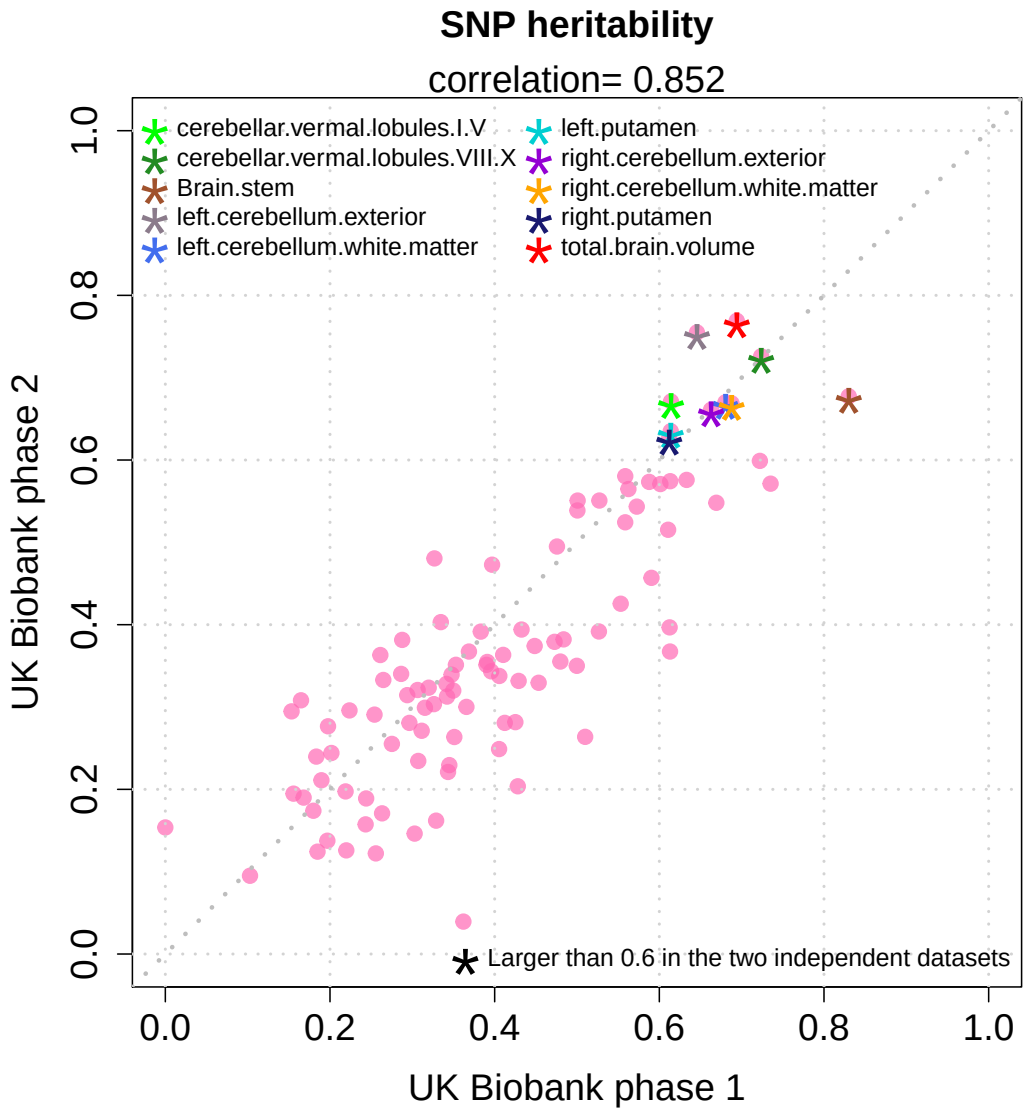
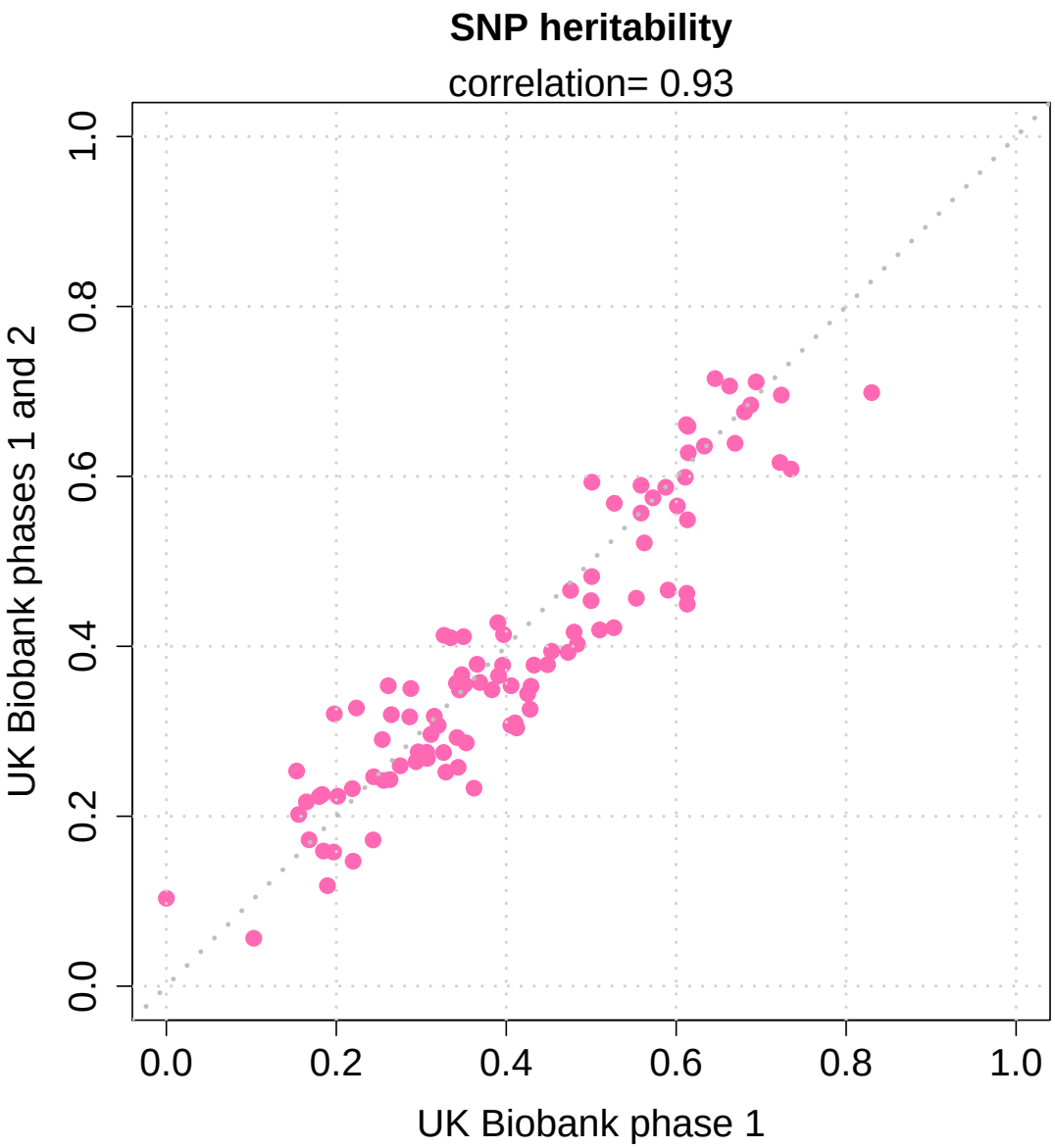
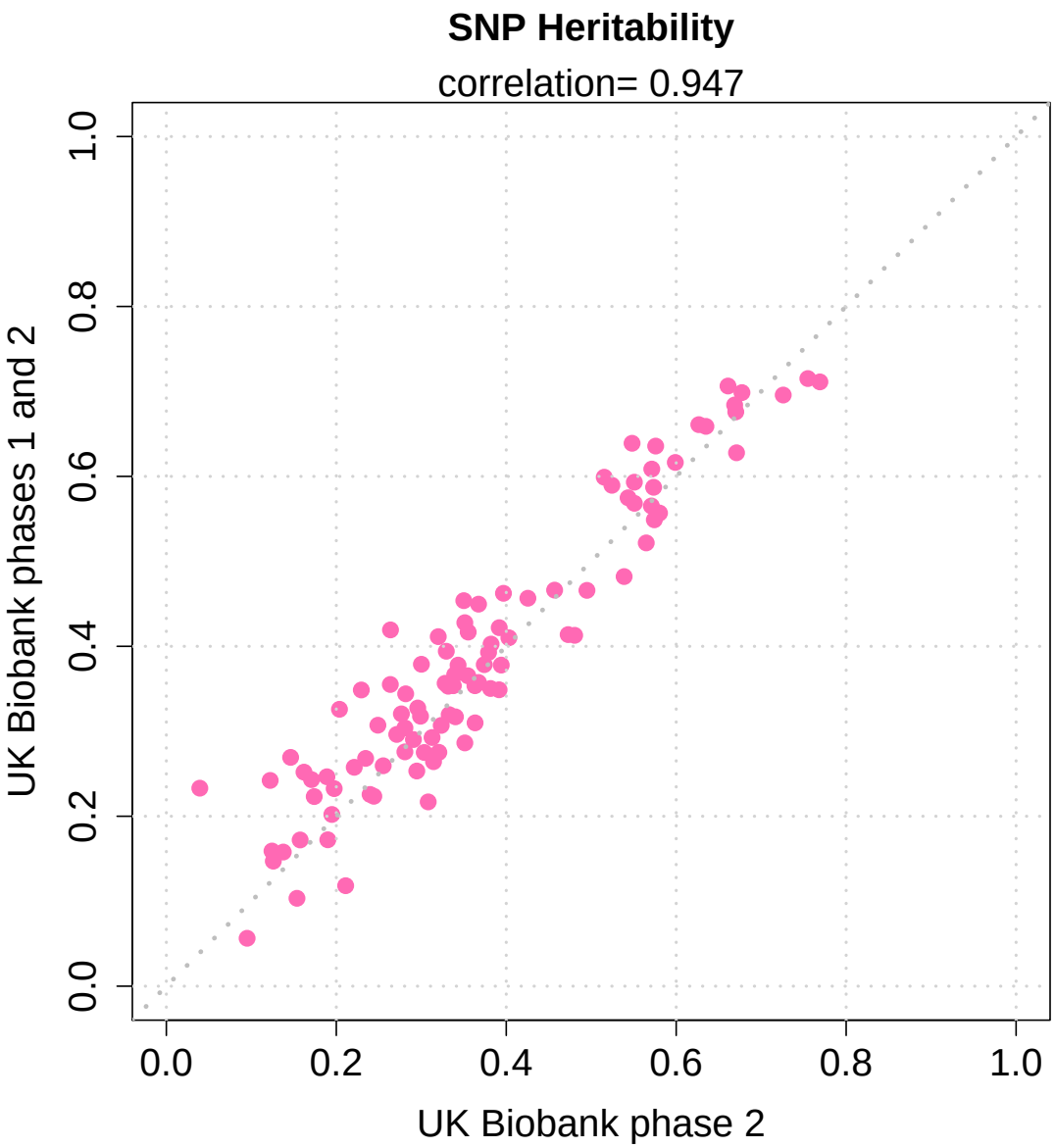
1 Supplementary figures



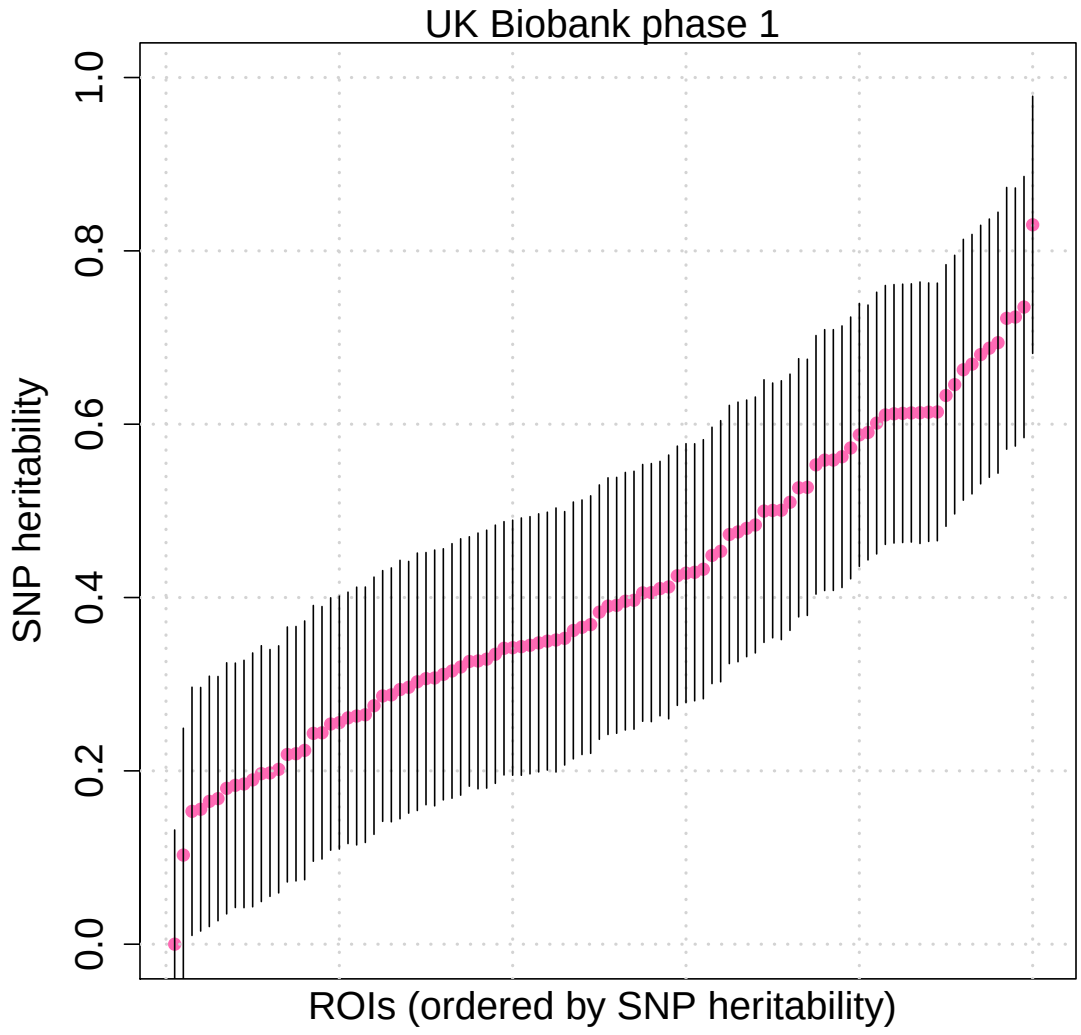
Supplementary Figure 1: Comparing SNP heritability estimates of ROI volumes in UK Biobank phase 1 (n=9,198 subjects) and phase 2 (n=10,431 subjects) data. The sample correlation coefficient of these estimates is 0.852.



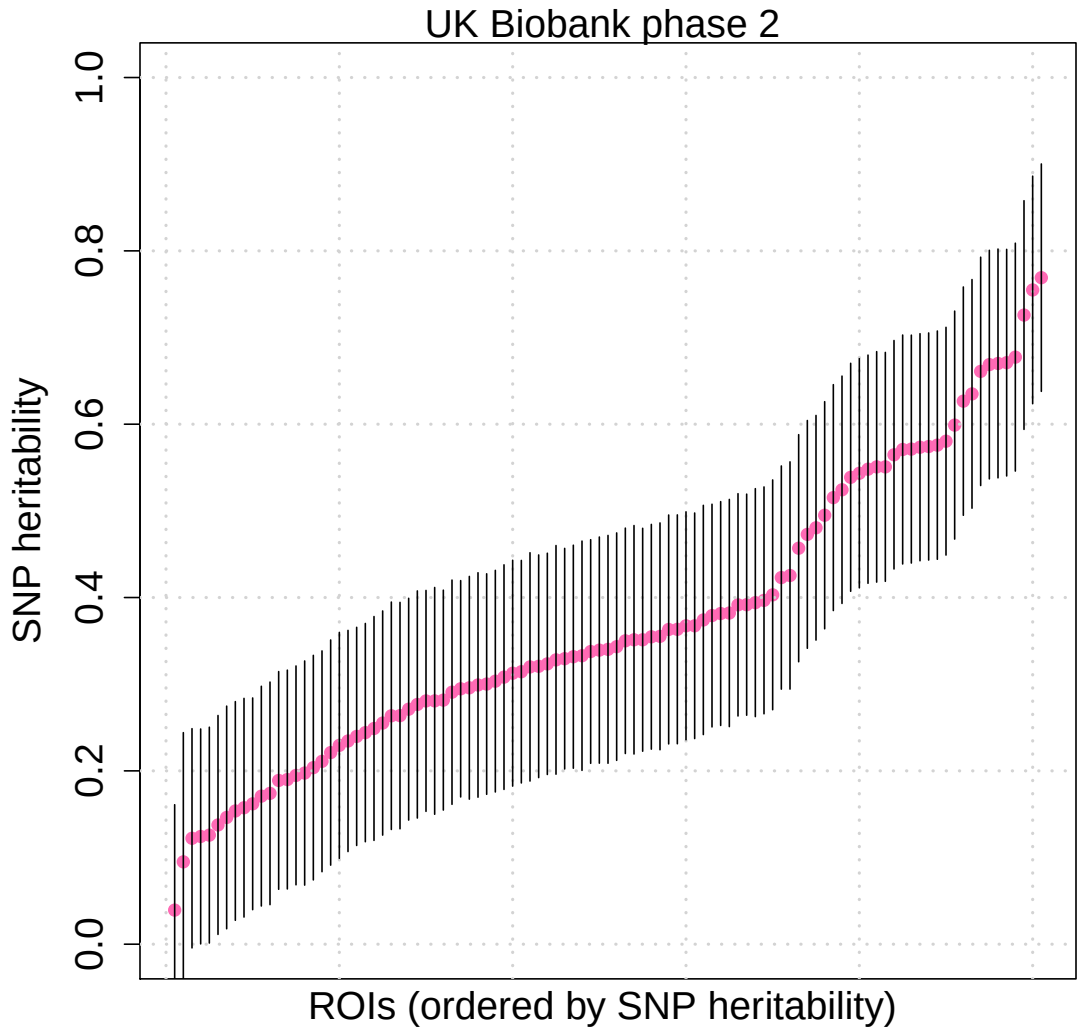
Supplementary Figure 2: Comparing SNP heritability estimates of ROI volumes in UK Biobank phase 1 (n=9,198 subjects) and the combined phases 1 and 2 (n=19,629 subjects) data. The sample correlation coefficient of these estimates is 0.930.



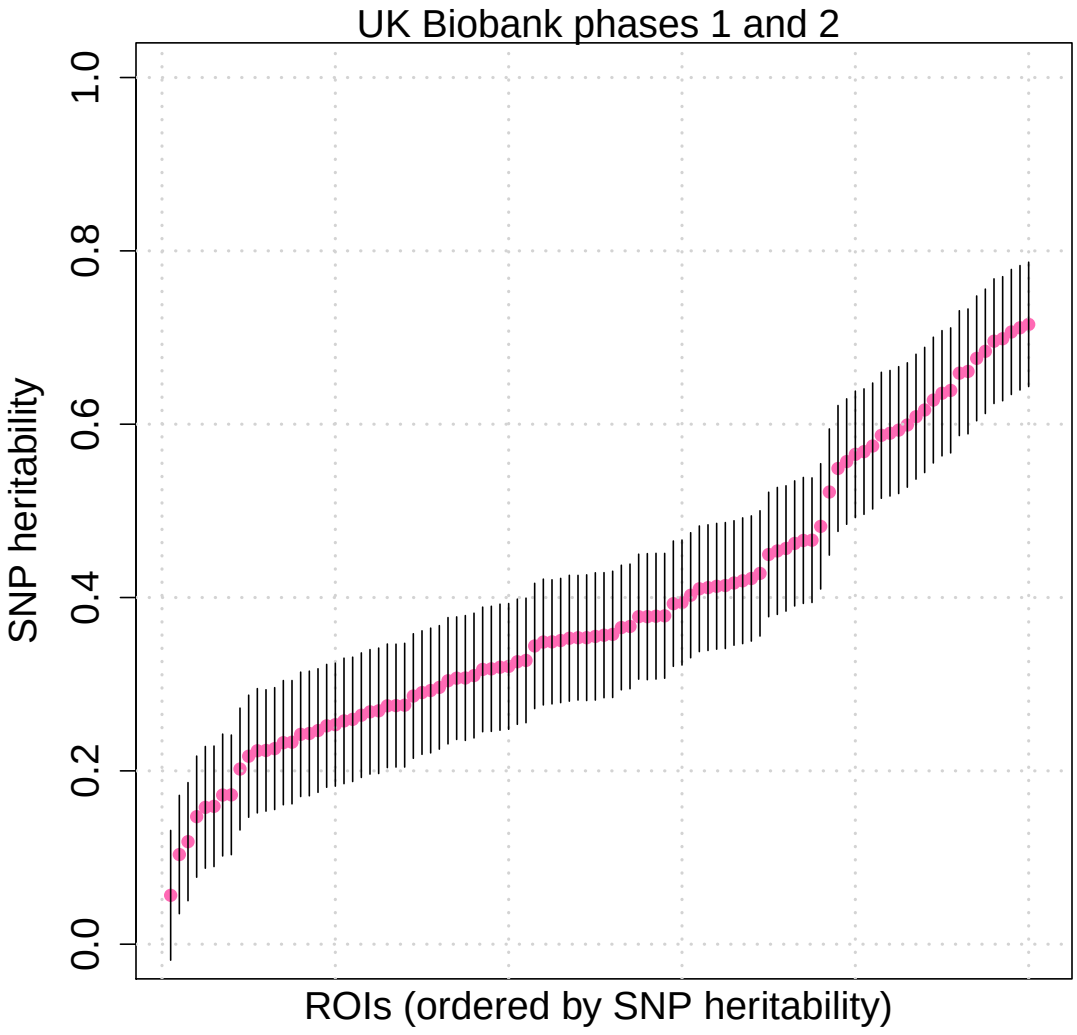
Supplementary Figure 3: Comparing SNP heritability estimates of ROI volumes in UK Biobank phase 2 (n=10,431 subjects) and the combined phases 1 and 2 (n=19,629 subjects) data. The sample correlation coefficient of these estimates is 0.947.



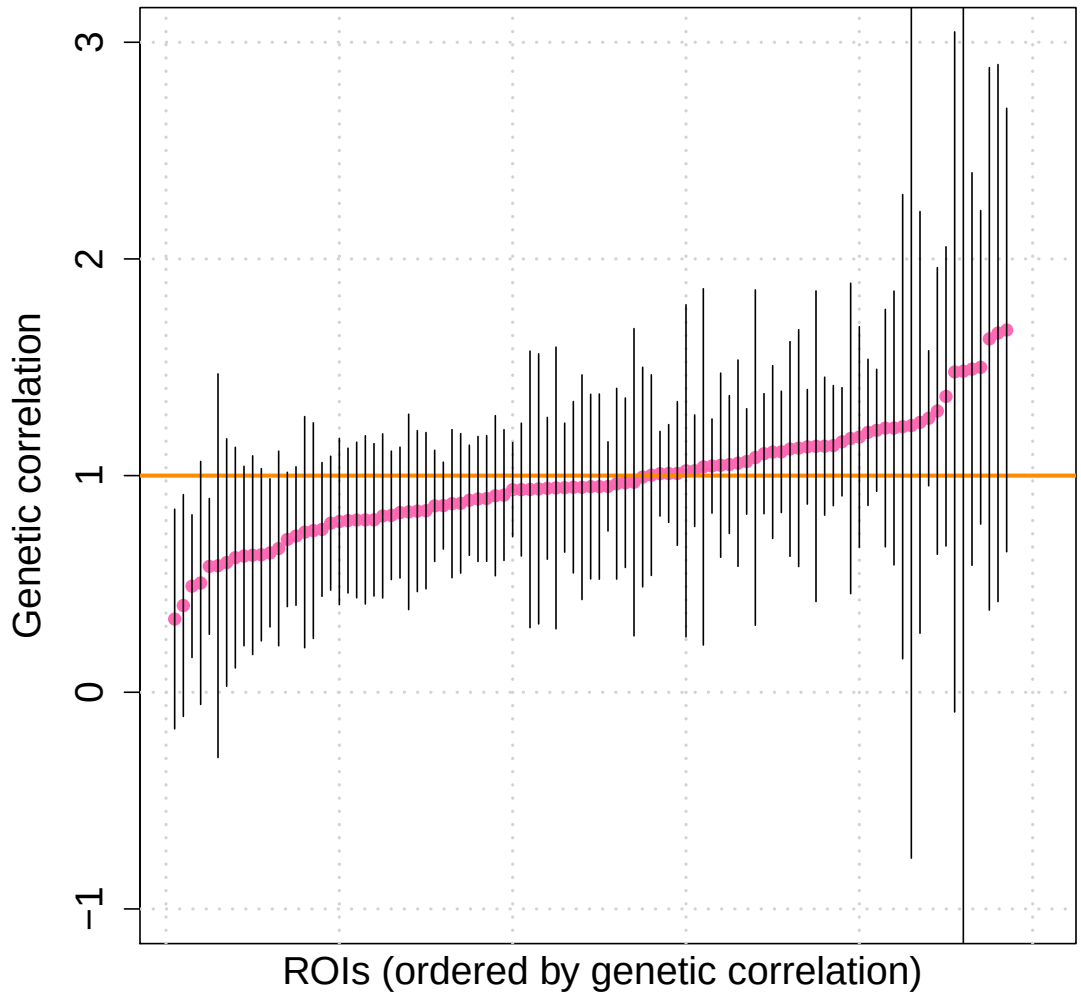
Supplementary Figure 4: SNP heritability estimates and 95% confidence intervals of ROI volumes in UK Biobank phase 1 (n=9,198 subjects) data.



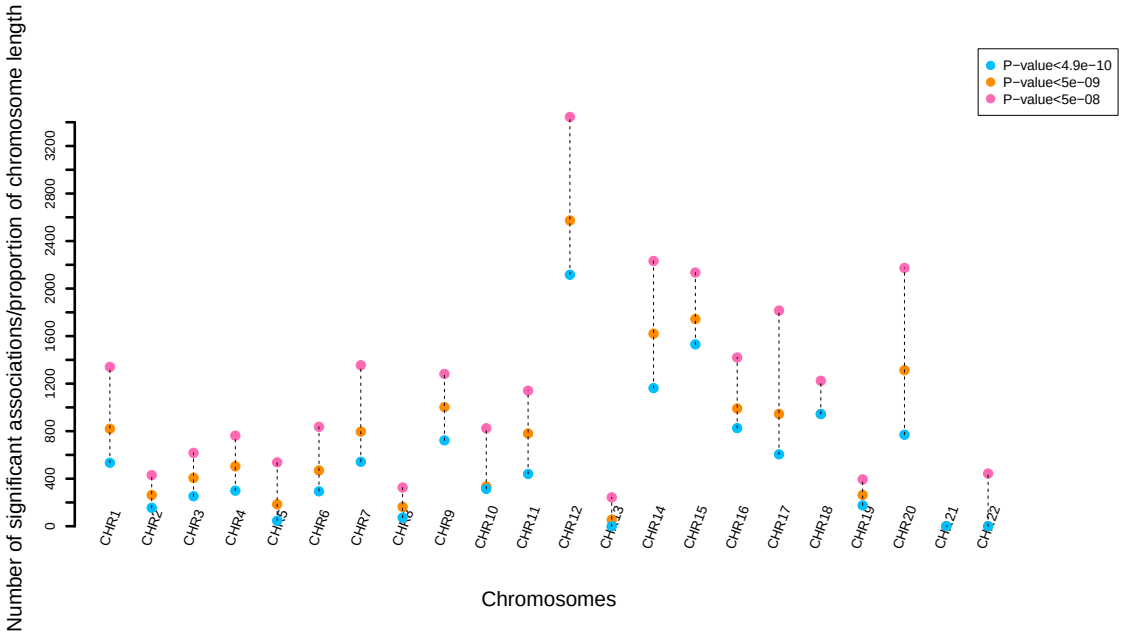
Supplementary Figure 5: SNP heritability estimates and 95% confidence intervals of ROI volumes in UK Biobank phase 2 (n=10,431 subjects) data.



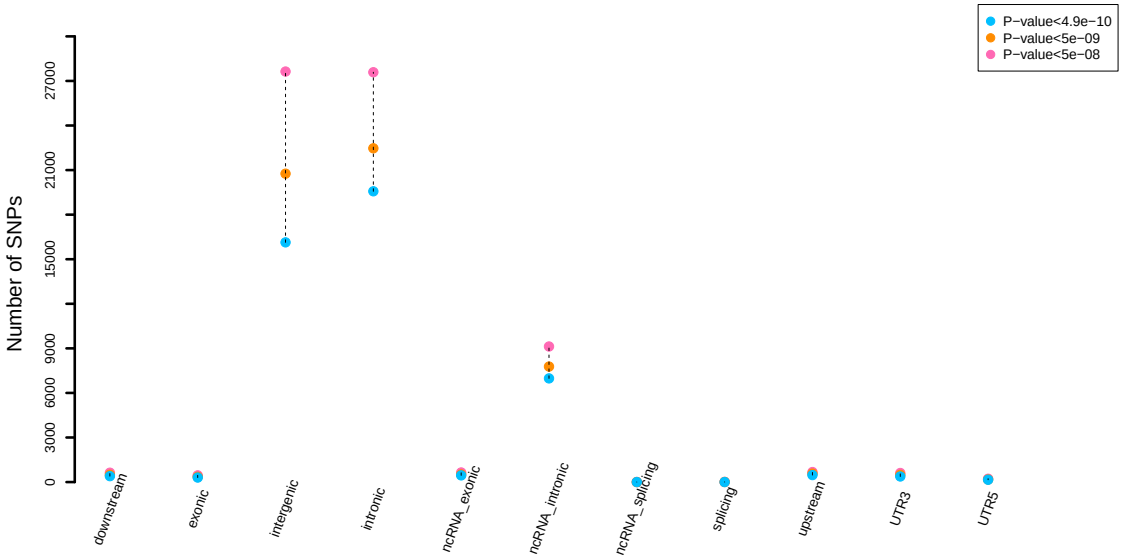
Supplementary Figure 6: SNP heritability estimates and 95% confidence intervals of ROI volumes in the combined UK Biobank phases 1 and 2 (n=19,629 subjects) data.



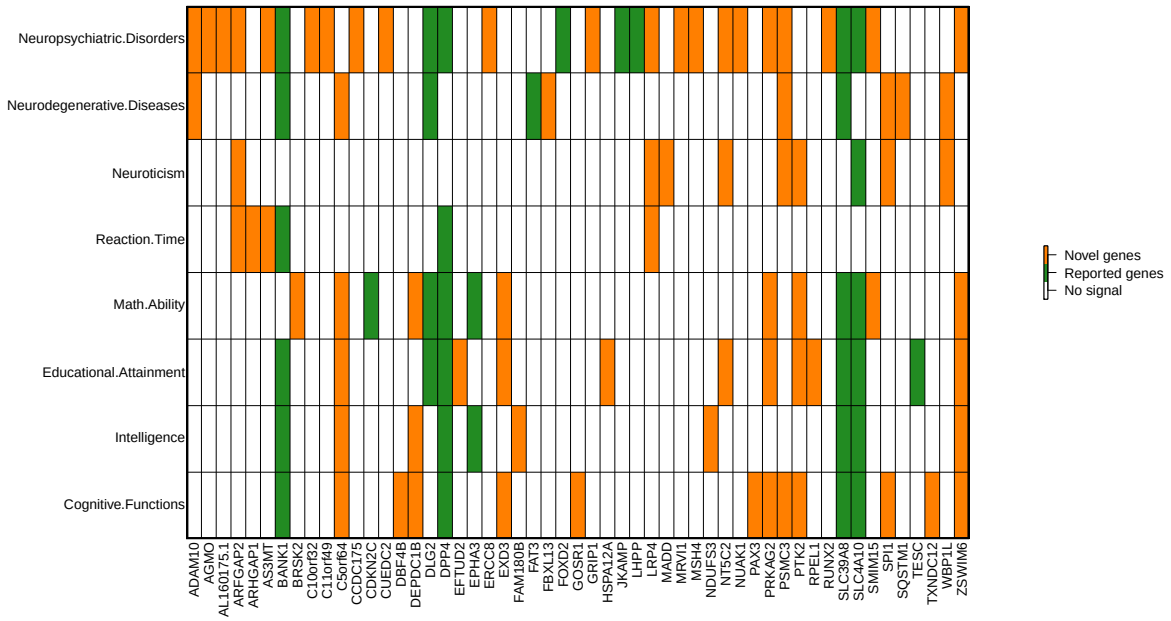
Supplementary Figure 7: Genetic correlation estimates and 95% confidence intervals for the 101 ROI volumes between UK Biobank phase 1 (n=9,198 subjects) and phase 2 (n=10,431 subjects) data.



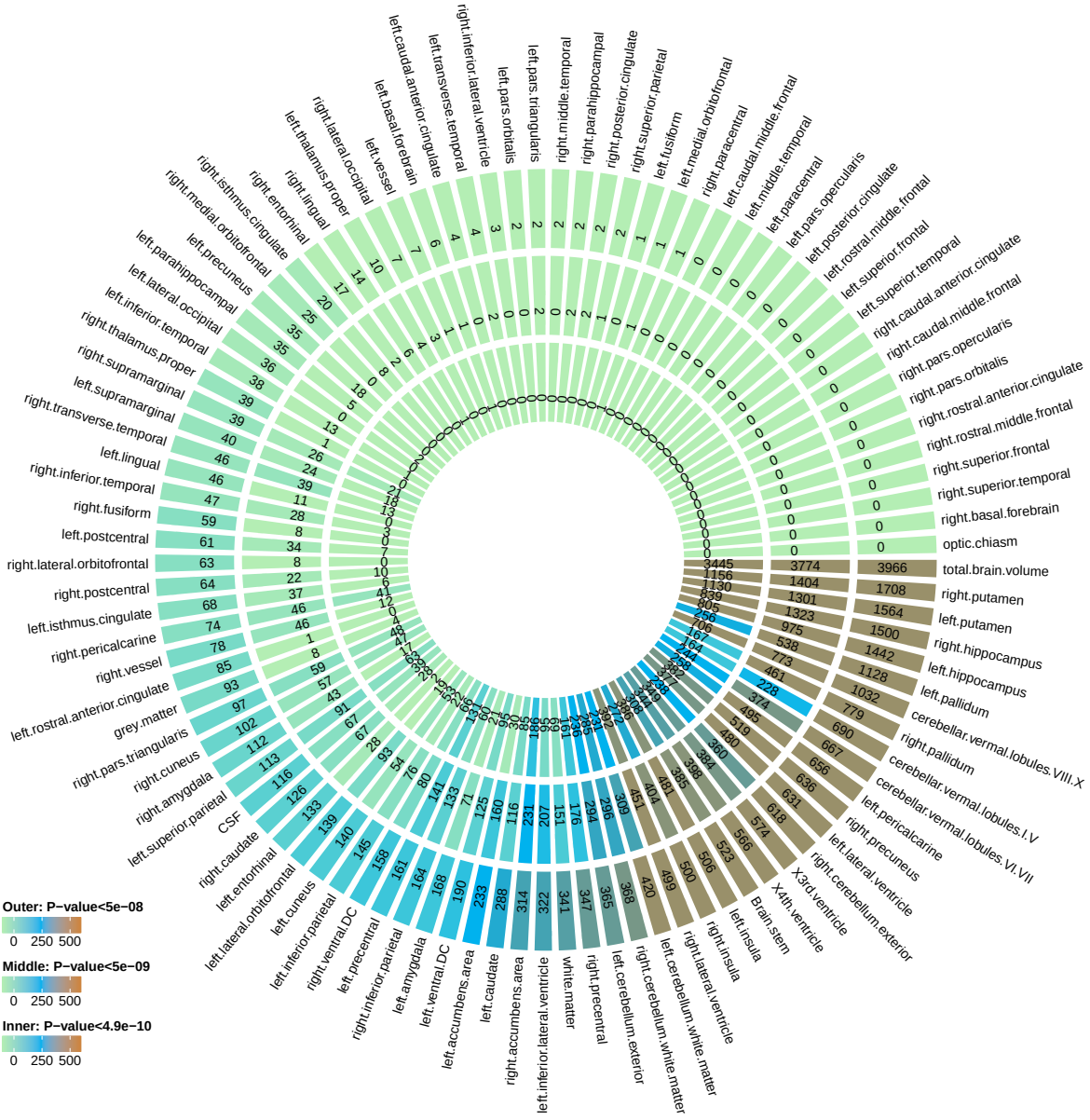
Supplementary Figure 8: Number of independent significant variant-level associations discovered in UKB GWAS (n=19,629 subjects) on each chromosome weighted by the chromosome length at different significance levels. The p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).



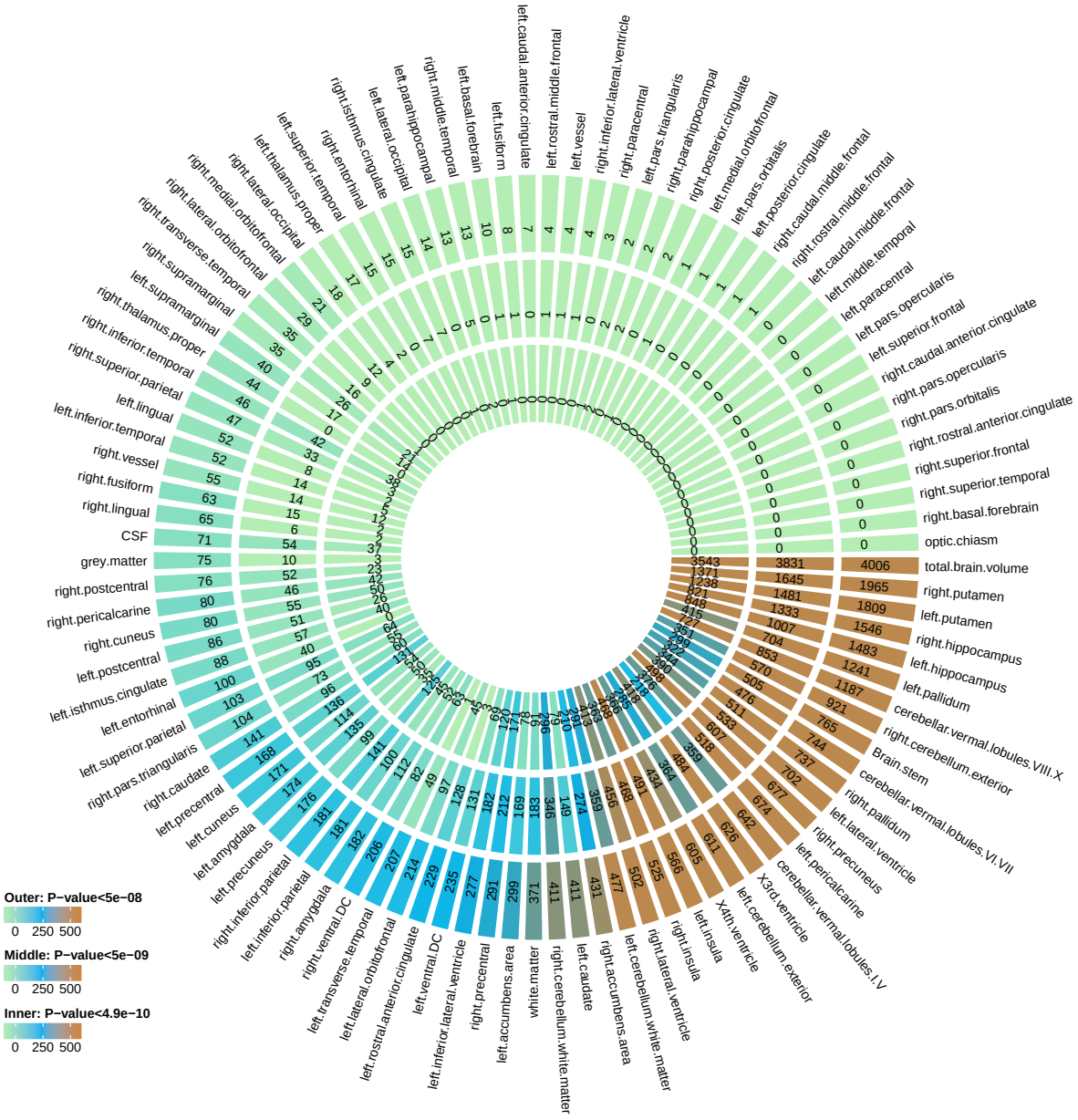
Supplementary Figure 9: Functional consequences of independent significant variants (SNPs) and variants in LD with them indicated by functional annotation assigned by ANNOVAR at different significance levels (n=19,629 subjects). The p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).



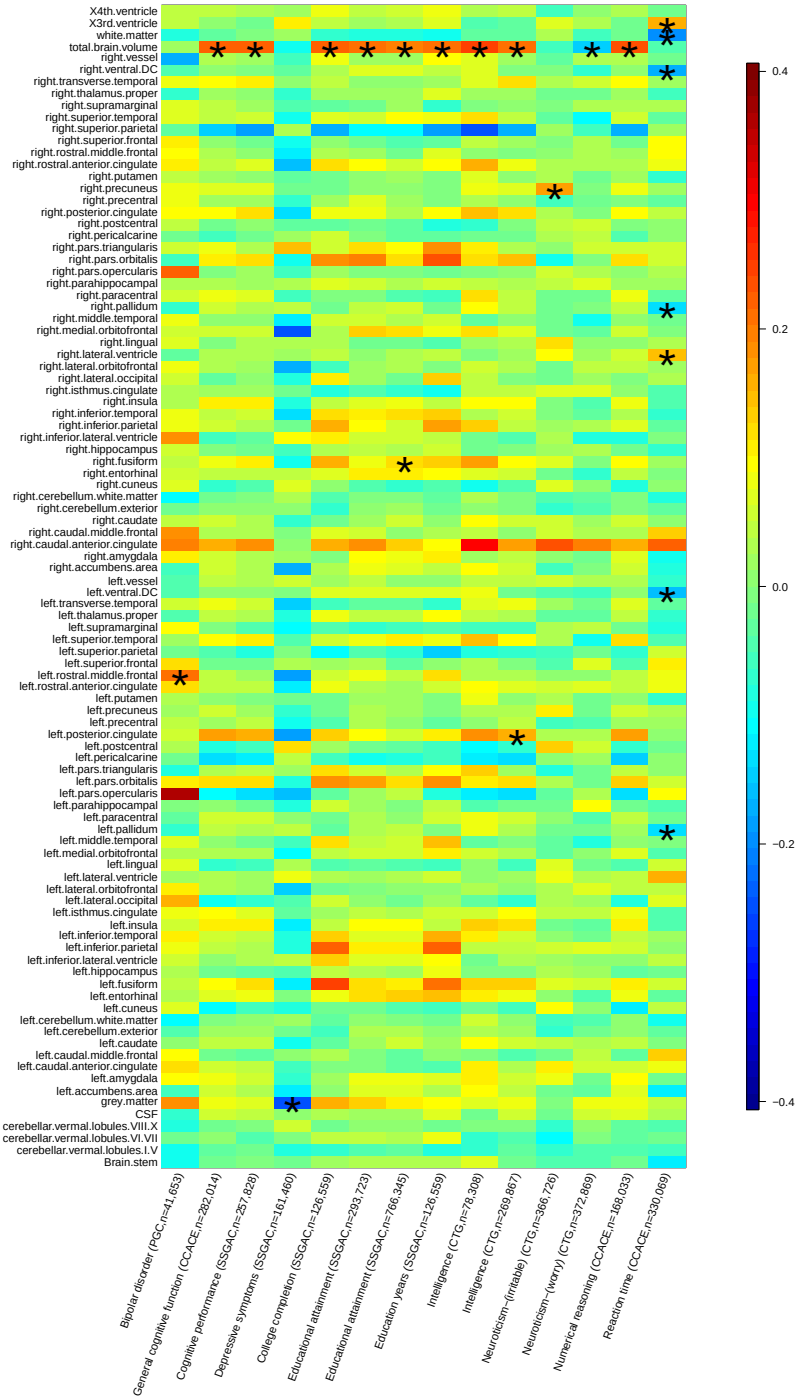
Supplementary Figure 10: Pleiotropic genes identified in functional mappings of ROI volumes (n=19,629 subjects) that have been linked to cognitive traits and mental health disease/disorders in previous GWAS. For each of the ROI-associated genes listed in the X axis, we manually checked the previously reported associations on the NHGRI-EBI GWAS catalog (version 2019-05-03, `https://www.ebi.ac.uk/gwas/`). The novel and previously reported genes of ROI volumes were labeled with two different colors (orange and green, respectively).



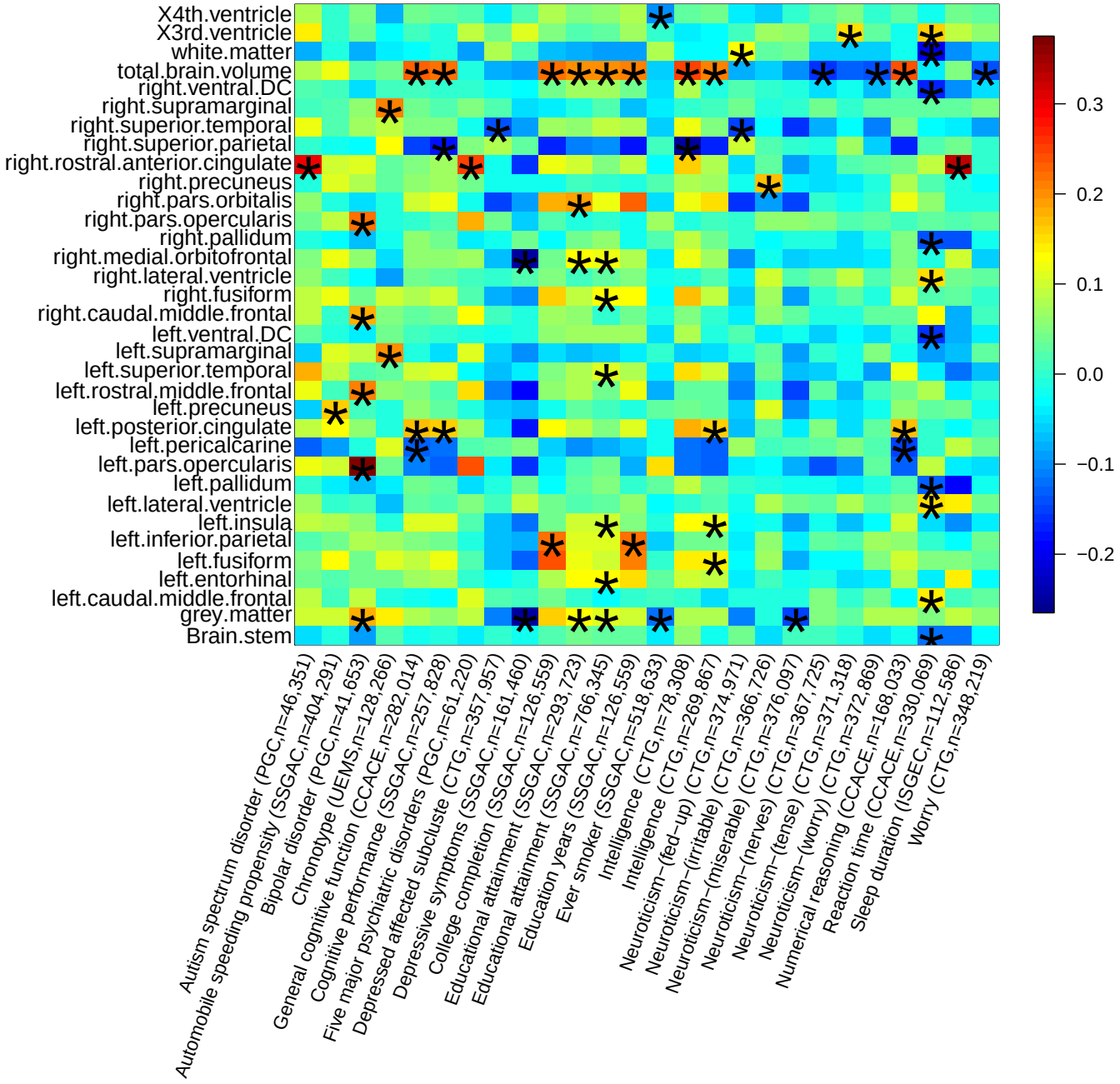
Supplementary Figure 11: Number of significant variant-level associations discovered in UKB GWAS (n=19,629 subjects) at different significance levels. Outer layer: P-value $< 5 \times 10^{-8}$; middle layer: P-value $< 5 \times 10^{-9}$; and inner layer: P-value $< 4.9 \times 10^{-10}$. The 4.9×10^{-10} threshold corresponds to adjusting for testing multiple imaging phenotypes with the Bonferroni correction. The p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).



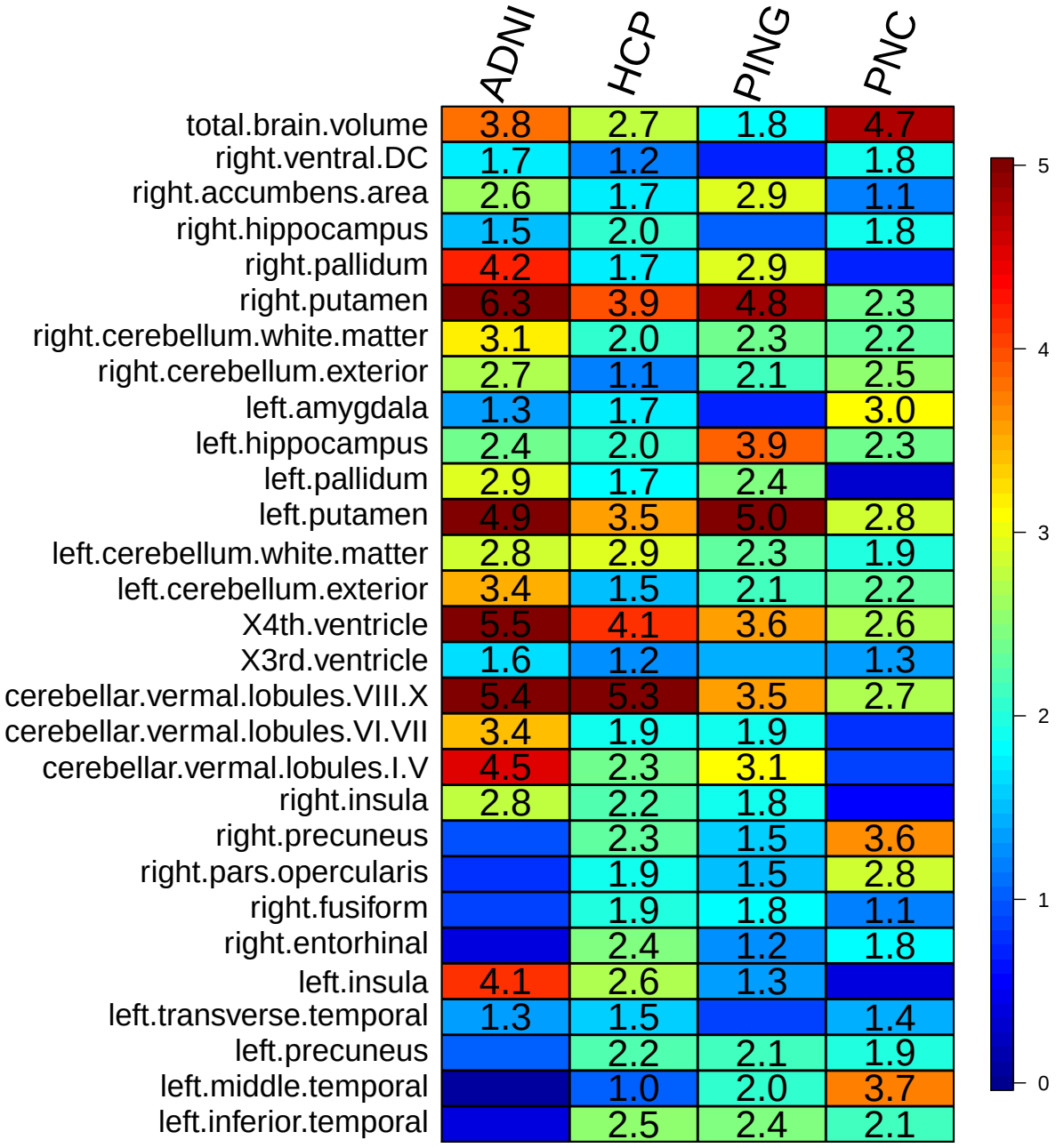
Supplementary Figure 12: Number of significant variant-level associations discovered in meta-analyzed GWAS (n=21,821 subjects) at different significance levels. Outer layer: P-value $< 5 \times 10^{-8}$; middle layer: P-value $< 5 \times 10^{-9}$; and inner layer: P-value $< 4.9 \times 10^{-10}$. The 4.9×10^{-10} threshold corresponds to adjusting for testing multiple imaging phenotypes with the Bonferroni correction. The two-sided t-test statistics were separately generated on UKB (n=19,629 subjects), PING (n=461 subjects), PNC (n=537 subjects), ADNI (n=860 subjects), and HCP (n=334 subjects) cohorts by PLINK (`https://www.cog-genomics.org/plink2/`), and then we meta-analyzed these summary results by METAL (`https://genome.sph.umich.edu/wiki/METAL`) with the sample-size weighted approach.



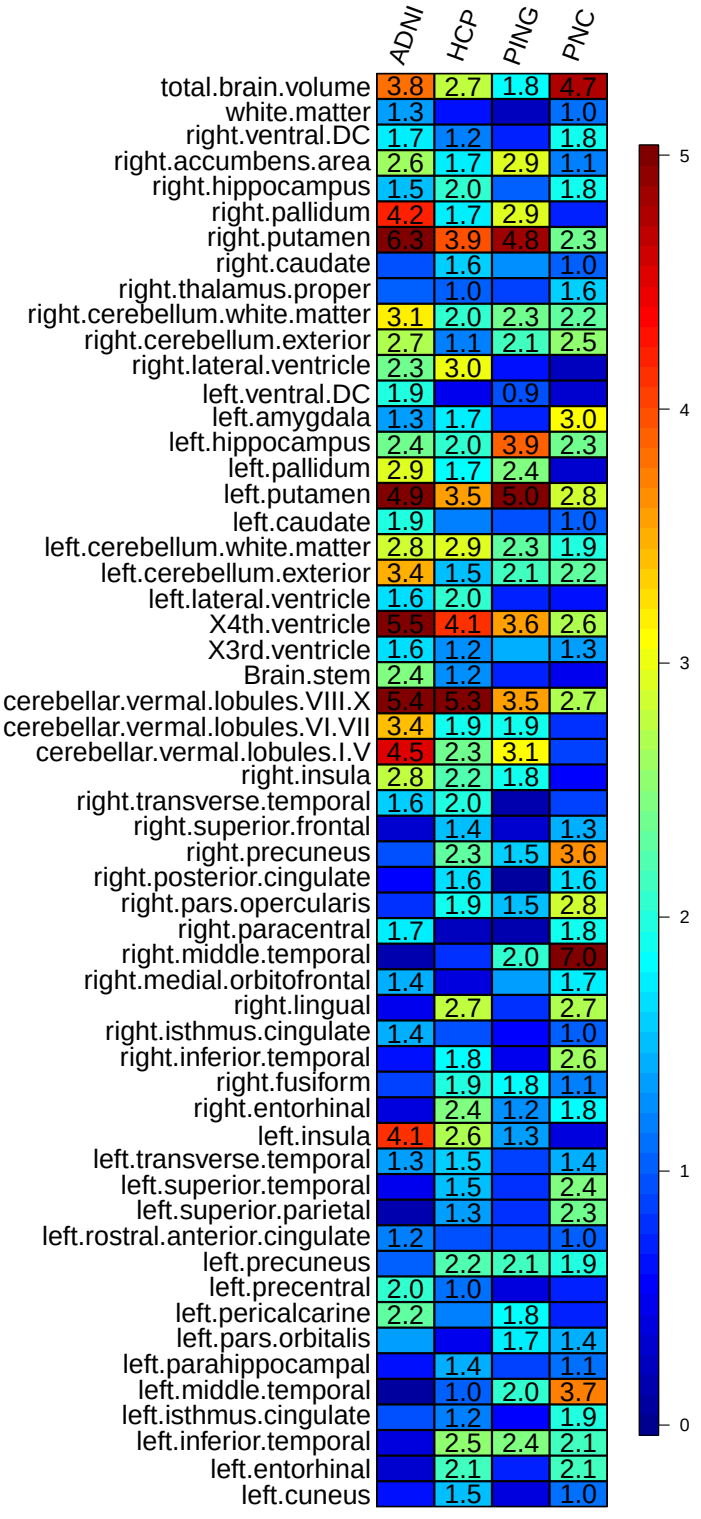
Supplementary Figure 13: Selected pairwise genetic correlations between ROI volumes (n=21,821 subjects) and other traits. Stars are significant associations after adjusting for multiple testing by the Benjamini-Hochberg procedure at 0.05 level. The LDSC software (`https://github.com/bulik/ldsc`) was used to estimate and test the pairwise genetic correlation using GWAS summary statistics. The two-sided test statistics were calculated using jackknife algorithms, more details can be found in Bulik-Sullivan et al. (`https://www.nature.com/articles/ng.3406`).



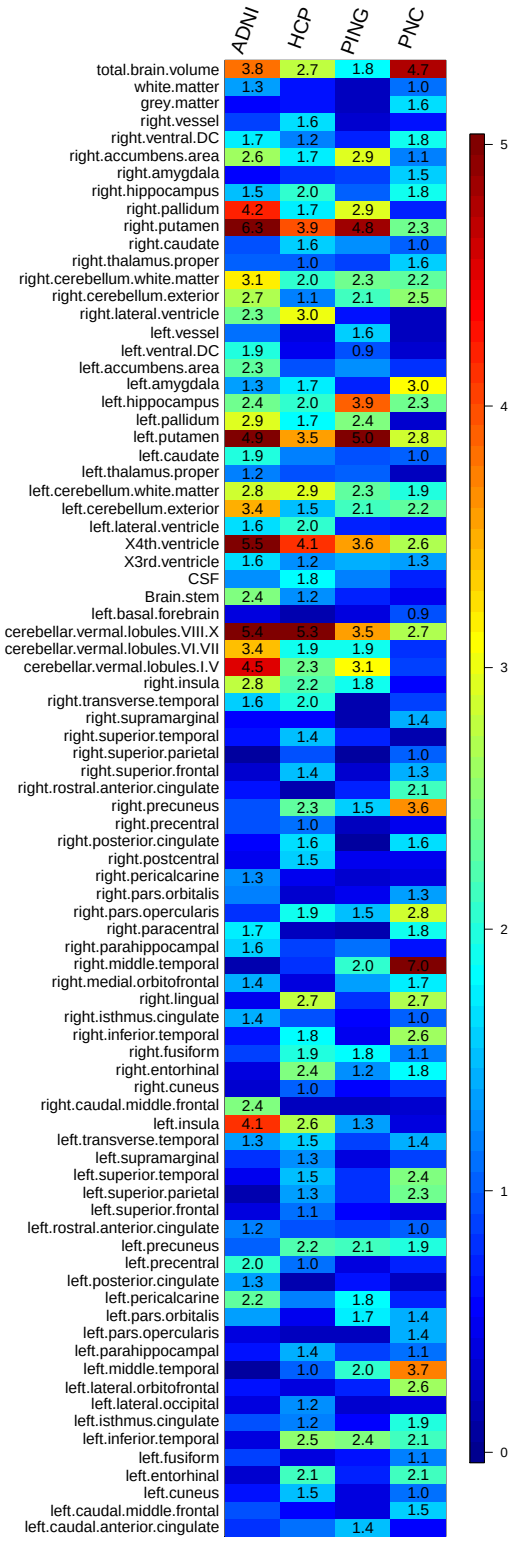
Supplementary Figure 14: Selected pairwise genetic correlation estimates between ROI volumes (n=21,821 subjects) and other traits. Stars are significant associations after adjusting for multiple testing by the Benjamini-Hochberg procedure at 0.1 level. The LDSC software (`https://github.com/bulik/ldsc`) was used to estimate and test the pairwise genetic correlation using GWAS summary statistics. The two-sided test statistics were calculated using jackknife algorithms, more details can be found in Bulik-Sullivan et al. (`https://www.nature.com/articles/ng.3406`).



Supplementary Figure 15: Prediction accuracy (incremental R-squared) of significant polygenic risk scores constructed by UKB-derived GWAS summary statistics (n=19,629 subjects) on the four independent datasets. We display the 29 ROIs that are significant in at least three of the four datasets after the Bonferroni correction.



Supplementary Figure 16: Prediction accuracy (incremental R-squared) of significant polygenic risk scores constructed by UKB-derived GWAS summary statistics (n=19,629 subjects) on the four independent datasets. We display the 56 ROIs that are significant in at least two of the four datasets after the Bonferroni correction.



Supplementary Figure 17: Prediction accuracy (incremental R-squared) of significant polygenic risk scores constructed by UKB-derived GWAS summary statistics (n=19,629 subjects) on the four independent datasets. We display the 84 ROIs that are significant in at least one of the four datasets after the Bonferroni correction.

2 Supplementary datasets and tables

All supplementary dataset and tables can be found in the zip file, here is a list the header of each dataset/table.

Supplementary Dataset 1: GWAS Manhattan plots for all the 101 ROI volumes (n=19,629 subjects).

Supplementary Dataset 2: GWAS QQ plots for all the 101 ROI volumes (n=19,629 subjects).

Supplementary Table 1: SNP heritability estimates of ROI volumes using UKB phase 1 data (n=9,198 subjects), UKB phase 2 data (n=10,431 subjects), UKB phases 1 and 2 data (n=19,629 subjects), and the genetic correlations between UKB phases 1 and 2. The SNP heritability estimates were calculated in linear mixed effect models and were tested using one-sided likelihood ratio test. The formal settings can be found in Yang et al. (`https://doi.org/10.1016/j.ajhg.2010.11.011`). We also provided the adjusted p-values after the Bonferroni correction. The LDSC software (`https://github.com/bulik/ldsc`) was used to estimate and test the pairwise genetic correlation. The two-sided test statistics were calculated using jackknife algorithms, more details can be found in Bulik-Sullivan et al. (`https://www.nature.com/articles/ng.3406`).

Supplementary Table 2: Number of significant variant-level associations, independent significant variant-level associations, and significant genetic risk loci discovered in UKB GWAS (n=19,629 subjects) at different significance levels. The p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 3: List of independent significant variant-level associations discovered in UKB GWAS (n=19,629 subjects) at different significance levels. The p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 4: Number of independent significant variant-level associations discovered in UKB GWAS (n=19,629 subjects) on each chromosome at different significance levels. The p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 5: List of significant genetic risk loci identified by UKB GWAS (n=19,629 subjects) at different significance levels. The p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 6: Number of significant genetic risk loci identified by UKB GWAS (n=19,629 subjects) on each chromosome at different significance levels. The p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 7: Summary of significant associations identified by UKB GWAS (n=19,629 subjects) at different significance levels. The p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 8: Independent significant (P-value $< 4.9 \times 10^{-10}$) variants and their correlated variants for ROI volumes that have previously been identified at P-value $< 9 \times 10^{-6}$ in GWAS of any traits listed on the NHGRI-EBI GWAS catalog (version 2019-01-31, `https://www.ebi.ac.uk/gwas/`). The independent significant variants for ROI volumes were identified by UKB GWAS (n=19,629 subjects), whose p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 9: Independent significant (P-value $< 4.9 \times 10^{-10}$) variants and their correlated variants for ROI volumes that have previously been identified at P-value $< 9 \times 10^{6}$ in GWAS of any brain volume and structure traits listed on the NHGRI-EBI GWAS catalog (version 2019-01-31, `https://www.ebi.ac.uk/gwas/`). The independent significant variants for ROI volumes were identified by UKB GWAS (n=19,629 subjects), whose p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 10: Independent significant (P-value $< 5 \times 10^{-8}$) variants and their correlated variants for ROI volumes that have previously been identified at P-value $< 9 \times 10^{-6}$ in GWAS of any traits listed on the NHGRI-EBI GWAS catalog (version 2019-01-31, `https://www.ebi.ac.uk/gwas/`). The independent significant variants for ROI volumes were identified by UKB GWAS (n=19,629 subjects), whose p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 11: List of variants that are within LD of independent significant variants discovered by UKB GWAS (n=19,629 subjects) at different significance levels. The p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 12: List of significant gene-level associations identified by MAGMA (P-value $< 2 \times 10^{-8}$) in UKB dataset (n=19,629 subjects). The p-values are raw p-values calculated using asymptotic sampling distribution via permutations. Details can be found in de Leeuw et al. (`https://doi.org/10.1371/journal.pcbi.1004219`) and the "Gene Analysis" section of the MAGMA manual (`https://ctg.cncr.nl/software/MAGMA/doc/manual_v1.07.pdf`).

Supplementary Table 13: List of novel genes identified by MAGMA gene-based association analysis (P-value $< 2 \times 10^{-8}$) in UKB dataset (n=19,629 subjects). The p-values are raw p-values calculated using asymptotic sampling distribution via permutations. Details can be found in de Leeuw et al. (`https://doi.org/10.1371/journal.pcbi.1004219`) and the "Gene Analysis" section of the MAGMA manual (`https://ctg.cncr.nl/software/MAGMA/doc/manual_v1.07.pdf`).

Supplementary Table 14: Functional consequences of independent variants (and variants in LD with them) indicated by functional annotation assigned by ANNOVAR at different significance levels. The independent variants were identified by UKB GWAS (n=19,629 subjects), whose p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 15: List of mapped genes identified in functional mapping of UKB GWAS (n=19,629 subjects) results at different significance levels. The UKB GWAS p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 16: Significant 3D chromatin (Hi-C) interactions identified in functional mapping of UKB GWAS (n=19,629 subjects) results at different significance levels. The UKB GWAS p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 17: List of mapped genes identified in functional mapping of UKB GWAS (n=19,629 subjects) results by 3D chromatin (Hi-C) interaction. The UKB GWAS p-values are raw p-values of two-sided t-test statistics generated by PLINK (`https://www.cog-genomics.org/plink2/`).

Supplementary Table 18: MAGMA gene property analysis for UKB GWAS (n=19,629 subjects) results and 14 brain tissues. The 14 brain tissues were from GTEx v7 RNA-seq database. Significant tissue groupings are highlighted in bold. The p-values are raw p-values of two-sided test statistics in linear regression. Details can be found in de Leeuw et al. (https://doi.org/10.1371/journal.pcbi.1004219) and the "Gene-level Analysis" section of the MAGMA manual (https://ctg.cncr.nl/software/MAGMA/doc/manual_v1.07.pdf).

Supplementary Table 19: Cell-type/tissue-specific chromatin-based annotation analysis for UKB GWAS (n=19,629 subjects) results. Significant tissue groupings are highlighted in bold. The p-values are raw p-values from one-sided test statistics calculated using jackknife algorithms. Details can be found in Finucane et al. (https://www.nature.com/articles/s41588-018-0081-4) and https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses.

Supplementary Table 20: DEPICT gene-set enrichment analysis for UKB GWAS (n=19,629 subjects) results. The p-values are raw p-values from one-sided tests and the test statistics were empirically calculated using re-sampling techniques with z-scores. Details can be found in Pers et al. (https://www.nature.com/articles/ncomms6890) and https://github.com/perslab/depict.

Supplementary Table 21: MAGMA gene-set analysis for UKB GWAS (n=19,629 subjects) results. Significant gene sets after correcting for testing multiple ROIs are highlighted in bold. The gene-set p-values are asymptotic p-values of one-sided test statistics in linear regression. Details can be found in de Leeuw et al. (https://doi.org/10.1371/journal.pcbi.1004219) and the "Gene-level Analysis" section of the MAGMA manual (https://ctg.cncr.nl/software/MAGMA/doc/manual_v1.07.pdf).

Supplementary Table 22: Number of significant variant-level associations discovered in meta-analyzed GWAS (n=21,821 subjects) at different significance levels. The two-sided t-statistics were separately generated on UKB (n=19,629 subjects), PING (n=461 subjects), PNC (n=537 subjects), ADNI (n=860 subjects), and HCP (n=334 subjects) cohorts by PLINK (https://www.cog-genomics.org/plink2/), and then we meta-analyzed the GWAS summary results by METAL (https://genome.sph.umich.edu/wiki/METAL) with the sample-size weighted approach.

Supplementary Table 23: Genetic correlation between several UKB ROIs volumes (TBV, left/right thalamus proper, left/right caudate, left/right putamen, left/right pallidum, left/right hippocampus, left/right accumbens area, n=21,821 subjects) and their corresponding traits studied in the ENIGMA consortium. The LD hub (`http://ldsc.broadinstitute.org/ldhub/`) was used to estimate and test the pairwise genetic correlation. The two-sided test statistics were calculated using jackknife algorithms, more details can be found in Bulik-Sullivan et al. (`https://www.nature.com/articles/ng.3406`).

Supplementary Table 24: Sources of the 50 sets of publicly available GWAS summary statistics used in this study.

Supplementary Table 25: Genetic correlation estimates and p-values between ROI volumes (n=21,821 subjects) and other traits. The LDSC software (`https://github.com/bulik/ldsc`) was used to estimate and test the pairwise genetic correlation. The two-sided test statistics were calculated using jackknife algorithms, more details can be found in Bulik-Sullivan et al. (`https://www.nature.com/articles/ng.3406`).

Supplementary Table 26: Significant Genetic correlation estimates and p-values between ROI volumes (n=21,821 subjects) and other traits. The LDSC software (`https://github.com/bulik/ldsc`) was used to estimate and test the pairwise genetic correlation. The two-sided test statistics were calculated using jackknife algorithms, more details can be found in Bulik-Sullivan et al. (`https://www.nature.com/articles/ng.3406`).

Supplementary Table 27: Prediction accuracy (incremental R-squared) and p-value of polygenic risk scores in the ten-fold cross-validation analysis within UKB (n=19,629 subjects). The p-values are asymptotic p-values of two-sided t-test statistics in linear regression.

Supplementary Table 28: Prediction accuracy (incremental R-squared) and p-value of polygenic risk scores constructed by UKB-derived GWAS summary statistics (n=19,629 subjects) on the four independent datasets. The p-values are asymptotic p-values of two-sided t-test statistics in linear regression.

Supplementary Table 29: Sample size and number of variants of the data used in each GWAS.

Supplementary Table 30: Demographic information of five datasets (UKB, ADNI, HCP, PING, and PNC).

3 Supplementary Note

Cohort information

In this study, we made use of data from five independent studies, whose demographic information is listed in Supplementary Table 30. The main GWAS was performed on the British individuals (self-reported ethnic background, Data-Field 21000) in UK Biobank study. For the other four cohorts, we only considered the unrelated European ancestry individuals (self-reported race, ethnic and family information) in GWAS. In polygenic risk score prediction, we used all available individuals (with genetic variants and phenotype data) to examine the prediction power of UKB GWAS results in testing data.

The raw MRI data were downloaded from each data resource. We processed the MRI data locally using consistent procedures via advanced normalization tools (ANTs) (Avants et al., 2011) to generate ROI volumes for each dataset. Normalization/standardization using the ANTs software had been detailed in Tustison et al. (2014) and Avants et al. (2011). We used the standard OASIS-30 Atropos template for registration and Mindboggle-101 atlases for labeling. Details of these templates and processing steps can be found in `https://mindboggle.info/data.html`, Klein and Tourville (2012) and Tustison et al. (2014).

Genotyping and quality control

We downloaded the imputed genetic variants data from UKB and HCP data resources, respectively. Genotype imputation was performed locally on the PNC, ADNI, and PING datasets using consistent procedures via MACH-Admix (Liu et al., 2013). A full description of the imputation procedures in PNC, ADNI, and PING datasets was detailed supplementary information of Zhao et al. (2018). We further performed the following genetic variants data quality controls on each dataset: 1) exclude subjects with more than 10% missing genotypes; 2) exclude variants with minor allele frequency less than 0.01; 3) exclude variants with larger than 10% missing genotyping rate; 4) exclude variants that failed the Hardy-Weinberg test at 1×10^{-7} level; and 5) remove variants with imputation INFO score less than 0.8.

For X chromosome analysis of UKB dataset, we performed quality controls using XWAS (Gao et al., 2015). Steps are detailed in `http://keinanlab.cb.bscb.cornell.edu/data/xwas/XWAS_manual_v3.0.pdf`. First, the general quality control steps were performed separately on males and females, including 1) exclude subjects with more than 10% missing genotypes; 2) exclude variants with minor allele frequency less than 0.01; 3) exclude variants with larger than 10% missing genotyping rate; 4) exclude variants that failed the Hardy-Weinberg test at 1×10^{-7} level; and 5) sex check. Next, the following X-specific quality control steps were performed on males and females together: 1) variants on chromosomes other than X were removed, as well as variants in the pseudoau-tosomal regions (PARs) on X; 2) variants were removed if they had significantly different MAF between male and female (p-value $< 1.76 \times 10^{-7}$, Bonferroni-corrected). The final number of genetic variants after all

quality controls is 283,120.

Number of significant variant-level associations of UKB GWAS

For the UKB GWAS, there were $26,425$ significant variant-level associations at the conventional 5×10^{-8} GWAS significance level and $14,448$ significant ones at the 4.9×10^{-10} significance level (that is, $5 \times 10^{-8}/101$, additionally adjusted for all 101 GWAS performed). TBV had the largest number of significant associations, which was $3,445$ at 4.9×10^{-10} significance level. In addition to TBV, left/right hippocampus, left/right putamen, and cerebellar vermal lobules VIII-X had more than 500 significant associations (Supplementary Table 2, Supplementary Fig. 11).

We then use FUMA (Watanabe et al., 2017) to define independent significant variants as significant variants that were independent of other significant variants. Left/right hippocampus, left/right putamen, and cerebellar vermal lobules VIII-X had at least 30 independent significant variants. Other ROIs that had at least 10 independent significant variants included left/right precentral, left/right cerebellum exterior, brain stem, X4th ventricle, left/right lateral ventricle, left/right cerebellum white matter, cerebellar vermal lobules I-V, cerebellar vermal lobules VI-VII, left pericalcarine, and TBV (Supplementary Table 4).

Enrichment and annotation analyses

First, gene property analysis was performed for 14 brain tissues to examine whether the tissue-specific gene expression levels were related to the associations between genes and ROI volumes by using MAGMA (de Leeuw et al., 2015). We detected five significant associations after the Bonferroni correction (that is, $14 \times 101 = 1,414$ tests), involving gene expression in brain cerebellar hemisphere and cerebellum tissues and gene's association significance with pallidum and putamen volumes (p-value $< 2.05 \times 10^{-5}$) (Supplementary Table 18). These results showed that genes with higher transcription levels on these brain tissues also had stronger genetic associations with brain ROI volumes. We also performed chromatin-based annotation analysis by stratified LDSC (Finucane et al., 2018). A total of 490 cell-type/tissue-specific DNase I hypersensitivity and activating histone marker annotations were examined (Online Methods). After the Bonferroni correction (that is, $490 \times 101 = 49,490$ tests), we found that H3K4me3 histone annotations from three brain tissues had significantly enriched contribution to per-SNP heritability of left/right accumbens area (p-value $< 6.58 \times 10^{-7}$) (Supplementary Table 19).

To gain more insights into the biological mechanisms, we used DEPICT (Pers et al., 2015) and MAGMA (de Leeuw et al., 2015) to conduct gene set analysis. DEPICT performed enrichment testing for $10,968$ reconstituted gene sets, see Supplementary Data 1 of Pers et al. (2015) for an overview of these gene sets. None of the candidates survived the Bonferroni correction (that is, $10,968 \times 101$ tests). At the 4.5×10^{-6} significance level (that is, not adjusted for testing multiple phenotypes), DEPICT showed suggestive evidence for enrichment

of 10 gene sets (Supplementary Table 20), such as "abnormal brain development" gene set (MP:0000913) enriched among variants associated with left cerebellum exterior, and "open neural tube" (MP:0000929) gene set enriched for left ventral DC. Similarly, MAGMA was performed for $10,678$ candidate gene sets, two of which were significant after the Bonferroni correction, and there were 35 more enriched gene sets when not adjusted for testing multiple ROIs (Supplementary Table 21).

Genetic correlation with other traits

In the genetic correlation (gc) analysis, we found that TBV had positive correlations with cognitive functions, education, intelligence, and numerical reasoning (gc range=$[0.20, 0.25]$, mean=0.22, p-value range=$[1.52\times10^{-11}, 3.45\times10^{-5}]$). Similar connections were also observed on several other ROIs. For example, left posterior cingulate showed positive correlation with intelligence (gc=0.16, p-value=1.38×10^{-4}), left rostral middle frontal showed positive correlation with BD (gc=0.22, p-value=5.85×10^{-5}), right fusiform had positive correlation with education (gc=0.13, p-value=1.0×10^{-4}), and right precuneus had positive correlation with neuroticism (gc=0.17, p-value=1.0×10^{-4}). Reaction time had positive genetic correlations with right lateral ventricle and X3rd ventricle (gc range=$[0.15, 0.16]$, p-value range=$[1.39\times10^{-5}, 2.04\times10^{-4}]$), and had negative correlations with left/right pallidum, left/right ventral DC, and WM (gc range=$[-0.20, -0.13]$, p-value range=$[3.80\times10^{-7}, 1.14\times10^{-4}]$). Negative genetic correlations were also found between depressive symptoms and GM (gc=-0.25, p-value=2.79×10^{-5}), and between neuroticism and TBV (gc=-0.14, p-value=2.0×10^{-4}). Gray matter volume loss has been revealed in depressed individuals (Belden et al., 2016; Grieve et al., 2013), and neuroticism and TBV may be negatively correlated (Jackson et al., 2011; Knutson et al., 2001).

PING Methods

Part of the data used in the preparation of this article were obtained from the Pediatric Imaging, Neurocognition and Genetics (PING) Study database (`http://ping.chd.ucsd.edu/`). PING was launched in 2009 by the National Institute on Drug Abuse (NIDA) and the Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD) as a 2-year project of the American Recovery and Reinvestment Act. The primary goal of PING has been to create a data resource of highly standardized and carefully curated magnetic resonance imaging (MRI) data, comprehensive genotyping data, and developmental and neuropsychological assessments for a large cohort of developing children aged 3 to 20 years. The scientific aim of the project is, by openly sharing these data, to amplify the power and productivity of investigations of healthy and disordered development in children, and to increase understanding of the origins of variation in neurobehavioral phenotypes. For up-to-date information, see `http://ping.chd.ucsd.edu/`.

ADNI Methods

Data used in the preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (`http://adni.loni.usc.edu`). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a 60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimers disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 subjects but ADNI has been followed by ADNI-GO and ADNI-2. To date these three protocols have recruited over 1500 adults, ages 55 to 90, to participate in the research, consisting of cognitively normal older individuals, people with early or late MCI, and people with early AD. The follow up duration of each group is specified in the protocols for ADNI-1, ADNI-2 and ADNI-GO. Subjects originally recruited for ADNI-1 and ADNI-GO had the option to be followed in ADNI-2. For up-to-date information, see `www.adni-info.org`.

Pediatric Imaging, Neurocognition and Genetics (PING) Authors

Connor McCabe[1], Linda Chang[2], Natacha Akshoomoff[3], Erik Newman[1], Thomas Ernst[2], Peter Van Zijl[4], Joshua Kuperman[5], Sarah Murray[6], Cinnamon Bloss[6], Mark Appelbaum[1], Anthony Gamst[1], Wesley Thompson[3], Hauke Bartsch[5].

Alzheimer's Disease Neuroimaging Initiative (ADNI) Authors

Michael Weiner[7], Paul Aisen[1], Ronald Petersen[8], Clifford R. Jack Jr[8], William Jagust[9], John Q. Trojanowki[10], Arthur W. Toga[11], Laurel Beckett[12], Robert C. Green[13], Andrew

J. Saykin[14], John Morris[15], Leslie M. Shaw[10], Zaven Khachaturian[16], Greg Sorensen[17], Maria Carrillo[18], Lew Kuller[19], Marc Raichle[15], Steven Paul[20], Peter Davies[21], Howard Fillit[22], Franz Hefti[23], Davie Holtzman[15], M. Marcel Mesulman[24], William Potter[25], Peter J. Snyder[26], Adam Schwartz[27], Tom Montine[28], Ronald G. Thomas[1], Michael Donohue[1], Sarah Walter[1], Devon Gessert[1], Tamie Sather[1], Gus Jiminez[1], Danielle Harvey[12], Matthew Bernstein[8], Nick Fox[29], Paul Thompson[11], Norbert Schuff[7], Charles DeCarli[12], Bret Borowski[8], Jeff Gunter[8], Matt Senjem[8], Prashanthi Vemuri[8], David Jones[8], Kejal Kantarci[8], Chad Ward[8], Robert A. Koeppe[30], Norm Foster[31], Eric M. Reiman[32], Kewei Chen[32], Chet Mathis[19], Susan Landau[9], Nigel J. Cairns[15], Erin Householder[15], Lisa Taylor-Reinwald[15], Virginia M.Y. Lee[10], Magdalena Korecka[10], Michal Figurski[10], Karen Crawford[11], Scott Neu[11], Tatiana M. Foroud[14], Steven Potkin[33], Li Shen[14], Kelley Faber[14], Sungeun Kim[14], Kwangsik Nho[14], Leon Thal[1], Richard Frank[34], Neil Buckholtz[35], Marilyn Albert[36], John Hsiao[35].

[1]UC San Diego, La Jolla, CA 92093, USA. [2]U Hawaii, Honolulu, HI 96822, USA. [3]Department of Psychiatry, University of California, San Diego, La Jolla, California 92093, USA. [4]Kennedy Krieger Institute, Baltimore, MD 21205, USA. [5]Multimodal Imaging Laboratory, Department of Radiology, University of California San Diego, La Jolla, California 92037, USA. [6]Scripps Translational Science Institute, La Jolla, CA 92037, USA. [7]UC San Francisco, San Francisco, CA 94143, USA. [8]Mayo Clinic, Rochester, MN 55905, USA. [9]UC Berkeley, Berkeley, CA 94720-5800, USA. [10]U Pennsylvania, Philadelphia, PA 19104, USA. [11]USC, University of Southern California, Los Angeles, CA 90033, USA. [12]UC Davis, Davis, CA 95616, USA. [13]Brigham and Women s Hospital/Harvard Medical School, Boston MA 02115, USA. [14]Indiana University, Indianapolis, IN 46202-5143, USA. [15]Washington University St. Louis, St. Louis, MO 63130, USA. [16]Prevent Alzheimers Disease 2020, Rockville, MD 20850, USA. [17]Siemens [18]Alzheimers Association, Chicago, IL 60601, USA. [19]University of Pittsburgh, Pittsburgh, PA 15260, USA. [20]Cornell University, Ithaca, NY 14850, USA. [21]Albert Einstein College of Medicine of Yeshiva University, Bronx, NY 10461, USA. [22]AD Drug Discovery Foundation, New York, NY 10019, USA. [23]Acumen Pharmaceuticals, Livermore, California 94551, USA. [24]Northwestern University, Evanston, IL 60208, USA. [25]National Institute of Mental Health, Bethesda, MD 20892-9663, USA. [26]Brown University, Providence, RI 02912, USA. [27]Eli Lilly, Indianapolis, Indiana 46285, USA. [28]University of Washington, Seattle, WA 98195, USA. [29]University of London, London WC1E 7HU, UK. [30]University of Michigan, Ann Arbor, MI 48109, USA. [31]University of Utah, Salt Lake City, UT 84112, USA. [32]Banner Alzheimers Institute, Phoenix, AZ 85006, USA. [33]UC Irvine, Irvine, CA 92697, USA. [34]General Electric [35]National Institute on Aging/National Institutes of Health, Bethesda, MD 20892, USA. [36]The Johns Hopkins University, Baltimore, MD 21218, USA.

LD Hub

We gratefully acknowledge all the studies and databases that made GWAS summary data available: ADIPOGen (Adiponectin genetics consortium), C4D (Coronary Artery Disease Genetics Consortium), CARDIoGRAM (Coronary ARtery DIsease Genome wide Replication and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics consortium), dbGAP (database of Genotypes and Phenotypes), DIAGRAM (DIAbetes Genetics Replication And Meta-analysis), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), EAGLE (EArly Genetics & Lifecourse Epidemiology Eczema Consortium, excluding 23andMe), EGG (Early Growth Genetics Consortium), GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOS (GEnetic Factors for OSteoporosis Consortium), GIANT (Genetic Investigation of ANthropometric Traits), GIS (Genetics of Iron Status consortium), GLGC (Global Lipids Genetics Consortium), GPC (Genetics of Personality Consortium), GUGC (Global Urate and Gout consortium), HaemGen (haemotological and platelet traits genetics consortium), HRgene (Heart Rate consortium), IIBDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), IMSGC (International Multiple Sclerosis Genetic Consortium), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), MESA (Multi-Ethnic Study of Atherosclerosis), PGC (Psychiatric Genomics Consortium), Project MinE consortium, ReproGen (Reproductive Genetics Consortium), SSGAC (Social Science Genetics Association Consortium) and TAG (Tobacco and Genetics Consortium), TRICL (Transdisciplinary Research in Cancer of the Lung consortium), UK Biobank. We gratefully acknowledge the contributions of Alkes Price (the systemic lupus erythematosus GWAS and primary biliary cirrhosis GWAS) and Johannes Kettunen (lipids metabolites GWAS).

References

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A. and Gee, J. C. (2011) A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, **54**, 2033–2044.

Belden, A. C., Irvin, K., Hajcak, G., Kappenman, E. S., Kelly, D., Karlow, S., Luby, J. L. and Barch, D. M. (2016) Neural correlates of reward processing in depressed and healthy preschool-age children. *Journal of the American Academy of Child & Adolescent Psychiatry*, **55**, 1081–1089.

Finucane, H. K., Reshef, Y. A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.-R., Lareau, C., Shoresh, N. et al. (2018) Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics*, **50**, 621–629.

Gao, F., Chang, D., Biddanda, A., Ma, L., Guo, Y., Zhou, Z. and Keinan, A. (2015) Xwas: a software toolset for genetic data analysis and association studies of the x chromosome. *Journal of Heredity*, **106**, 666–671.

Grieve, S. M., Korgaonkar, M. S., Koslow, S. H., Gordon, E. and Williams, L. M. (2013) Widespread reductions in gray matter volume in depression. *NeuroImage: Clinical*, **3**, 332–339.

Jackson, J., Balota, D. A. and Head, D. (2011) Exploring the relationship between personality and regional brain volume in healthy aging. *Neurobiology of Aging*, **32**, 2162–2171.

Klein, A. and Tourville, J. (2012) 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience*, **6**, 171.

Knutson, B., Momenan, R., Rawlings, R. R., Fong, G. W. and Hommer, D. (2001) Negative association of neuroticism with brain volume ratio in healthy humans. *Biological Psychiatry*, **50**, 685–690.

de Leeuw, C., Mooij, J., Heskes, T. and Posthuma, D. (2015) Magma: Generalized gene-set analysis of gwas data. *PLoS Comput Biol*, **11**, e1004219.

Liu, E. Y., Li, M., Wang, W. and Li, Y. (2013) Mach-admix: genotype imputation for admixed populations. *Genetic Epidemiology*, **37**, 25–37.

Pers, T. H., Karjalainen, J. M., Chan, Y., Westra, H.-J., Wood, A. R., Yang, J., Lui, J. C., Vedantam, S., Gustafsson, S., Esko, T. et al. (2015) Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications*, **6**, 5890.

Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., Kandel, B. M., van Strien, N., Stone, J. R., Gee, J. C. et al. (2014) Large-scale evaluation of ants and freesurfer cortical thickness measurements. *Neuroimage*, **99**, 166–179.

Watanabe, K., Taskesen, E., Van Bochoven, A. and Posthuma, D. (2017) Functional mapping and annotation of genetic associations with fuma. *Nature Communications*, **8**, 1826.

Zhao, B., Ibrahim, J. G., Li, Y., Li, T., Wang, Y., Shan, Y., Zhu, Z., Zhou, F., Zhang, J., Huang, C. et al. (2018) Heritability of regional brain volumes in large-scale neuroimaging and genetic studies. *Cerebral Cortex*, **29**, 2904–2914.