysed GWASs of ICV and HC from Iceland ($n = 15\,497$), ICV from the UKB ($n = 37\,100$), and ICV and HC from enhancing neuroimaging genetics through meta-analysis (ENIGMA) consortium and Early Growth Genetics (EGG) consortium³ ($n = 26\,577$) to search for sequence variants associated with ICV (Fig. 1). The meta-analysis included 42.9 million imputed variants available in the Icelandic³² and UK samples and 9.7 million in the ENIGMA/EGG sample.³ We applied a weighted Bonferroni approach that uses significance thresholds for variants based on their functional annotation class.⁶⁴ The ICV meta-analysis yielded 57 variants that met the threshold for genome-wide significance, of which 30 represent novel ICV associations (Fig. 2, Supplementary Table 1 and Supplementary Figs. 2–31).

Previous GWAS studies on ICV or HC have reported 36 variants^{10,3,11,65–67} (Supplementary Table 1). Of those 36 variants, we replicated the association of 34 variants ($P_{\text{threshold}} < 0.05$, Supplementary Table 1). Altogether, 64 ICV associations at 51 loci, including 30 novel, were significant and explain 5.0% of the variability in ICV (Supplementary Table 1 and Supplementary Fig. 32 compares effect sizes between the studies). The largest effect on ICV was conferred by a low-frequency variant at 6p21.2 (rs180819997-A, $\beta = -0.191$ SD, $P = 2.2 \times 10^{-11}$, EAF =

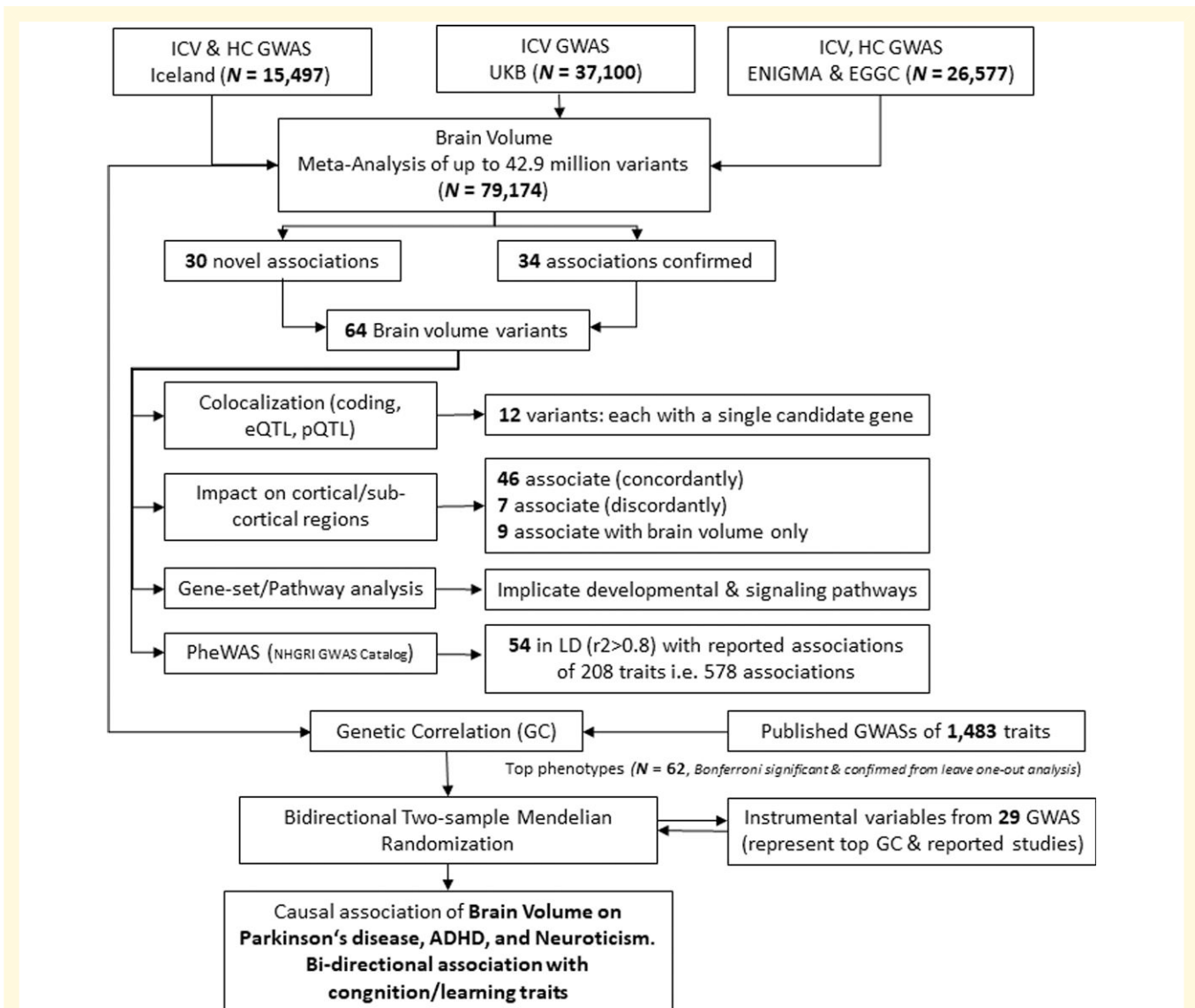


Figure 1 Workflow of the study. A GWAS meta-analysis of ICV by combining GWAS summary data from Iceland, UKB and ENIGMA + EGGC (total $n = 79\,174$) was performed. Our analysis identified 30 novel associations and confirmed 34 associations with ICV. For these 64 ICV associated variants, we performed *cis*-colocalization studies, studied their impact on cortical and sub-cortical regions (volumes), performed a PheWAS by looking up the variants and correlated variants up in the GWAS catalogue. Additionally, we studied the involvement of ICV associated genes in known pathways and gene-set terms. Finally, to understand the causal path of wide range of diseases, we used ICV associated variants (as instrumental variables) to study their impact on genetically correlated traits for Mendelian randomization analysis.

1.05%, [Supplementary Table 1](#)). This variant has recently been shown to associate with height⁶⁸ ($P = 4.9 \times 10^{-9}$). However, the effect sizes were not reported,⁶⁸ and we tested the association with height using UKB + Icelandic data ($\beta = -0.07$, $P = 3.3 \times 10^{-9}$, $n = 511\,260$). The effect estimates (in standard units) of rs180819997-A on ICV (adjusted for height) is larger than that on height ($P_{\text{heterogeneity}} = 9.0 \times 10^{-5}$).

Identification of candidate genes

To search for genes mediating the effects at the ICV loci, we determined whether the lead ICV variants or correlated ones ($r^2 >$

0.8) affected amino-acid sequence (coding), mRNA expression (*cis*-eQTL) and/or protein levels in plasma (*cis/trans*-pQTL). Ten of the 64 ICV variants are in high LD ($r^2 > 0.8$) with coding variants (missense or loss-of-function) in *MYCL*, *CDKN1B*, *HOOK2*, *FRZB*, *TGOLN2*, *XRN1*, *TNNC1*, *GLI3*, *ZNF789* and *LRRC24* ([Supplementary Table 1](#)).

We performed colocalization analysis of the 64 ICV variants with variants affecting transcript abundance using sequencing of RNA from whole blood samples ($n = 13\,173$) as well as 49 tissues from GTEx (v8).⁶⁹ Altogether, we tested 3310 moderately/highly expressed genes present within 1 Mb of the 64 variants in 50 tissues; performed 75 728

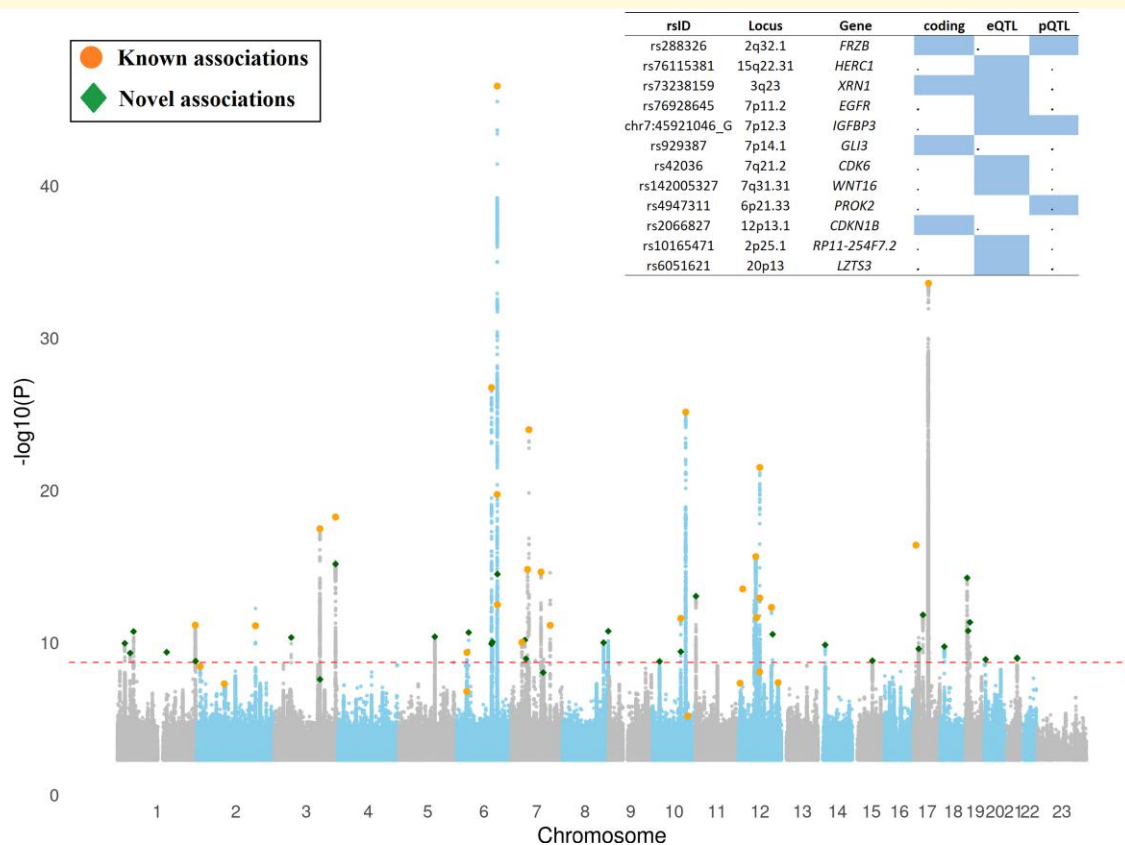


Figure 2 Manhattan-plot showing association results for ICV ($n = 79\,174$) with 42.91 million sequence variants (SNPs, In-dels and SVs). Each dot represents the position of a tested marker for the association. The x-axis represents the chromosomal position of the tested marker (where 23 refers to chromosome X), and the y-axis represents the significance ($-\log_{10}P$) of the observed association. The horizontal dotted line represents $P = 1.2 \times 10^{-9}$ ($0.05/42.9 \times 10^6$). Novel associations are highlighted with diamond shape, whereas bold filled dots represent associations of variants already reported in the scientific literature (Supplementary Table 1). The basic GWAS descriptive statistics calculated for ICV meta-analysis through LDSC^{61,62} are $h^2 = 0.0626$ (0.0044 SE), $\lambda_{GC} = 1.2818$, mean $\chi^2 = 1.4252$ and intercept = 1.0397 (0.0136 SE).

independent tests (combination of variant \times gene \times tissue tested, $P_{\text{threshold}} < 0.05/75\,728 = 6.6 \times 10^{-7}$). Twenty-six ICV variants colocalize (or are in high LD $r^2 > 0.8$) with the top *cis*-eQTLs in any of the tissues analysed, regulating the expression of 71 genes (Supplementary Table 2A).

We also performed pQTL analysis using SOMAscan proteomics assays (SomaLogic, Inc.), measuring 4907 aptamers targeting 4719 proteins in 35 559 Icelanders to test for association between ICV variants and protein levels in plasma.⁷⁰ Five of the 64 ICV variants are in high LD ($r^2 > 0.8$) with top pQTL ($P_{\text{threshold}} < 0.05/4907/64 = 1.6 \times 10^{-7}$) (Supplementary Table 2B). Of these five, two are *cis*-pQTL (*FRZB* and *IGFBP3*) and three are *trans*-pQTLs (*HS6ST2*, *PROK2* and *CR2*).

We subsequently integrated the results of these three analyses (*cis*-coding, *cis*-eQTL and *cis/trans*-pQTL) to search for genes mediating the effect of variants on ICV. In the case of 12 ICV variants, our analysis implicated a single protein coding gene (*CDKN1B*, *GLI3*, *FRZB*, *LZTS3*, *XRN1*, *WNT16*, *HERC1*, *IGFBP3*, *EGFR*, *CDK6* and *PROK2*) and one long non-coding RNA (*RP11-254F7.2*) (Supplementary Table 1).

Phenome-wide association study (PheWAS)

We determined whether the 64 ICV variants colocalize ($r^2 > 0.8$) with variants reported to associate with 5004 traits in the NHGRI GWAS Catalogue⁴⁵ (URL). Of the 64 ICV variants, 54 colocalize with 578 reported associations other than ICV (Supplementary Table 3) of which, 20 colocalize with various blood trait variants, 17 with anthropomorphic measurement variants (height, BMI and waist circumference), 14 with brain region variants (volume, area and thickness), 12 with personality traits and cognitive measures, 11 with cardiovascular disorder variants and the rest colocalized with variants associated with neurological disorders, autoimmune disorders and reproductive traits (Supplementary Table 3).

Impact on cortical and sub-cortical regions

ICV correlates phenotypically (Supplementary Table 4A), and genetically⁸ with the volumes of various cortical and

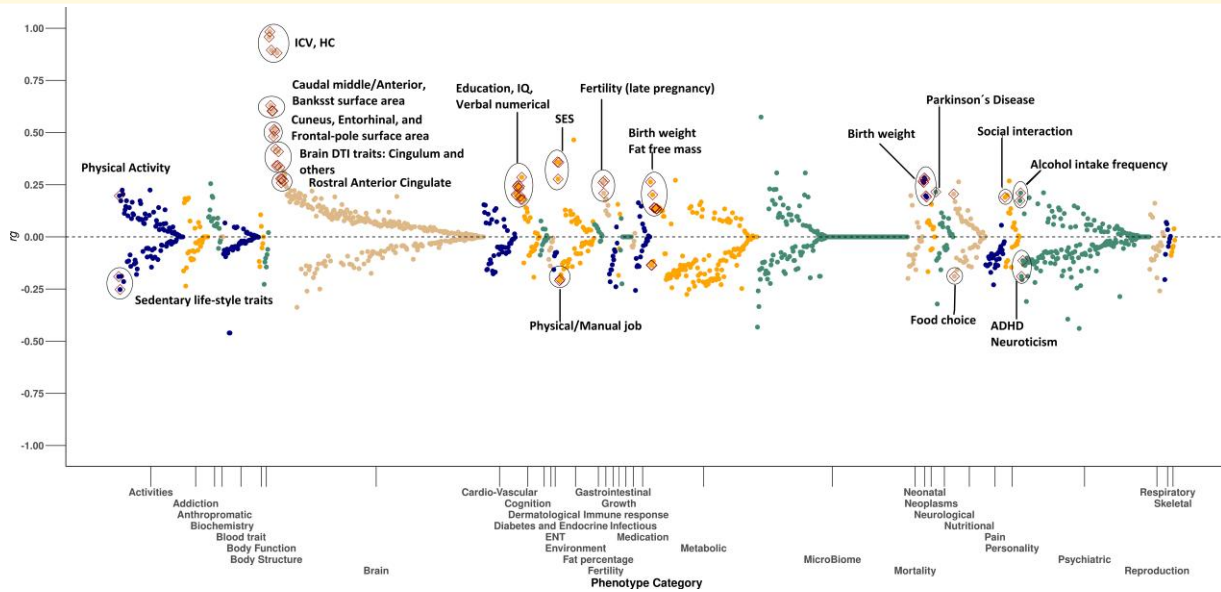


Figure 3 Phenome-wide bivariate genetic correlation between ICV and 1483 published GWAS studies estimated through LDSC.^{61,62} Each dot is an estimate of genetic correlation (r_g) between ICV GWAS meta-analysis and one of the tested GWAS traits (binned into phenotype categories), where the x-axis represents phenotype (category) and the y-axis shows its genetic correlation (r_g). The significant associations ($P_{\text{threshold}} < 0.05/1483 = 3.37 \times 10^{-5}$) are highlighted with diamond shape.

sub-cortical regions. We tested the 64 ICV variants for their association with 115 cortical and sub-cortical volumes (adjusted for ICV) extracted from structural brain MRI (sMRI) data of 37100 participants from the UKB. Of the 64 ICV variants, 53 associate with at least one sMRI trait ($P_{\text{threshold}} < 0.05/64/115 = 6.8 \times 10^{-6}$) (Supplementary Table 4B).

Gene set enrichment and pathway analysis

To understand the molecular mechanism and impact of ICV variants on known pathways/gene-sets, we performed gene enrichment/pathway analysis using the molecular signature database⁷¹ through MAGMA.⁷² MAGMA uses full GWAS summary statistics to test for regional association of genes and then uses all genes as an input for the pathway and enrichment analysis by adjusting for known confounders.⁷² The analysis found enrichment of 90 pathways/gene set terms ($P_{\text{threshold}} < 0.05/9753 = 5.1 \times 10^{-6}$) including terms describing insulin signalling, brain ventricle development, oncogenesis, neurogenesis, growth/development of cells, metabolic processes, and facial and skull development (Supplementary Table 5).

Genetic correlation analysis

We used LD score regression to perform a phenome-wide genetic correlation analysis between the ICV meta-analysis and 1483 published GWAS studies (Supplementary Table 6A).^{61,62} These studies include GWAS data of brain

anatomy and physiology, neurological, metabolic, anthropometric, cardiovascular and blood traits. Our analysis highlighted 62 of the 1483 GWAS traits showing genetic correlation with ICV meta-analysis after correcting for multiple testing and leave-one-sample-out analysis ($P_{\text{threshold}} < 0.05/1483 = 3.4 \times 10^{-5}$, Fig. 3, Supplementary Table 6A). Among the positive genetic correlations, we replicated genetic correlation of GWASs of the brain's cortical and sub-cortical regional volumes,¹² educational attainment,¹⁰ cognitive performance¹⁰ (verbal and numerical reasoning) and Parkinson's disease.⁹ Furthermore, ICV exhibited positive genetic correlation with social interaction, neonatal traits, nutritional choice and higher alcohol intake. We also replicated the previously reported negative genetic correlation of ICV with ADHD and neuroticism.⁸ Moreover, ICV correlated negatively with having a physical job, loneliness and sedentary lifestyle (Supplementary Table 6A). Some of these genetic correlations are expected, such as other brain anatomy phenotypes. We further investigate ADHD and Parkinson's disease, since these are the only disease/disorder phenotypes that survive our multiple testing threshold. We also focus on cognition and learning phenotypes, since we have access to large cohorts in Iceland and UKB for further analysis.

Phenotypic correlations

We tested the phenotypic correlations between ICV and selected traits: ADHD, Parkinson's, neuroticism and cognitive/learning traits in Icelandic and UKB data. The UKB Parkinson's disease cases have larger ICV (cases = 83,

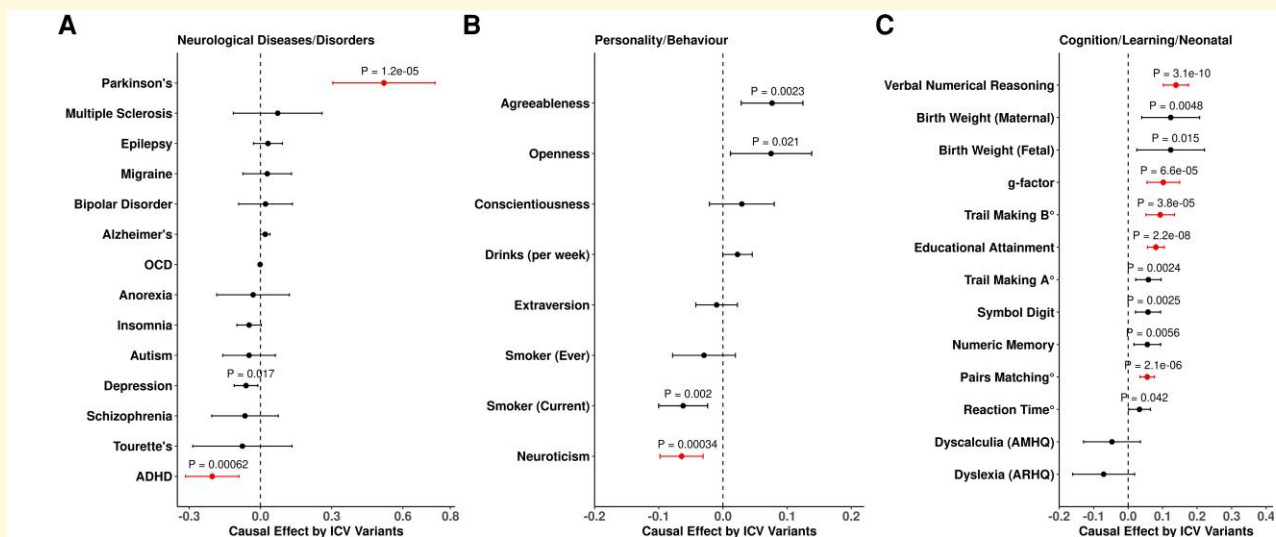


Figure 4 Causal association of instrumental variants from ICV on 34 tested traits. (A) neurological diseases/disorders, (B) personality/behaviour traits, (C) cognitive/learning/birth weight traits. The analysis was performed using a two-sample Mendelian randomization (MR) approach, the instrumental variables and their effect sizes are based on results for ICV variants versus their effects from largest available studies of the genetically correlated traits (Supplementary Table 8A). IVW (inverse variance weighted) method was used to estimate the causal effect, additionally Egger analysis was performed to detect whether IVW estimates are biased i.e. intercept is different from zero (Supplementary Table 8A). The Bonferroni significant associations ($P < 0.05/34 = 1.47 \times 10^{-3}$) are highlighted, ‘°’ refers to traits for which effect estimates were flipped for better representation (Supplementary Figs. 33–46, Supplementary Table 8A and 8B).

controls = 37 154, $\beta = 0.30$ SD, $P = 0.0072$), which is consistent with a previous report,⁷³ and in Iceland, we find that ADHD cases have smaller HC (cases = 5489, controls = 45 291, $\beta = -0.15$, $P = 1.3 \times 10^{-26}$). The phenotypic correlations for the tested traits are in keeping with the genetic correlation results (Supplementary Table 6B).

Polygenic risk score analysis

We constructed two PRSs for ICV, one excluding UKB (ICV-PRS-exUKB) and the other excluding Iceland (ICV-PRS-exICE), to allow estimation of phenotypic variance explained in the UKB and Icelandic samples, respectively. ICV-PRS-exUKB explained 5.9% of phenotypic variance in UKB ($N_{ICV} = 37\,100$), and ICV-PRS-exICE explained 8.8% of phenotypic variance in Iceland ($N_{ICV} = 1328$) (Supplementary Fig. 1 and Supplementary Table 7).

Mendelian randomization analysis

A recent study examined the shared heritability of common neurological disorders using genetic correlation data.⁷⁴ We employed a two-sample MR approach using the 64 ICV associated variants and top variants from available meta-analyses as instrumental variables (IVs) to study the causal effect on traits genetically correlated with ICV as well as the traits studied by Anttila *et al.*⁷⁴ (There are large meta-analyses available in several categories: neurological diseases/disorders, personality and behavioural traits, and cognition/learning/neonatal

traits)^{9,48–60,75–77} (Supplementary Table 8A, $P_{\text{threshold}} < 0.05/35 = 1.4 \times 10^{-3}$).

The MR results were consistent with positive causal effects of ICV variants on Parkinson’s disease ($\beta = 0.520$, $P = 1.22 \times 10^{-5}$), and cognitive traits; including verbal numerical reasoning/fluid intelligence ($\beta = 0.139$, $P = 3.07 \times 10^{-10}$), g factor ($\beta = 0.102$, $P = 6.62 \times 10^{-5}$), trail making test B ($\beta = -0.093$, $P = 3.79 \times 10^{-5}$), educational attainment ($\beta = 0.073$, $P = 9.18 \times 10^{-8}$), and pairs matching ($\beta = -0.055$, $P = 2.11 \times 10^{-6}$) and a negative causal effect on ADHD ($\beta = -0.203$, $P = 6.16 \times 10^{-4}$), and on neuroticism ($\beta = -0.064$, $P = 3.37 \times 10^{-4}$) (Fig. 4, Supplementary Table 8B). For Parkinson’s disease, this corresponds to a 68% increase in disease risk per SD of ICV and an 18% decrease in the risk of ADHD per SD of ICV. The effect estimates from the 35 tested outcome studies are mostly made up of samples of Caucasian origin and may contain some overlap of samples from the UKB or ENIGMA or EGG used in our ICV meta-analysis. To evaluate whether these results are driven by sample overlap, we performed MR analyses using ICV GWAS data with leave-one-sample-out (i.e. ICV meta-analysis excluding UKB, or Iceland, or ENIGMA + EGG GWAS (Supplementary Table 8A). All MR associations remained significant in the leave-one-sample out analyses (Supplementary Table 8B). Egger analysis of ICV variants with eight significant traits from inverse variance weighted (IVW) analyses revealed no evidence of variant pleiotropy, i.e. the intercepts were not significantly different from zero (Fig. 5, Supplementary Figs. 33–46 and Supplementary Table 8B).

To test for reverse causation, we used as IVs GWAS significant variants from Parkinson's disease⁹ ($n=90$), neuroticism¹⁴ ($n=135$), insomnia⁷⁸ ($n=780$), ADHD⁵⁰ ($n=9$), depression⁴⁸ ($n=98$), schizophrenia⁵¹ ($n=113$), bipolar disorder⁵² ($n=139$), anorexia⁵³ ($n=8$), Alzheimer's disease⁷⁵ ($n=30$), epilepsy⁵⁷ ($n=10$), multiple sclerosis⁷⁹ ($n=275$), smoking and alcohol drinking behaviour⁵⁸ ($N_{\text{smoking}}=268$; $N_{\text{drinking}}=72$), migraine⁸⁰ ($n=121$), educational attainment⁶⁰ ($n=1262$), intelligence⁶⁰ ($n=242$) and birth weight⁵⁹ ($n=144$) as exposures for their potential causal effects on ICV (Supplementary Table 8C).

The exposures of Parkinson's disease, neuroticism and migraine had a nominally significant effect on ICV in our MR analysis, but ADHD was not significant as an exposure for ICV ($P_{\text{threshold}} < 0.05/29 = 1.7 \times 10^{-3}$, $\beta_{\text{ADHD}} = 0.019$, $P_{\text{ADHD}} = 0.456$; $\beta_{\text{Parkinson's}} = 0.040$, $P_{\text{Parkinson's}} = 0.001$; $\beta_{\text{neuroticism}} = 0.126$, $P_{\text{neuroticism}} = 0.011$; $\beta_{\text{migraine}} = -0.042$, $P_{\text{migraine}} = 0.016$; Supplementary Fig. 47 and Supplementary Table 8D). Effects of birth weight, insomnia, cognitive and learning trait IVs associated with ICV after accounting for multiple testing ($\beta_{\text{birth-weight}} = 0.216$, $P_{\text{birth-weight}} = 1.16 \times 10^{-6}$; $\beta_{\text{insomnia}} = -0.192$, $P_{\text{insomnia}} = 7.07 \times 10^{-6}$; $\beta_{\text{education}} = 0.264$, $P_{\text{education}} = 4.11 \times 10^{-33}$; $\beta_{\text{cognitive performance}} = 0.190$, $P_{\text{cognitive performance}} = 8.62 \times 10^{-9}$; Supplementary Table 8D).

Discussion

Our GWAS meta-analysis of human ICV doubles the number of identified variants and provides novel insights into the biology of brain structure. Our analyses of transcriptome, proteome and coding variants highlight that 12 of the 64 ICV variants likely affect ICV via a single candidate gene each. Three of these genes (*GLI3*, *CDK6* and *FRZB*) have a priori been associated with phenotypes closely aligned with ICV/skull size. Fifty-five variants are associated with various other traits, including personality/cognition/learning, cardiovascular disorders, neurological and autoimmune disorders. We also observe a general confluence of effects in analysis using multiple markers, such as genetic correlation and MR.

Three of the markers are related to genes previously associated with microcephaly or skull bone development, phenotypes closely related to ICV. One of these is a common missense variant in *GLI3* (p.Asp1137Asn) associating with larger ICV. Rare loss of function mutations in *GLI3* have previously been associated with a premature fusing of the skull (craniosynostosis).⁸¹ Therefore, we speculate that p.Asp1137Asn may associate with a delayed fusing of the skull. The second variant is common and associates with smaller ICV and further associates with lower expression of *CDK6*. This finding is consistent with a reported recessive association between a missense variant in *CDK6* and microcephaly.⁸² Thirdly, we find a common variant in *FRZB* (p.His488Gln) that associates with larger ICV and higher *FRZB* protein expression. *FRZB* plays a role in

osteogenesis.^{83,84} The role of *FRZB* in osteogenesis suggests that *FRZB* may exert its impact on ICV by influencing skull development. Genome-wide significant associations with brain morphology have been reported in or near 5 of the 12 genes (*FRZB*,⁸⁵ *EGFR*,⁸⁶ *IGFBP3*,⁸⁷ *GLI3*⁸⁷ and *CDK6*⁸⁸). Other interesting associations reported in the GWAS catalogue include Parkinson's disease (*LZB3*), educational attainment (*FRZB* and *GLI3*) and ADHD/Externalising behaviour (*HERC1*). Other, less related phenotypes were also associated directly with most of the markers. The pheWAS of the ICV variants reveals that 55 of the 64 variants associate with a wide range of diseases and traits, including personality/cognition/learning, cardiovascular disorders, neurological and autoimmune disorders. Particularly, one previously studied marker is the one that tags the inversion polymorphism located at 17q21.31. The inversion has two haplotypes in Caucasian populations, H1 and H2. H1 associates with Parkinson's disease⁹ and larger ICV. One of the genes affected by the inversion polymorphism is *MAPT*, a candidate gene in Parkinson's disease. H2, the inverted haplotype, associates with smaller ICV, neuroticism¹⁴ and negatively with cognitive traits.⁶⁰ In order to understand the general confluence of the variants, we further analysed the data using genetic correlation and MR.

The phenome-wide genetic correlation analysis for ICV and 1483 published GWAS studies revealed genetic correlations with 62 traits. Of these 62 traits, we find two that are diseases or disorders, namely ADHD and Parkinson's disease. One of those is related to neurodevelopment while the other is related to neurodegeneration. It is a key question whether variants associated with structural changes in the brain cause neurological disorders or alternatively whether genetic predisposition to certain neurological or neurodevelopmental disorders impacts brain structure or development. We attempted to dissect the causal relationships between ICV and genetically correlated traits using bi-directional MR analyses.

It is well established that ADHD correlates with smaller HC.⁸⁹ However, few studies report a relationship between Parkinson's disease and ICV.⁷³ Here, we observe that Parkinson's cases have greater ICV than controls. Our MR analysis is consistent with ICV associated variants having a causal effect on these diseases. A fundamental assumption of MR analysis is the absence of horizontal pleiotropy,⁹⁰ i.e. the ICV sequence variants used as instruments should not systematically associate with another phenotype than ICV that causally affects the outcome. This assumption is inherently impossible to validate. However, because of the strength of the relationship between the effects of ICV variants and their disease effects, i.e. for each SD of ICV, the risk of Parkinson's disease increases by 68% and the ADHD risk decreases by 18%, it is likely that alternative phenotypes driving this relationship would have to be strongly correlated with ICV. The reverse relationship between the effects of Parkinson's disease and ADHD variants and their effect on ICV was weaker, suggesting that ICV, or

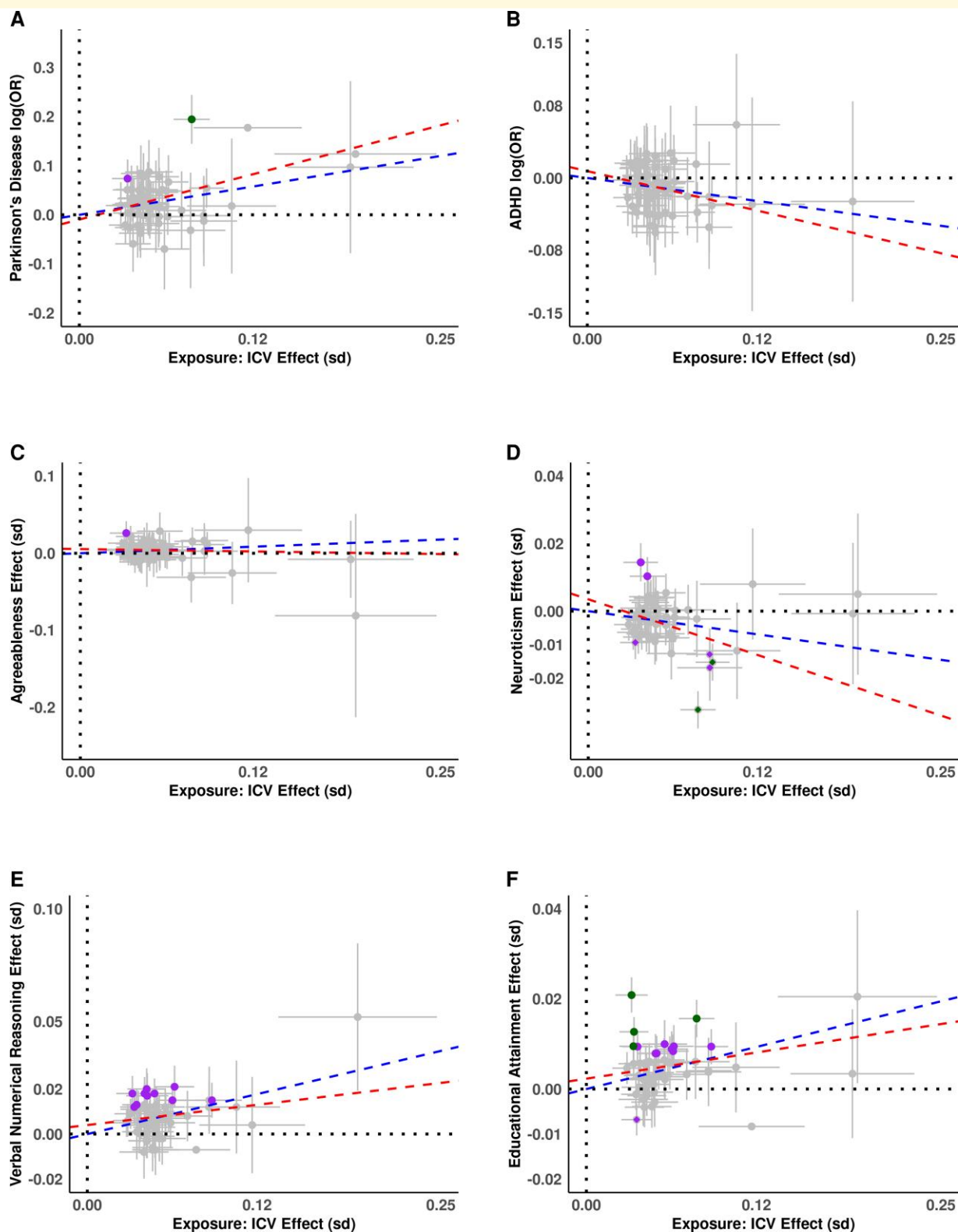


Figure 5 Effect versus effect plots of top associations from MR analysis. On x-axis are effect size for ICV and on y-axis (not always symmetric around zero) for; (A) Parkinson's disease as log(odds ratio), (B) ADHD as log(odds ratio), (C) Agreeableness as beta in SD, (D) neuroticism as beta in SD, (E) verbal numerical reasoning as beta in SD and (F) educational attainment as beta in SD. All effects are plotted for alleles with increasing ICV. Blue line represents the estimated slope from IVW (inverse variance weighted regression), and red line is estimated from MR Egger analysis including the intercept. Green dots represent conventional GWAS associations ($P < 5.0 \times 10^{-8}$) for respective y-axis trait, while purple dots are Bonferroni significant associations ($P < 0.05/64 = 7.8 \times 10^{-4}$) for respective the y-axis trait. See [Supplementary Figs. 33–46](#) for individual trait plots.

its close correlates, are likely to drive or contribute more to the relationship rather than these disorders affecting ICV.

Nalls *et al.*⁹ have previously reported a significant causal effect of educational attainment on Parkinson's disease via MR (effect = 0.162, SE = 0.040, $P = 2.06 \times 10^{-4}$). The reported causal effect of educational attainment with Parkinson's disease is weaker than that of ICV with Parkinson's disease (effect = 0.537, SE = 0.105, $P = 4.74 \times 10^{-6}$, [Supplementary Table 8B](#)). In comparison, the causal effect of ICV on educational attainment is small (effect = 0.08, SE = 0.012, $P = 2.24 \times 10^{-8}$, [Supplementary Table 8B](#)). The difference in significance is largely due to statistical power of these exposure phenotypes, where the educational attainment phenotype's sample size is over a million, compared with 80K for ICV. With this evidence we conclude that ICV is a more probable explanation as an exposure conferring risk for Parkinson's disease compared with educational attainment. This largest GWAS meta-analysis of ICV to date highlights 64 associations, of which 30 are novel. We implicate 12 genes through co-localization analyses. Our MR analyses revealed that ICV, or a closely correlated trait, has a causal effect on a neurodevelopmental disorder (ADHD) as well as on a neurodegenerative disease (Parkinson's). These findings highlight the relationship between anatomical variation and neurological and developmental disorders, underscoring the potential for applying brain volume measures, combined with genetics, to gain a foothold in understanding the complex structure-function relationships of the brain.

Acknowledgements

We thank all the participants of the study who contributed to the research, also we thank our colleagues at deCODE genetics who contributed to handling, and genotyping of biological samples, and those who helped to analyse the whole genome sequencing data. We thank the UK Biobank for providing the genotype and ICV data of the participants. The project was approved through UK Biobank license number 24 898. We also thank Simon Haworth, Beate S. Pourain and co-authors for providing ENIGMA + EGG GWAS summary data available on web-portal (URLs) for research. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund (<https://commonfund.nih.gov/GTEx>) of the Office of the Director of the National Institutes of Health, and by NCI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: [gs://gtex_analysis_v8] the GTEx Portal on 09/30/2019.

Funding

This research was in part funded by the European Commission through projects CoMorMent (H2020 grant no 847776) and painFACT (H2020 grant no 848099).

Competing interests

M.S.N., G.E., M.B., R.S.G., G.B.W., G.A.J., A.T.S., G.B., S.H.M., B.A. U.U., S.A.G., G.H.H., E.F., I.J., G.T., H.H., U.T., P.S., D.F.G., H.S., T.E.T., M.O.U. and K.S. are employees of deCODE genetics/Amgen.

Supplementary material

[Supplementary material](#) is available at *Brain Communications* online.

References

- van Der Lee SJ, Knol MJ, Chauhan G, *et al.* A genome-wide association study identifies genetic loci associated with specific lobar brain volumes. *Commun Biol.* 2019;2(1):285.
- Smit DJ, Luciano M, Bartels M, *et al.* Heritability of head size in Dutch and Australian twin families at ages 0–50 years. *Twin Res Hum Genet.* 2010;13(4):370–380.
- Haworth S, Shapland CY, Hayward C, *et al.* Low-frequency variation in TP53 has large effects on head circumference and intracranial volume. *Nat Commun.* 2019;10(1):357.
- Hshieh TT, Fox ML, Kosar CM, *et al.* Head circumference as a useful surrogate for intracranial volume in older adults. *Int Psychogeriatr.* 2016;28(1):157–162.
- Stefansson H, Meyer-Lindenberg A, Steinberg S, *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature.* 2014;505(7483):361–366.
- Sonderby IE, Gústafsson Ó, Doan NT, *et al.* Dose response of the 16p11.2 distal copy number variant on intracranial volume and basal ganglia. *Mol Psychiatry.* 2020;25(3):584–602.
- Walters GB, Gústafsson O, Sveinbjörnsson G, *et al.* MAP1B mutations cause intellectual disability and extensive white matter deficit. *Nat Commun.* 2018;9(1):3456.
- Klein M, Walters RK, Demontis D, *et al.* Genetic markers of ADHD-related variations in intracranial volume. *Am J Psychiatry.* 2019;176(3):228–238.
- Nalls MA, Blauwendraat C, Vallerga CL, *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: A meta-analysis of genome-wide association studies. *Lancet Neurol.* 2019;18(12):1091–1102.
- Jansen PR, Nagel M, Watanabe K, *et al.* Genome-wide meta-analysis of brain volume identifies genomic loci and genes shared with intelligence. *Nat Commun.* 2020;11(1):5606.
- Zhao B, Luo T, Li T, *et al.* Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat Genet.* 2019;51(11):1637–1644.
- Grasby KL, Jahanshad N, Painter J, *et al.* The genetic architecture of the human cerebral cortex. *Science.* 2020;367(6484):eaay6690.
- Stefansson H, Helgason A, Thorleifsson G, *et al.* A common inversion under selection in Europeans. *Nat Genet.* 2005;37(2):129–137.
- Nagel M, Jansen PR, Stringer S, *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genet.* 2018;50(7):920–927.
- Trampush JW, Yang MLZ, Yu J, *et al.* GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: A report from the COGENT consortium. *Mol Psychiatry.* 2017;22(3):336–345.

16. Wiberg A, Ng M, Al Omran Y, *et al.* Handedness, language areas and neuropsychiatric diseases: Insights from brain imaging and genetics. *Brain*. 2019;142(10):2938-2947.
17. Jónsson BA, Bjornsdottir G, Thorgeirsson TE, *et al.* Brain age prediction using deep learning uncovers associated sequence variants. *Nat Commun*. 2019;10(1):5409.
18. Mahmood S, Ahmad W, Hassan MJ. Autosomal recessive primary microcephaly (MCPH): Clinical manifestations, genetic heterogeneity and mutation continuum. *Orphanet J Rare Dis*. 2011;6(1):39.
19. Makhdoom EUH, Waseem SS, Iqbal M, *et al.* Modifier genes in microcephaly: A report on WDR62, CEP63, RAD50 and PCNT variants exacerbating disease caused by Biallelic mutations of ASPM and CENPJ. *Genes (Basel)*. 2021;12(5):731.
20. Nicholas AK, Khurshid M, Désir J, *et al.* WDR62 Is associated with the spindle pole and is mutated in human microcephaly. *Nat Genet*. 2010;42(11):1010-1014.
21. Kousar R, Nawaz H, Khurshid M, *et al.* Mutation analysis of the ASPM gene in 18 Pakistani families with autosomal recessive primary microcephaly. *J Child Neurol*. 2010;25(6):715-720.
22. Elsaïd MF, Kamel H, Chalhouh N, *et al.* Whole genome sequencing identifies a novel occluding mutation in microcephaly with band-like calcification and polymicrogyria that extends the phenotypic spectrum. *Am J Med Genet Part A*. 2014;164(6):1614-1617.
23. Rump P, Jazayeri O, van Dijk-Bos KK, *et al.* Whole-exome sequencing is a powerful approach for establishing the etiological diagnosis in patients with intellectual disability and microcephaly. *BMC Med Genomics*. 2015;9(1):7.
24. Otori S, Mitsuhashi S, Ben-Haim R, *et al.* A novel PAK1 variant causative of neurodevelopmental disorder with postnatal macrocephaly. *J Hum Genetics*. 2020;65(5):481-485.
25. Horn S, Au M, Basel-Salmon L, *et al.* De novo variants in PAK1 lead to intellectual disability with macrocephaly and seizures. *Brain*. 2019;142(11):3351-3359.
26. Ji Y, Zhang X, Wang Z, *et al.* Genes associated with gray matter volume alterations in schizophrenia. *Neuroimage*. 2021;225:117526.
27. Sonderby IE, van der Meer D, Moreau C, *et al.* 1q21.1 Distal copy number variants are associated with cerebral and cognitive alterations in humans. *Transl Psychiatry*. 2021;11(1):182.
28. Mlakar J, Korva M, Tul N, *et al.* Zika Virus associated with microcephaly. *N Engl J Med*. 2016;374(10):951-958.
29. Alfaro-Almagro F, Jenkinson M, Bangerter NK, *et al.* Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage*. 2018;166:400-424.
30. Gudbjartsson DF, Helgason Hs, Gudjonsson SA, *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet*. 2015;47(5):435-444.
31. Jonsson H, Sulem P, Kehr B, *et al.* Whole genome characterization of sequence diversity of 15,220 Icelanders. *Sci Data*. 2017;4(1):170115.
32. Eggertsson HP, Kristmundsdottir S, Beyter D, *et al.* GraphTyper2 enables population-scale genotyping of structural variation using pan-genome graphs. *Nat Commun*. 2019;10(1):5402.
33. Kong A, Masson G, Frigge ML, *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet*. 2008;40(9):1068-1075.
34. Bycroft C, Freeman C, Petkova D, *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.
35. Auton A, Brooks LD, Durbin RM, *et al.* A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
36. McCarthy S, Das S, Kretschmar W, *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283.
37. 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.
38. Bulik-Sullivan BK, Loh PR, Finucane HK, *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.
39. Kong A, Thorleifsson G, Frigge ML, *et al.* The nature of nurture: Effects of parental genotypes. *Science*. 2018;359(6374):424-428.
40. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*. 1999;9(2):179-194.
41. Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*. 1999;9(2):195-207.
42. Klein A, Tourville J. 101 Labeled brain images and a consistent human cortical labeling protocol. *Front Neurosci*. 2012;6:171.
43. Mikaelsson E, Thorleifsson G, Stefansson L, *et al.* Genetic variants associated with platelet count are predictive of human disease and physiological markers. *Commun Biol*. 2021;4(1):1132.
44. Skuladottir AT, Bjornsdottir G, Thorleifsson G, *et al.* A meta-analysis uncovers the first sequence variant conferring risk of Bell's palsy. *Sci Rep*. 2021;11(1):4188.
45. 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.
46. Watanabe K, Stringer S, Frei O, *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*. 2019;51(9):1339-1348.
47. Kurilshikov A, Medina-Gomez C, Bacigalupe R, *et al.* Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat Genet*. 2021;53(2):156-165.
48. Howard DM, Adams MJ, Clarke TK, *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343-352.
49. Lee JJ, Wedow R, Okbay A, *et al.* 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.
50. Demontis D, Walters RK, Martin J, *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet*. 2019;51(1):63-75.
51. Lam M, Chen CY, Li Z, *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet*. 2019;51(12):1670-1678.
52. Stahl EA, Breen G, Forstner AJ, *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet*. 2019;51(5):793-803.
53. Watson HJ, Yilmaz Z, Thornton LM, *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat Genet*. 2019;51(8):1207-1214.
54. Grove J, Ripke S, Als TD, *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431-444.
55. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry*. 2018;23(5):1181-1188.
56. Yu D, Sul JH, Tsetsos F, *et al.* Interrogating the genetic determinants of Tourette's syndrome and other tic disorders through genome-wide association studies. *Am J Psychiatry*. 2019;176(3):217-227.
57. International League Against Epilepsy Consortium on Complex Epilepsies. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nat Commun*. 2018;9(1):5269.
58. Liu M, Jiang Y, Wedow R, *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*. 2019;51(2):237-244.

59. Warrington NM, Beaumont RN, Horikoshi M, *et al.* Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet.* 2019;51(5):804-814.
60. Lee JJ, Wedow R, Okbay A, *et al.* 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.
61. 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.
62. Bulik-Sullivan BK, Loh PR, Finucane HK, *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-295.
63. Yavorska OO, Burgess S. Mendelianrandomization: An R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol.* 2017;46(6):1734-1739.
64. Sveinbjornsson G, Albrechtsen A, Zink F, *et al.* Weighting sequence variants based on their annotation increases power of whole-genome association studies. *Nat Genet.* 2016;48(3):314-317.
65. Adams HH, Hibar DP, Chouraki V, *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci.* 2016;19(12):1569-1582.
66. Ikram MA, Fornage M, Smith AV, *et al.* Erratum: Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat Genet.* 2012;44(6):539-544.
67. Wu C. Multi-trait genome-wide analyses of the brain imaging phenotypes in UK biobank. *Genetics.* 2020;215(4):947-958.
68. 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.
69. Consortium G. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369(6509):1318-1330.
70. Ferkingstad E, Sulem P, Atlason BA, *et al.* Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet.* 2021;53(12):1712-1721.
71. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. *Cell Syst.* 2015;1(6):417-425.
72. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* 2015;11(4):e1004219.
73. Laansma MA, Bright JK, Al-Bachari S, *et al.* International multicenter analysis of brain structure across clinical stages of Parkinson's disease. *Mov Disord.* 2021;36(11):2583-2594.
74. Anttila V, Bulik-Sullivan B, Finucane HK, *et al.* Analysis of shared heritability in common disorders of the brain. *Science.* 2018;360(6395):eaap8757.
75. Jansen IE, Savage JE, Watanabe K, *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet.* 2019;51(3):404-413.
76. Gormley P, Anttila Vi, Winsvold BS, *et al.* Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat Genet.* 2016;48(8):856-866.
77. Jansen PR, Watanabe K, Stringer S, *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat Genet.* 2019;51(3):394-403.
78. Watanabe K, Jansen PR, Savage JE, *et al.* Genome-wide meta-analysis of insomnia in over 2.3 million individuals indicates involvement of specific biological pathways through gene-prioritization. *medRxiv.* 2020;2020.12.07.20245209v1. <http://doi.org/10.1101/2020.12.07.20245209>
79. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science.* 2019;365(6460):eaav7188.
80. Hautakangas H, Winsvold BS, Ruotsalainen SE, *et al.* Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. *Nat Genet.* 2022;54(2):152-160.
81. Hurst JA, Jenkins D, Vasudevan PC, *et al.* Metopic and sagittal synostosis in Greig cephalopolysyndactyly syndrome: Five cases with intragenic mutations or complete deletions of GLI3. *Eur J Hum Genet.* 2011;19(7):757-762.
82. Hussain MS, Baig SM, Neumann S, *et al.* CDK6 Associates with the centrosome during mitosis and is mutated in a large Pakistani family with primary microcephaly. *Hum Mol Genet.* 2013;22(25):5199-5214.
83. Loughlin J, Dowling B, Chapman K, *et al.* Functional variants within the secreted frizzled-related protein 3 gene are associated with hip osteoarthritis in females. *Proc Natl Acad Sci USA.* 2004;101(26):9757-9762.
84. Jin C, Jia L, Huang Y, *et al.* Inhibition of lncRNA MIR31HG promotes osteogenic differentiation of human adipose-derived stem cells. *Stem Cells.* 2016;34(11):2707-2720.
85. Shadrin AA, Kaufmann T, van der Meer D, *et al.* Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *Neuroimage.* 2021;244:118603.
86. Satizabal CL, Adams HHH, Hibar DP, *et al.* Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat Genet.* 2019;51(11):1624-1636.
87. van der Meer D, Kaufmann T, Shadrin AA, *et al.* The genetic architecture of human cortical folding. *Sci Adv.* 2021;7(51):eabj9446.
88. van der Meer D, Frei O, Kaufmann T, *et al.* Understanding the genetic determinants of the brain with MOSTest. *Nat Commun.* 2020;11(1):3512.
89. Aagaard K, Bach CC, Henriksen TB, Larsen RT, Matthiesen NB. Head circumference at birth and childhood developmental disorders in a nationwide cohort in Denmark. *Paediatr Perinat Epidemiol.* 2018;32(5):458-466.
90. Burgess S, Thompson SG. *Mendelian Randomization: Methods for Causal Inference Using Genetic Variants.* CRC Press; 2021.