

Polygenic Architecture of Human Neuroanatomical Diversity

Anne Biton, Nicolas Traut, Jean-Baptiste Poline, Benjamin Aribisala, Mark Bastin, Robin Bülow, Simon R. Cox, Ian J Deary, Masaki Fukunaga, Hans J Grabe, et al.

▶ **To cite this version:**

Anne Biton, Nicolas Traut, Jean-Baptiste Poline, Benjamin Aribisala, Mark Bastin, et al.. Polygenic Architecture of Human Neuroanatomical Diversity. Cerebral Cortex, 2020, 30 (4), pp.2307-2320. 10.1093/cercor/bhz241 . pasteur-02915091

HAL Id: pasteur-02915091

https://pasteur.hal.science/pasteur-02915091

Submitted on 13 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

OXFORD UNIVERSITY PRESS | Cerebral Cortex

Polygenic architecture of human neuroanatomical diversity.

Journal:	*Cerebral Cortex*
Manuscript ID	CerCor-2019-00297.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	09-Aug-2019
Complete List of Authors:	Biton, Anne; Institut Pasteur, Unité de Génétique Humaine et Fonctions Cognitives, Département de Neuroscience; Institut Pasteur, Hub Bioinformatique et Biostatistique – C3BI, USR 3756 IP CNRS Traut, Nicolas; Institut Pasteur, Neuroscience Poline, Jean-Baptiste; McGill University, Faculty of Medicine Department of Neurology and Neurosurgery Montreal Neurological Institute and Hospital, McConnell Brain Imaging Centre Aribisala, Benjamin; University of Edinburgh, Centre for Cognitive Ageing and Cognitive Epidemiology; University of Edinburgh, Centre for Clinical Brain Sciences; Lagos State University, Department of Computer Science Bastin, Mark; University of Edinburgh, Medical physics; University of Edinburgh, Centre for Clinical and Brain Sciences; University of Edinburgh, Brain Research Imaging Centre Bülow, Robin; University Medicine Greifswald, Institute of Diagnostic Radiology and Neuroradiology Cox, Simon; University of Edinburgh, Centre for Cognitive Ageing and Cognitive Epidemiology; University of Edinburgh, Department of Psychology Deary, Ian; University of Edinburgh, Centre for Cognitive Ageing and Cognitive Epidemiology; University of Edinburgh, Department of Psychology Fukunaga, Masaki; National Institute for Physiological Sciences, Division of Cerebral Integration Grabe, Hans-Jörgen; University Medicine Greifswald, Department of Psychiatry and Psychotherapy; German Centre of Neurodegenerative Diseases (DZNE) site Greifswald/Rostock, German Centre of Neurodegenerative Diseases (DZNE) site Greifswald/Rostock Hagenaars, Saskia; University of Edinburgh, Centre for Cognitive Ageing and Cognitive Epidemiology; University of Edinburgh, Department of Psychology Hashimoto, Ryota; National Center of Neurology and Psychiatry, Department of Pathology of Mental Diseases Kikuchi, Masataka; Osaka University Graduate School of Medicine, Department of Genome Informatics Maniega, Susana; University of Edinburgh, Centre for Cognitive Ageing and Cognitive Epidemiology; University of Edinburgh, Centre for Clinical Brain Sciences; University of Edinburgh, Brain Research Imaging Centre Nauck, Matthias; University Medicine Greifswald, Institute of Clinical Chemistry and Laboratory Medicine; University Medicine Greifswald, DZHK (German Centre for Cardiovascular Research), Partner Site

	Greifswald Royle, Natalie; University of Edinburgh, Centre for Cognitive Ageing and Cognitive Epidemiology; University of Edinburgh, Centre for Clinical Brain Sciences; University of Edinburgh, Brain Research Imaging Centre Teumer, Alexander; University Medicine Greifswald, Institute for Community Medicine Valdes Hernandez, Maria; University of Edinburgh, Centre for Cognitive Ageing and Cognitive Epidemiology; University of Edinburgh, Centre for Clinical Brain Sciences; University of Edinburgh, Brain Research Imaging Centre Völker, Uwe; University Greifswald, Interfaculty Institute of Genetics and Functional Genomics; University Medicine Greifswald, DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald Wardlaw, Joanna; University of Edinburgh, Centre for Clinical Brain Sciences Wittfeld, Katharina; German Center for Neurodegenerative Diseases (DZNE), Site Rostock/Greifswald Yamamori, Hidenaga; Osaka University Graduate School of Medicine, Department of Psychiatry Bourgeron, Thomas; Institut Pasteur, Laboratory of Human Genetics and Cognitive Functions Toro, Roberto; Institut Pasteur, Neuroscience; Center for Research and Interdisciplinarity (CRI),
Keywords:	Neuroimaging, Genetics, Polygenic architecture, Subcortical structures, Heritability

SCHOLARONE™
Manuscripts

Polygenic architecture of human neuroanatomical diversity.

Anne Biton[1,2*], Nicolas Traut[1*], Jean-Baptiste Poline[3], Benjamin S. Aribisala[4,5,6], Mark E. Bastin[4,5,7], Robin Bülow[8], Simon R. Cox[4,9], Ian J. Deary[4,9], Masaki Fukunaga[10,11], Hans J. Grabe[12,13], Saskia Hagenaars[4,9,14], Ryota Hashimoto[15], Masataka Kikuchi[16], Susana Muñoz Maniega[4,5,7], Matthias Nauck[17,18], Natalie A. Royle[4,5,7], Alexander Teumer[19], Maria Valdes Hernandez[4,5,7], Uwe Völker[18,20], Joanna M. Wardlaw[4,5,7], Katharina Wittfeld[12,13], Hidenaga Yamamori[21], Alzheimer's Disease Neuroimaging Initiative†, Thomas Bourgeron[1★], Roberto Toro[1,22★]

[1] Human Genetics and Cognitive Functions Unit, Institut Pasteur, UMR 3571, CNRS, Université Paris Diderot, Sorbonne Paris Cité, Paris (75015) France, [2] Hub de Bioinformatique et Biostatistique – Département Biologie Computationnelle, Institut Pasteur, USR 3756 CNRS, Paris, France, [3] McGill University, Montreal, Quebec, Canada, [4] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK, [5] Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK, [6] Department of Computer Science, Lagos State University, Lagos, Nigeria, [7] Brain Research Imaging Centre, University of Edinburgh, Edinburgh, UK, [8] Institute of Diagnostic Radiology and Neuroradiology, University Medicine Greifswald, Greifswald, Germany, [9] Department of Psychology, University of Edinburgh, Edinburgh, UK, [10] Division of Cerebral Integration, National Institute for Physiological Sciences, Okazaki, Japan, [11] Department of Physiological Sciences, School of Life Sciences, The Graduate University for Advanced Studies (SOKENDAI), Okazaki, Japan, [12] Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany, [13] German Centre of Neurodegenerative Diseases (DZNE) site Greifswald/Rostock, Germany, [14] Social Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, Denmark Hill, London, UK, [15] Department of Pathology of Mental Diseases, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan, [16] Department of Genome Informatics, Graduate School of Medicine, Osaka University, Osaka, Japan, [17] Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany, [18] DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, University Medicine, Greifswald, Germany, [19] Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany, [20] Interfaculty Institute of Genetics and Functional Genomics, University Greifswald, Greifswald, Germany, [21] Department of Psychiatry, Graduate School of Medicine, Osaka University, Osaka, Japan, [22] Center for Research and Interdisciplinarity (CRI), Université Paris Descartes, Paris, France.

*Equal contribution, ★Equal contribution, Correspondence: rto@pasteur.fr
†Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:
http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Corresponding author:
Roberto Toro
Institut Pasteur
25-28 Rue du Dr Roux
75015 Paris
France
Telephone: +33 (0)1 44 38 93 48
E-mail: rto@pasteur.fr

Running title: Polygenic architecture of human neuroanatomical diversity.

Abstract

We analysed the genomic architecture of neuroanatomical diversity using magnetic resonance imaging and single nucleotide polymorphism (SNP) data from >26,000 individuals from the UK Biobank project and 5 other projects that had previously participated in the ENIGMA consortium. Our results confirm the polygenic architecture of neuroanatomical diversity, with SNPs capturing from 40% to 54% of regional brain volume variance. Chromosomal length correlated with the amount of phenotypic variance captured, r~0.64 on average, suggesting that at a global scale causal variants are homogeneously distributed across the genome. At a local scale, SNPs within genes (~51%) captured ~1.5 times more genetic variance than the rest; and SNPs with low minor allele frequency (MAF) captured less variance than the rest: the 40% of SNPs with MAF<5% captured <1/4th of the genetic variance. We also observed extensive pleiotropy across regions, with an average genetic correlation of r_G~0.45. Genetic correlations were similar to phenotypic and environmental correlations, however, genetic correlations were often larger than phenotypic correlations for the left/right volumes of the same region. The heritability of differences in left/right volumes was generally not statistically significant, suggesting an important influence of environmental causes in the variability of brain asymmetry. Our code is available at https://github.com/neuroanatomy/genomic-architecture.

Keywords:

Neuroimaging, Genetics, Polygenic architecture, Subcortical structures, Heritability

Introduction

Imaging genetics studies have traditionally emphasised the role of candidate genes and specific loci. The underlying hypothesis is that phenotypic diversity results from the action of a reduced number of genes, close to the Mendelian paradigm where a phenotype is determined by a single locus. However, the amount of neuroanatomical variance captured by candidate genes or genome-wide significant loci has remained extremely small. Genome-wide association studies (GWAS) aiming at identifying associated variants through international collaborative efforts such as ENIGMA and CHARGE have revealed only few statistically significant associated loci (Stein et al. 2012, Hibar et al. 2015, Satizabal et al. 2017), capturing <1% of phenotypic variance. This is in contrast with the large heritability of neuroanatomical diversity estimated by twin and extended pedigree studies (Blokland et al. 2012, Wen et al. 2016) which show that about 80% of neuroanatomical variability is captured by genetic factors.

The genetic architecture of neuroanatomical diversity could result, alternatively, from the aggregated effect of thousands of different loci spread across the genome, a paradigm closer to the infinitesimal model proposed by Fisher (1918). Thanks to the development of methods to estimate heritability from whole-genome genotyping data (reviewed by Yang et al. 2017), several groups have estimated that genotyped variants, taken together, were able to capture up to 55% of the phenotypic variance, retrieving about two-thirds of the heritability estimated by family studies (Toro et al. 2015, Ge et al. 2015, 2016, Zhao et al. 2018, Elliott et al. 2018). These results are compatible with the hypothesis of a highly polygenic architecture, where phenotypes are influenced by large numbers of loci of effect sizes too small to reach genome-wide significance (usually $p<5\times10^{-8}$). While information about the function of a few candidate genes can have a strong explanatory power in the case of a few causal loci, a polygenic approach calls for alternative, system-wide, sources of biological insight.

Our aim was to better understand the role of polygenic causes in the determination of neuroanatomical diversity. In a previous work, we used a sample of N = 2,011 subjects with whole-genome genotyping

data from the IMAGEN consortium combined with structural magnetic resonance imaging (MRI) data analysed according to the pipelines of the ENIGMA consortium, and showed that genomic complex trait analysis (GCTA) was able to capture a substantial proportion of the variability in regional brain volume – up to 55% (Toro et al. 2015). Due to sample size limitations, however, standard errors were high (about ± 25%) and statistical power for post-hoc analyses was limited. The advent of collaborative efforts such as the ENIGMA consortium, and large scale projects such as UK Biobank, in particular, has allowed researchers to greatly increase the number of subjects used for heritability estimation (see for example, Elliott et al. 2018, Zhao et al. 2018). Here we replicate and follow up on our original results using a sample ten times larger, N = 26,818, which combines data from the UK Biobank project as well as 5 other projects that had previously participated in the ENIGMA consortium (IMAGEN, SHIP, TREND, Lothian, ADNI).

We used the same regional brain volumes estimates than those used in the ENIGMA consortium studies by Stein et al. (2012) and Hibar et al. (2015): several subcortical structures, total brain volume and intracranial volume. In addition, we also studied height and intelligence scores. Brain volume correlates with height and intelligence scores (Taki et al. 2012), which are both known to be heritable (Yang et al. 2010, Plomin and Deary 2015). We aimed thus at determining to which extent the heritability of regional brain volumes was given by their relationship with height (i.e., affected by the same genetic factors that determine body size), or if different genetic factors affected them specifically. Similarly for intelligence scores, we aimed at better understanding its relationship with brain volume.

For all phenotypes, we estimated to what extent genome-wide SNPs were able to capture the inter-individual variability in regional brain volumes, that is, we estimated the proportion of phenotypic variance captured by SNPs across the genome (also called SNP heritability). Additionally, we used GWAS data to compute genome-wide polygenic scores, which provide a phenotypic prediction at the individual level. The analyses of the influence of the complete genome on our phenotypes were complemented with analyses on a series of genomic partitions: genic versus non-genic; preferential

expression in the central nervous system or by cell type; low, medium or high minor allele frequency. This type of analysis can reveal whether specific genomic regions are enriched in the amount of variance they capture. Finally, we looked at the pleiotropy across phenotypes. For this, we computed genetic correlations and phenotypic correlations for all pairs of phenotypes. For brain regions, we also compared the genetic and phenotypic correlations between the left and right parts of the same structure as a means to estimate the role of genetics and environment in brain asymmetry.

Our results confirm the observation that polygenic factors play an important role in the determination of neuroanatomical variability, and show different ways in which biological information can be obtained to better understand polygenic effects.

Material and Methods

Data sharing

We obtained whole-genome genotyping from N = 26,818 subjects from 6 different projects: UK Biobank, IMAGEN, ADNI, Lothian Birth Cohort 1936, SHIP and TREND. Extensive efforts have been made to homogenise the neuroanatomical measurements across sites, which were described in the ENIGMA 1 and 2 reports (Stein et al. 2012, Hibar et al. 2015), and the UK Biobank neuroimaging analysis group (Alfaro-Almagro et al. 2018). The UK Biobank project (https://imaging.ukbiobank.ac.uk) is a large, long-term biobank study in the United Kingdom aiming at investigating the contributions of genetic predisposition and environmental exposure to the development of disease. The study is following about 500,000 volunteers enrolled at ages from 40 to 69 years old, 54% females. IMAGEN (https://imagen-europe.com) is a project to identify and characterise specific genetically-influenced alterations in reinforcer sensitivity and executive control which manifest in adolescence and carry the risk for overt psychopathology later in life. It includes general population 13 to 17 year old adolescents (49%

of females) from Germany, France, Ireland, and the United Kingdom. ADNI, the Alzheimer's Disease Neuroimaging Initiative (http://adni.loni.usc.edu), is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of Alzheimer's disease in the United States of America. The dataset combines data from the initial five-year study (ADNI-1), and the follow-ups ADNI-GO, ADNI-2, and ADNI-3. It includes subjects 54 to 90 years old, 42% female. The Lothian Birth Cohort 1936 (https://www.lothianbirthcohort.ed.ac.uk) is a follow-up of the Scottish Mental Surveys of 1947, which tested the intelligence of almost every child born in 1936 and attending school in Scotland in the month of June 1947. It includes subjects 71 to 73 years old, 47% female. The SHIP and TREND cohorts contain data from the Study of Health in Pomerania (SHIP, http://www2.medizin.uni-greifswald.de/cm/fv/ship.html), a population-based epidemiological study consisting of two independent cohorts SHIP and SHIP-TREND. These projects investigate common risk factors, subclinical disorders and diseases in a population of northeast Germany. The dataset included data from subjects 21 to 90 years old, 44% female for TREND, 48% female for SHIP. All data sharing was approved by our local ethical board as well as by those of the participating projects wherever required. The list of projects and their respective number of subjects is described in Table 1.

Regional brain volumes

The measurements of regional brain volume coming from projects that had previously participated in ENIGMA (IMAGEN, Lothian Birth Cohort 1936 , TREND, SHIP and ADNI) were the same that had been used in Stein et al. (2012) and Hibar et al. (2015). For the UK Biobank subjects, the estimation of the volumes was performed using FreeSurfer 6.0 (https://surfer.nmr.mgh.harvard.edu). For comparison, we also included the estimates obtained using FSL FIRST (https://fsl.fmrib.ox.ac.uk/fsl) that were made available by UK Biobank (the processing pipeline is described in https://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf). In addition to the subjects excluded in the quality control made by the UK Biobank, we excluded 52 additional subjects showing an extreme relationship

Project	Total sample size	Percentage of females	Mean age (standard deviation)	Sample size included	Total number of variants	Number of variants included
IMAGEN	2,011	49%	14.6 (0.4)	1,736	573,299	267,151
Lothian Birth Cohort 1936	1,005	47%	72.7 (0.7)	544	529,015	256,417
TREND	858	44%	50.0 (13.5)	813	2,389,858	597,902
SHIP	963	48%	56.5 (12.6)	941	863,230	271,635
ADNI	1,189	42%	74.2 (7.1)	986	331,088	227,005
UK Biobank*	20,792	54%	62.6 (7.5)	19,270	734,447	490,061

Table 1. Sample sizes and number of variants per project. *The UK Biobank dataset was split in two parts: N = 14,144 subjects were used for heritability analyses (13,086 included), and N = 6,678 subjects were used for validation of genome-wide polygenic scores (6,184 included).

between total brain volume and intracranial volume. For this, we used a kernel density estimator to fit a probability density function to the intracranial versus brain volume data, and tagged as outliers all subjects with a local density inferior to 1% of the maximum density for UK Biobank, and 2% for other datasets.

The regions included in our analyses were: nucleus accumbens (labelled as Acc), amygdala (Amy), putamen (Pu), pallidum (Pa), caudate nucleus (Ca), hippocampus (Hip) and thalamus (Th); along with brain volume (BV) and intracranial volume (ICV). In addition to these regions, we investigated height and intelligence scores (IS), available from the UK Biobank and IMAGEN projects. It is important to note that the fluid intelligence score in UK Biobank (a 2 minute test aiming at evaluating the capacity to solve

problems that require logic and reasoning ability, independent of acquired knowledge; see https://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=100027), is not the same as the intelligence score used by the IMAGEN project, which was obtained using the WISC test.

Genotype filtering

All genetic analyses were performed for each project independently. Genotyping data was converted to the hg19 reference wherever required using UCSC LiftOver (http://genome.ucsc.edu/cgi-bin/hgLiftOver). We used genotyped autosomal SNPs (single nucleotide polymorphisms) which passed UK Biobank quality control for all batches (http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf). Additionally, we removed SNPs in 24 regions with long range linkage disequilibrium (LD, see Price et al. 2008). SNPs were then filtered to exclude those with minor allele frequency (MAF) <0.1%, missing rate >1%, or Hardy-Weinberg disequilibrium with a $p<10^{-6}$. Individuals were removed when >10% of their SNPs were missing. We finally pruned SNPs which were in linkage disequilibrium with a variance inflation factor >10, which corresponds to a multiple R^2 for the regression over linked SNPs <0.9. The filtering was made using PLINK v1.90b3.46 (Purcell et al. 2007).

Genetic relationship matrices (GRM)

GRMs were computed based on autosomal chromosomes using GCTA v1.91.3 (Yang, Lee, et al. 2011). We included only one of each pair of subjects with an estimated relatedness >0.025 (approximately corresponding to cousins two to three times removed). GRMs were computed per chromosome and then merged for the whole genome.

Population structure

Genetic variance estimates based on genomic estimates of relatedness are sensitive to cryptic relatedness and population structure. These factors can influence the phenotypic similarity beyond the estimated degree of genetic relatedness (Yang, Manolio, et al. 2011, Browning and Browning 2011). In addition to the exclusion of subjects with a degree of genetic relatedness greater than 0.025, we used the first 10 principal components of the GRM as covariates in our statistical analyses.

Genetic variance

We estimated the amount of phenotypic variance captured by SNPs using a linear mixed model with age, sex, imaging centre, and the first 10 principal components of the GRM as fixed effect covariates, and a random effect with a covariance matrix corresponding to the GRM (GCTA GREML method, Yang et al. 2011). We estimated SNP heritability as the ratio of the genetic variance to the phenotypic variance, with genetic variance being the variance of the random component, and the phenotypic variance being the sum of random component and residual component with fixed effects removed. We used GCTA v1.91.3 (Yang, Lee, et al. 2011) for those computations and did not constrain genetic variance estimates to lie in the range 0% to 100%, in order to obtain unbiased estimates (option *--reml-no-constrain*).

Genetic correlation

Genetic correlation was estimated using GCTA REML bivariate analysis (Lee, Yang, et al. 2012) in constrained mode (option *--reml-bivar*). Both phenotypic and genetic correlation were adjusted for age, sex, imaging centre, and the first 10 principal components of the GRM. We compared genetic and phenotypic correlations using the delta method to estimate standard errors (R package msm https://cran.r-project.org/web/packages/msm/). We report estimates with their standard errors (s.e.).

Genetic variance partitioning

In its simplest form, GCTA allows to estimate the amount of variance captured by the matrix of genetic relationships, assuming that each SNP captures the same amount of variance. Through genomic partitions we can create different genetic relationship matrices based on non-overlapping regions of the genome. The SNPs on each of these partitions can capture then a different amount of variance (although, as before, SNPs within a given partition are supposed to capture all the same amount of variance). We grouped SNPs in the following partitions:

1. Partition based on genic status. Using 66,632 gene boundaries from the UCSC Genome Browser hg19 assembly, we made a first set with all SNPs within these boundaries, two further sets that included also SNPs 0 to 20 kbp and 20kbp to 50 kbp upstream and downstream of each gene, and a last set including the SNPs not located in regions less than 50 kbp upstream or downstream of genes. Both exonic and intronic regions were included in the genic regions. These partitions do not correspond exactly to those used by Toro et al. (2015) which were: one with strict genic/non-genic boundaries (0 kbp), another with genic ± 20 kbp versus the rest, and finally genic ± 50 kbp versus the rest.
2. Partition based on preferential central nervous system (CNS) expression (Raychaudhuri et al. 2010, Lee, DeCandia, et al. 2012) using ± 50 kbp as gene boundaries, and based on markers of brain cell types as defined by two recent scRNA-Seq studies (Skene et al. 2018, 2018, Li et al. 2018) using ± 20 kbp as gene boundaries.
3. Partition based on allele frequency. A partition based on MAF with 4 groups: from 0.1 to 5%, from 5 to 20%, from 20 to 35% and from 35 to 50%. In Toro et al (2015), only the last 3 partitions were included, covering the range from 5 to 50% of MAF.

Genetic variance partition enrichment

Once the variance captured by each partition was computed, we were able to estimate the significance of the difference in variance captured by an individual partition against a model where each SNP captures exactly the same amount of variance (a null hypothesis of no enrichment). In this latter case, the amount of variance captured by each partition should be directly proportional to the number of SNPs it contains.

We tested whether any of the partitions captured more variance than what could be expected given its number of SNPs. A Z-score was computed by comparing the SNP-set genetic estimated variance V_{G_i} of partition i to the SNP-set genetic variance $f_i \cdot V_{G_{tot}}$ expected under no enrichment:

$$Z = \frac{V_{G_i} - f_i \cdot V_{G_{tot}}}{\sqrt{Var(V_{G_i} - f_i \cdot V_{tot})}} \ , \qquad \text{(Eq. 1)}$$

where f_i is the fraction of the SNPs included in partition i. We estimated the variance of the observed enrichment as:

$$Var(V_{G_i} - f_i \cdot V_{G_{tot}}) = Var(V_{G_i}) + f_i^2 \cdot Var(V_{G_{tot}}) - 2 \cdot f_i \cdot Cov(V_{G_i}, V_{G_{tot}}) . \qquad \text{(Eq. 2)}$$

Here, V_{G_i} represents the genetic variance of partition i and $V_{G_{tot}} = \sum_i V_{G_i}$. We preferred the analytic estimation over the estimation by permutations, because the permutation methods presented some limitations due to the difficulty of preserving the SNP structure of the partitions and due to the computational resources needed for the permutations (Fig. S1, S2).

Meta-analytical combination of the estimates of each project

The independent estimates obtained from each of the projects were combined into a single one using an inverse variance weighting method. We validated the meta-analytical approach by comparing the distribution of genetic variance estimates from simulated heritable phenotypes with their theoretical normal distribution in sub-samples of ADNI and UK Biobank datasets.

Accounting for measurement error

The estimation of regional brain volumes, height and intelligence scores is submitted to measurement errors. In particular, small, poorly defined brain regions are more difficult to measure accurately than larger ones. Measurement errors enter as environmental variance in the decomposition of phenotypic variance into genetic and environmental variance, and hence decrease the heritability estimates (V_G/V_P). Heritability estimates unbiased for measurement errors can be obtained with repeated measurements (Ponzi et al. 2018), however, the datasets we used to measure brain volumes provide only one MRI per subject. We used datasets from CoRR, the Consortium for Reliability and Reproducibility (Zuo et al. 2014), which includes multiple MRIs per individual to estimate intraclass correlation coefficients (ICC) (Fisher 1970) for each regional volume. We segmented the volumes as we did in the UK Biobank dataset, using FreeSurfer 6.0. ICCs were computed on the volumes of 836 subjects with two MRI sessions after covarying for age, sex and scanning site. For intelligence, we estimated ICC from 1,301 subjects in UK Biobank who took the fluid intelligence test 3 times. The ICC for height was considered to be equal to 1 (no measurement error). Adjusted phenotypic variance and covariance estimates were obtained by multiplying the raw estimates by their corresponding ICC (see Supplemental Materials).

Genome-wide polygenic scores

We used the SNP effects estimated in the association analysis of 13,086 UK Biobank subjects to estimate the phenotypes of 6,184 additional unrelated subjects with MRI data from the latest release of the UK Biobank project. The scores were estimated from the filtered SNPs (not LD pruned). SNPs under various association p-value thresholds were selected and the ones in LD with a more significantly associated SNP were clumped. The p-value threshold that produced the best fit with the target dataset was selected. We used the software PRSice associated with PLINK for the computation of genome-wide polygenic scores (Euesden, Lewis, and O'Reilly 2015). Each phenotype was regressed on age, sex, scanning centre, and the 10 first principal components of the GRM. The analyses were performed on the residuals of this linear regression. We then estimated the proportion of variance captured by genome-wide polygenic scores in each phenotype using the coefficient of determination R^2. For comparison, we also computed the predicted height of the 6,184 subjects using SNP effects from an independent group of 277,756 unrelated (with a pairwise estimated relatedness <0.025 in the GRM) UK Biobank subjects.

Results

Genome-wide variants capture a large proportion of the diversity of regional brain volumes, height and intelligence scores

The heritability estimates for intracranial volume, total brain volume, as well as the volume of subcortical structures were substantial and statistically significant (Fig. 1, Table S1.1). Estimates for intracranial volume and total brain volume were large: V_G/V_P(ICV) = 53 ± 4.5% (all our variance estimates are reported as estimation ± s.e.), V_G/V_P (BV) = 52 ± 4.5%. Similarly, the genetic variance estimates for subcortical structures were all above 40%: accumbens (V_G/V_P = 40 ± 4.5%), amygdala (V_G/V_P = 45 ±

4.5%), putamen (V_G/V_P = 48 ± 4.5%), palladium (V_G/V_P = 40 ± 4.5%), caudate (V_G/V_P = 52 ± 4.5%), thalamus (V_G/V_P = 54 ± 4.5%), hippocampus (V_G/V_P = 53 ± 4.5%). All estimates were highly statistically significant with $p<10^{-11}$ in all cases (log-likelihood ratio statistics from 49 to 242 in UK Biobank alone) (Fig. 1, Table S1.2). The V_G/V_P estimate for caudate nucleus (V_G/V_P = 52 ± 4.4%) was statistically significant and very different from what was observed in the IMAGEN cohort (V_G/V_P = 4 ± 24%) which, as mentioned in Toro et al. (2015), may reflect an age bias specific to IMAGEN (where individuals were on average 14 years old).

The heritability estimate (V_G/V_P) for height (which combines only UK Biobank and IMAGEN) was large: V_G/V_P = 71 ± 3.1% and close to those obtained from twin studies (Polderman et al. 2015) and from previous SNP heritability estimates (V_G/V_P = 68.5 ± 0.4% in Ge et al. 2017). The fluid intelligence score in UK Biobank had the lowest heritability among all 11 studied phenotypes, V_G/V_P = 35 ± 4.8%. This estimate is similar to the one obtained by Davies et al. (2011) on 30,801 UK Biobank subjects (V_G/V_P = 31% ± 1.8%). The verbal and performance intelligence scores (VIQ and PIQ) in IMAGEN, although not based on the same test as fluid intelligence (FI) in UK Biobank, aim at capturing a similar phenotype. The estimate for fluid intelligence in UK Biobank seem smaller than those for verbal and performance intelligence quotients in IMAGEN (~56 ± 26%), however, they were not statistically significantly different.

Our estimates were obtained using a meta-analytic approach to combine the results of the different projects using inverse variance weighting. In all cases, the meta-analytic estimates closely corresponded with the values obtained for the UK Biobank project (Table S1.2), because of its large sample size which accounted for ~94% of the weighted estimates. Standard errors agreed well with the theoretical values

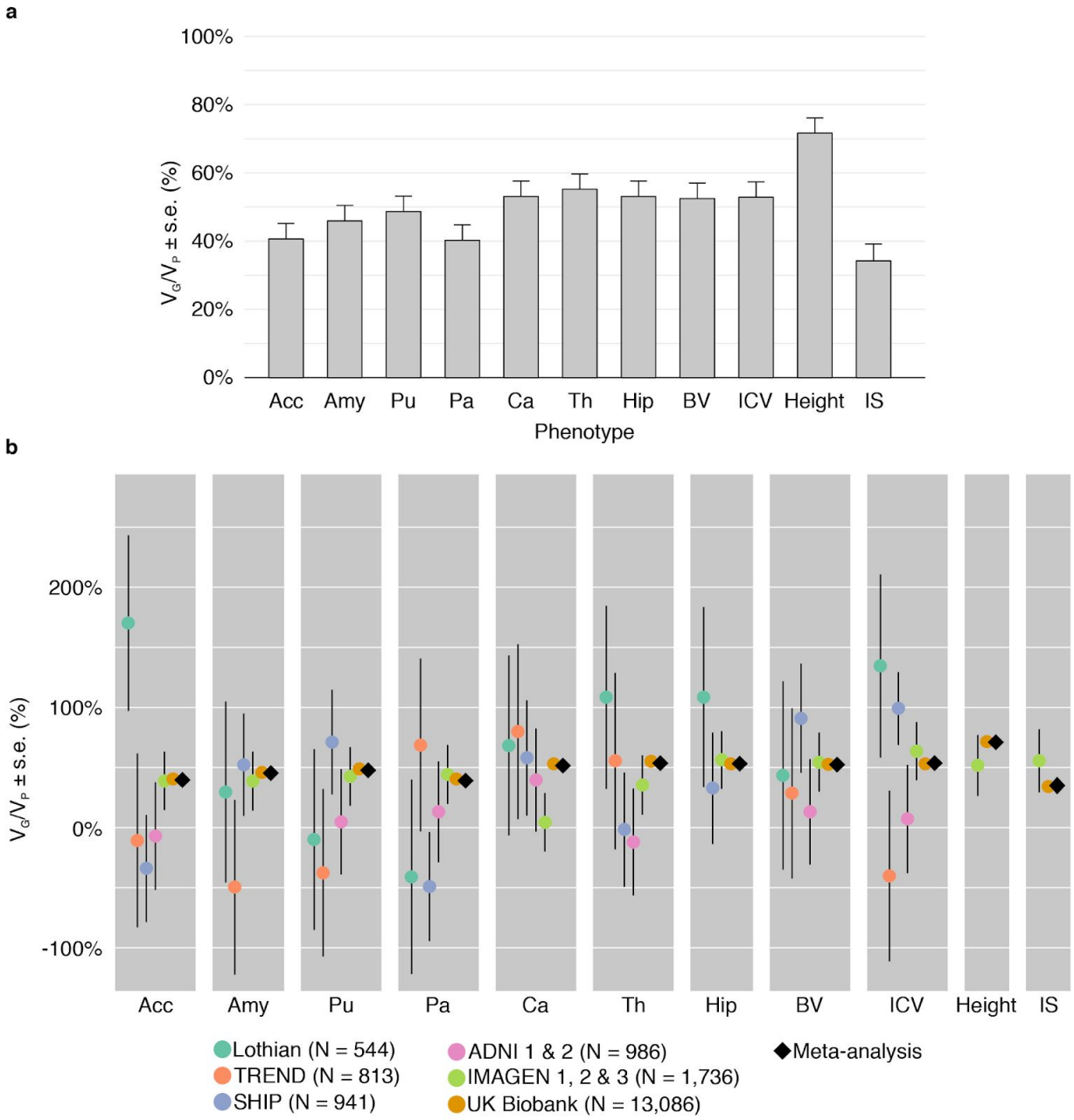


Figure 1. a. Proportion of variance captured by common genotyped variants (V_G/V_P) for brain regions, height and intelligence score. Meta-analytic estimates were obtained using inverse variance weighting of the estimates of the different projects studied. **b.** V_G/V_P estimates for each project. The estimates were obtained using GCTA, without constraining the results to lie in the 0% to 100% range. The diamond shows the meta-analytic estimation. Age, sex, center, and the first 10 principal components were included as covariates. The error bars show the standard errors of the V_G/V_P estimates.

proposed by Visscher et al. (2014), and implemented in the GREML statistical power calculator (http://cnsgenomics.com/shiny/gctaPower): they were from 1.01 to 1.05 times higher than the theoretical values when taking into account the variance of the genetic relationships, except for ADNI where they were 1.21 times higher). As expected, the standard errors decreased with sample size, from about 76% for ~550 individuals, 43% for ~1,000 individuals, 24% for ~1,750 individuals (IMAGEN), down to 4% for UK Biobank with more than 13,000 individuals. Our result from simulated phenotypes showed that the estimates of V_G/V_P obtained with sub-samples of N = 800, N = 400, N = 200 and even N = 100 subjects from the UK Biobank project were unbiased relative to the expected normal distribution. The simulations based on the ADNI project, however, showed a significant bias towards positive values when the sub-samples included N = 100 to N = 400 subjects, probably due to the heterogeneity of the population (Figures S12-13 and Table S6-7).

Similarly to what we had observed previously (Toro et al. 2015), the genetic variance estimates were not significantly affected by population structure: the non-inclusion of the 10 first PCs did not impact the estimates of variance, which changed on average by less than 4% (Fig. S3).

Finally, our heritability estimates for the UK Biobank and ADNI projects were comparable to those of previous studies of these projects by Elliott et al. (2018) and Zhao et al. (2018) (see Fig. S6 and Supplemental Materials).

Genetic partitions show significant enrichment of the proportion of variance captured by specific gene sets

Partition per chromosome (autosomes): We observed a strong correlation between V_G/V_P estimates and chromosome size, which was significant for all phenotypes except the pallidum (Fig. 2). The

correlation coefficients ranged from 0.30 ± 0.21 to 0.80 ± 0.13, r = 0.63 on average, capturing 41% of the variance (estimated as R^2).

Partition between genic and non-genic regions: The genic SNP set (± 0 kbp) contained 51% of all genotyped SNPs and captured on average 69% of the variance attributable to SNPs of most of the studied phenotypes, significantly more than what we would expect from its length (FDR<5%). The only exceptions were fluid intelligence and nucleus accumbens, for which it explained respectively 59 ± 9% and 55 ± 6% of the total genetic variance (Fig. 3, Table S4). Height was the only phenotype for which we observed an enrichment of V_G/V_P captured by one of the non-genic SNP sets: the set of SNPs within 0 ± 20 kbp of the genic set. This set contained 15% of all SNPs but explained 27% of the variance of the height phenotype attributable to SNPs (FDR corrected p<0.01). In total, the variants located between 0 and 50 kbp away from genes captured 36 ± 4% of the genetic variance of height (FDR corrected p<0.05).

Partition by involvement in preferential CNS expression and in brain cell types: No statistically significant enrichment was observed for CNS expression nor brain cell type partitions.

Partition by allele frequency: SNPs within the low MAF partition (MAF<5%) captured less genetic variance per SNP than those with medium and high MAF (the 3 partitions with MAF>5%) (Fig. 3), as previously described by Speed et al. (2017). Our four MAF-based partitions included the following average proportions of total SNPs: (1) MAF from 0.1% to 5%: comprising 40% of SNPs, (2) 5% to 20%: with 30% of SNPs, (3) 20% to 35%: with 12% of SNPs, and (4) 35% to 50%: with 9% of SNPs. Although the partition of SNPs with low MAF contained 40% of the SNPs, it captured on average only about 16% of the total genetic variance. This is less than expected in the GCTA model where each SNP captures the same amount of phenotypic variance, but slightly more than expected in a neutral

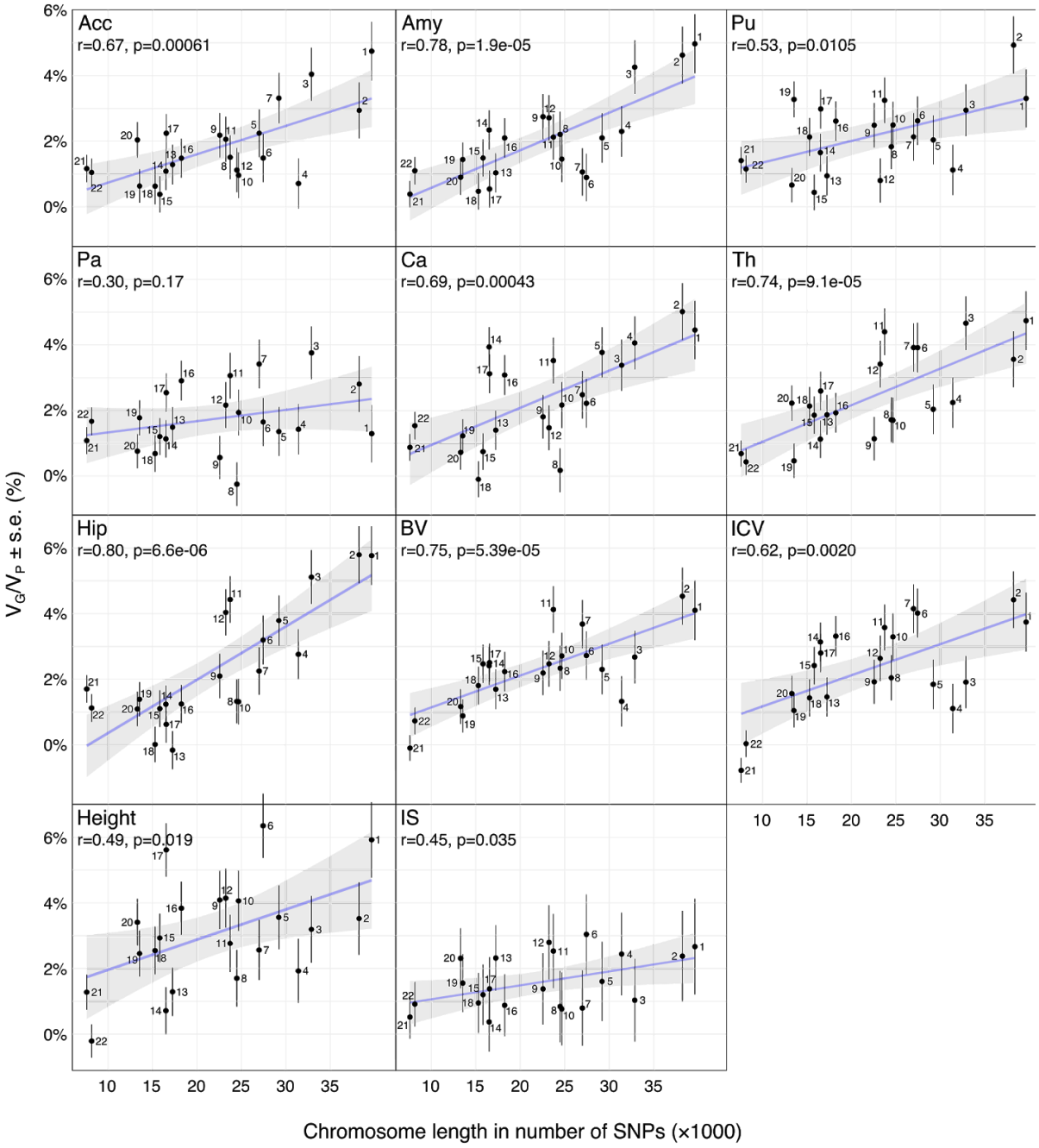


Figure 2. Scatter plots of the number of SNPs per chromosome versus V_G/V_P estimates computed for each chromosome. V_G/V_P estimates were obtained by partitioning SNPs across chromosomes and computed using the GCTA REML unconstrained method for total subcortical volumes. Age, sex, center, and the top 10 principal components were included as covariates. The error bars show the standard errors of the V_G/V_P estimates.

theory of evolution, where the captured variance is proportional to the size of the MAF bin (Yang et al. 2017, Visscher et al. 2012).

Genetic factors influencing volume are shared among brain regions

We observed extensive pleiotropy across brain regions, with an average genetic correlation of r_G~0.45 (Fig. 4). Genetic correlations, which represent the correlation between genetic effects of two phenotypes, were computed for each pair of brain regions, ICV, Height and IS, using data from the UK Biobank project. Figure 4 shows the genetic correlation matrix together with the phenotypic correlation matrix (see also Table S5).

Genetic correlations were in general similar to phenotypic and environmental correlations, however, genetic correlations were often larger than phenotypic correlations for the left/right volumes of the same region. The concordance between phenotypic and genetic correlation was high (R^2 = 0.80, Fig 5c), consistent with the report by Sodini et al. (2018) for other traits, and the correlation matrices were similar (Fig. S7). Phenotypic correlations can be decomposed as a sum of genetic and environmental correlations. The concordance between genetic and environmental correlations was also strong (R^2 = 0.47, Fig. S8).

Environmental factors are important in shaping brain asymmetry

For all brain regions, with the exception of the nucleus accumbens, the differences between their left and right volumes appeared to be of environmental origin. When considering total volumes (left plus right), the differences between genetic and phenotypic correlations were not different from zero (Z-test FDR > 50%, Table S5). However, the situation was different when considering the regional volume asymmetries (left minus right). In that case, the genetic correlations of the asymmetries were not statistically different from

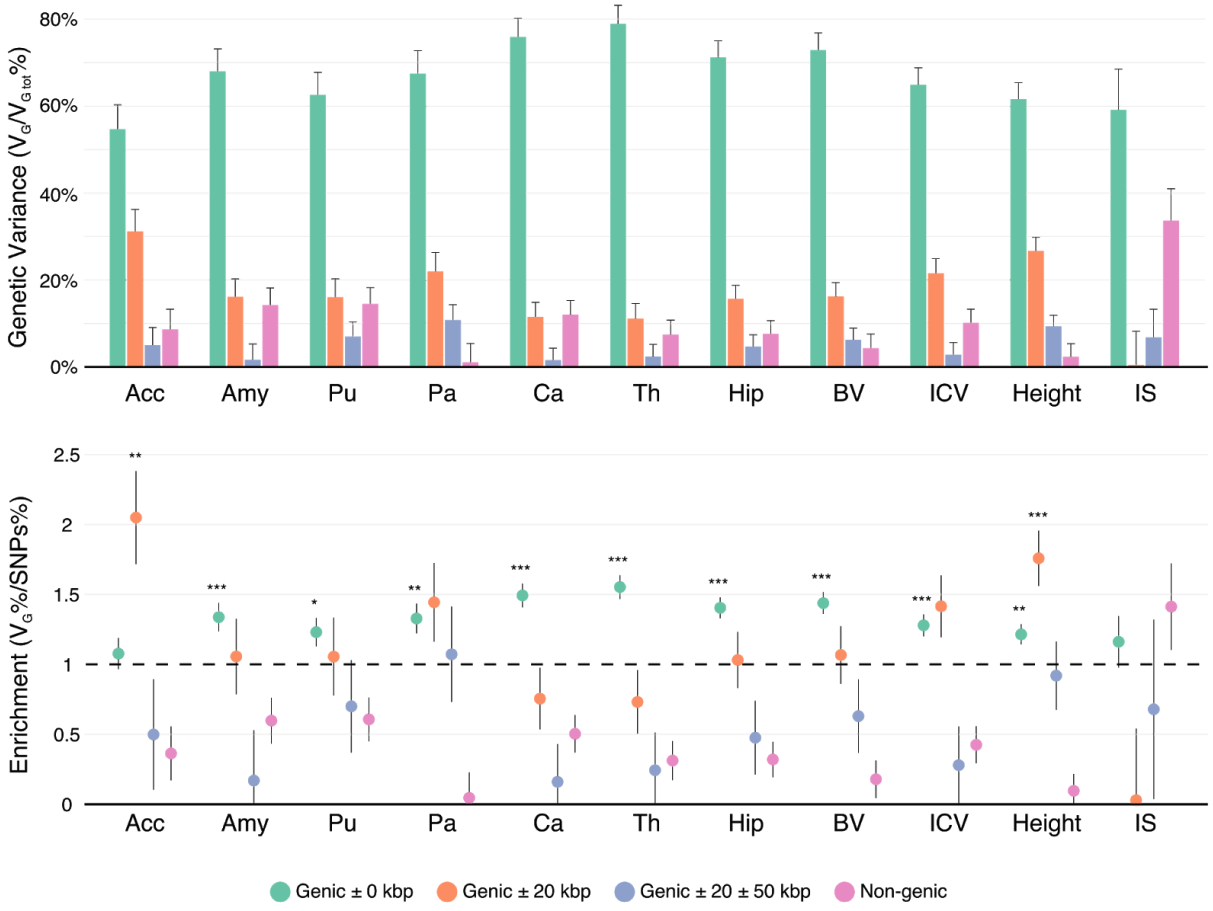


Figure 3. Variance enrichment for partitions based on closeness to genic regions. Meta-analytic estimates were obtained using inverse variance weighting of the estimates of the different projects studied. Top: V_G/V_{Gtot} estimates computed for four sets of SNPs based on their distance to gene boundaries: all SNPs within the boundaries of the 66,632 gene boundaries from the UCSC Genome Browser hg19 assembly, two further sets that included also SNPs 0 to 20 kbp and 20kbp to 50 kbp upstream and downstream of each gene, and a remaining set containing SNPs located farther than 50kb from one of the gene boundaries. V_G/V_{Gtot} estimates were computed using the GCTA REML unconstrained method for height, intelligence, and brain, intracranial and total subcortical volumes. The error bars represent the standard errors. Bottom: Enrichment of variance captured by each partition. The y-axis shows the ratio of the fraction of genetic variance explained by each partition divided by the fraction of SNPs contained in each partition. If all SNPs explained a similar amount of variance, this ratio should be close to 1 (dashed line). A Z-test was used to compare the ratios to 1 and p-values were FDR adjusted (*p<0.05, **p<0.01, ***p<0.001).

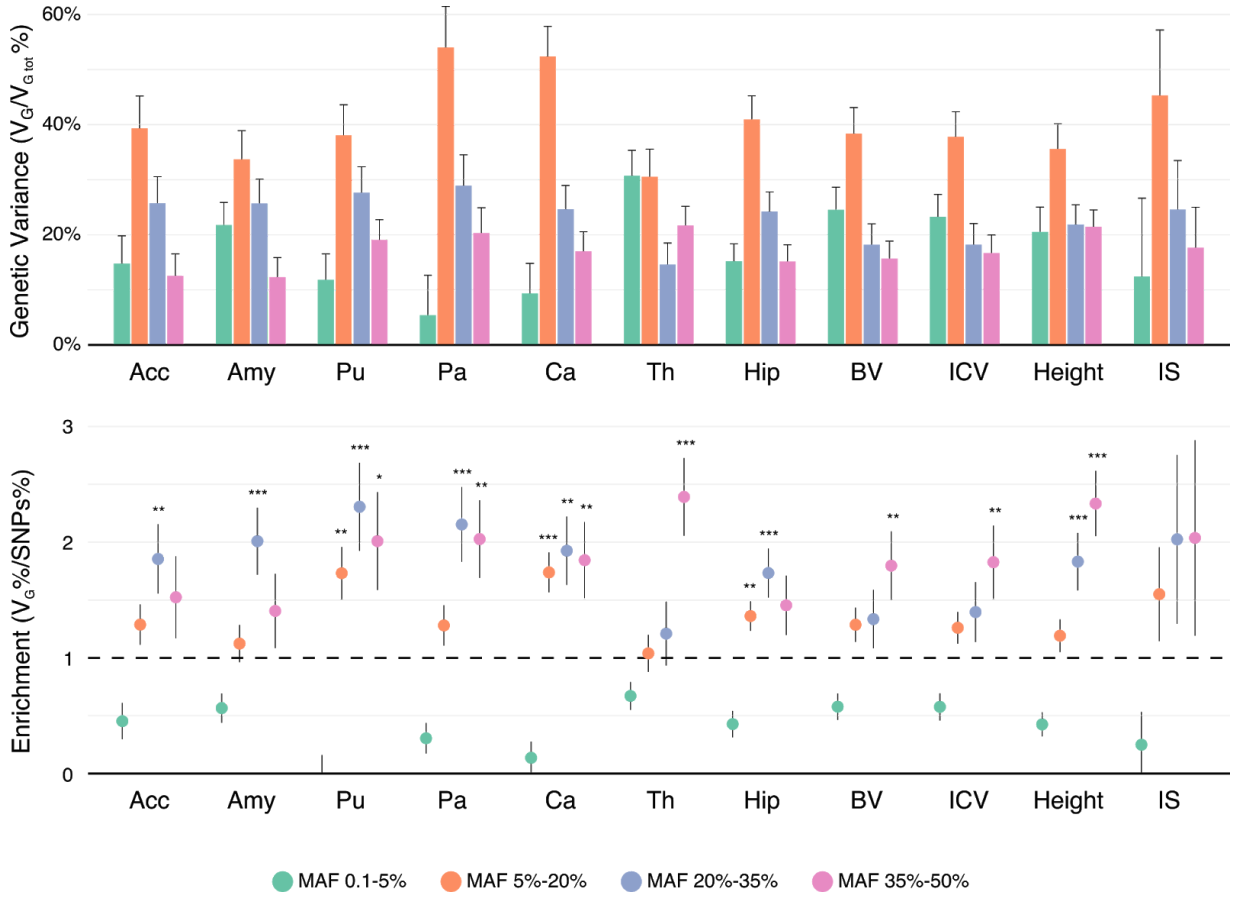


Figure 4. Variance enrichment for partitions based on minor allele frequency (MAF). Meta-analytic estimates were obtained using inverse variance weighting of the estimates of the different projects studied. Top: V_G/V_{Gtot} estimates computed for four sets of SNPs based on their MAF: from 0.1 to 5%, from 5 to 20%, from 20 to 35% and from 35 to 50%. $V_G/V_{G\,tot}$ estimates were computed using the GCTA REML unconstrained method for height, intelligence, and brain, intracranial and total subcortical volumes. The error bars represent the standard errors. Bottom: Enrichment of variance captured by each partition. The y-axis shows the ratio of the fraction of genetic variance explained by each partition divided by the fraction of SNPs contained in each partition. If all SNPs explained a similar amount of variance, this ratio should be close to 1 (dashed line). A Z-test was used to compare the ratios to 1 and p-values were FDR adjusted (*p<0.05, **p<0.01, ***p<0.001).

r_G = 1, and were in all cases statistically significantly larger than the phenotypic correlations (Z-test FDR<5%, Fig. S9 and Table S5), which suggests an important role of the environment in shaping hemispheric asymmetry. To further test this hypothesis, we measured the V_G/V_P of the differences between left and right volumes of each structure. Only the heritability of the volume asymmetry of the nucleus accumbens was significantly different from zero (V_G/V_P = 16 ± 4.5%). Except for this structure, the differences in volumes between right and left hemispheres of all the other brain regions seemed to be only of environmental origin.

Genetic factors influencing brain volume are shared with height and intelligence scores, but less between height and intelligence scores

The genetic correlation between height and brain volume ($r_{G:BV.height}$ = 0.164 ± 0.053) was smaller than the genetic correlation between brain volume and intelligence ($r_{G:BV.IS}$ = 0.343 ± 0.080). However, the difference was not statistically significant: $r_{G:BV.IS}$ - $r_{G:BV.height}$ = 0.179 ± 0.095, 95% CI: -0.008, 0.365 (computed using the formula xxxvii in Pearson and Filon 1898). The phenotypic and genetic correlations between intelligence scores and height were, however, the smallest we observed across all phenotypes: r_P = 0.101 ± 0.009 and r_G = 0.048 ± 0.073. This is different to what we had previously observed in the IMAGEN cohort (Toro et al. 2015), where the correlation between brain volume and intelligence scores was higher than the correlation between brain volume and height.

Heritability and genetic correlation estimates remain statistically significant after accounting for measurement errors

The estimates of heritability, phenotypic and environmental correlation can be biased by noise in the volume measurements, which could in particular explain the differences between genetic and phenotypic correlations of left and right volumes. We investigated this hypothesis by removing the effects of measurement errors estimated from the Consortium for Reliability and Reproducibility (Zuo et al. 2014)

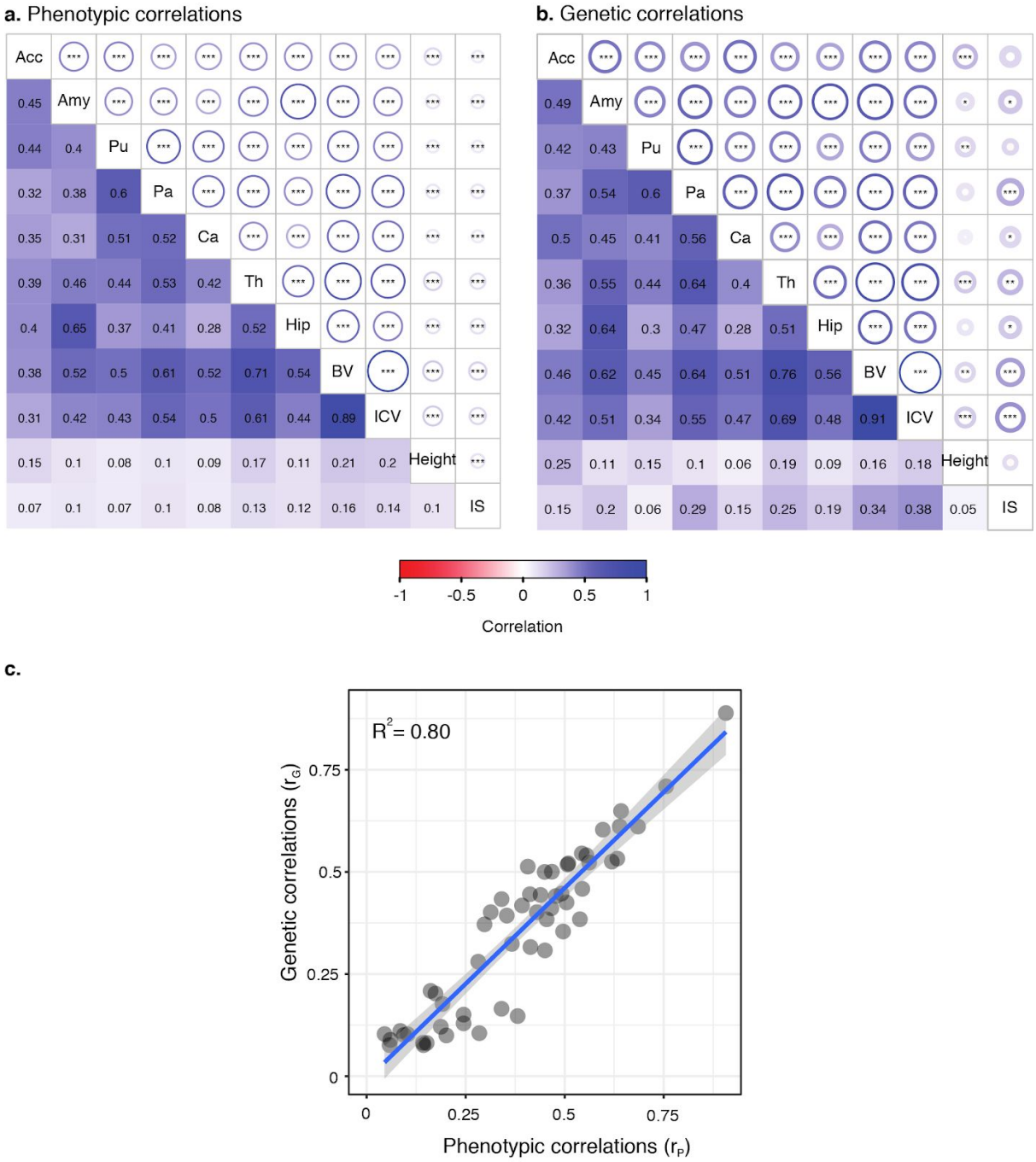


Figure 5. Phenotypic and genetic correlations. Significant phenotypic (a) and genetic (b) correlations were observed for most phenotypes. Correlation estimates are shown in the lower triangular part of the matrices, statistical significance in the upper triangular part. Circle radius represents correlation strength, stars indicate statistical significance of the correlation being non null (*P<0.05, **P<0.01, ***P<0.001). The scatter plot (c) of phenotypic versus genetic correlations.

for FreeSurfer segmentations, and from the repeated measures of intelligence score in UK Biobank. Intraclass correlation coefficient (ICC) for intelligence scores was estimated at 63% whereas the lowest ICCs for subcortical structures were found for accumbens (83%), pallidum (87%) and amygdala (89%). After correction for measurement error, height remained the most heritable phenotype although it was the only one that we did not correct. Disattenuated estimates of V_G/V_P for intelligence became 54 ± 7.8%, which is similar to the disattenuated heritabilities of brain volumes (between 46 ± 5.3% for pallidum and 61 ± 5.0% for thalamus) (Fig. S10). Genetic correlations between left and right hemispheres of subcortical structures remained statistically significantly greater than the adjusted phenotypic correlations except for pallidum and accumbens (Table S5 and Fig. S11). Estimates of V_G/V_P for the differences between left and right volumes remained low and not significantly different from 0 for all subcortical structures (between 5.4 ± 14% for pallidum and 16 ± 11% for thalamus) except for accumbens (34 ± 9.8%).

Genome-wide polygenic scores captured a statistically significant but small proportion of phenotypic variance

Genome-wide polygenic scores based on a GWAS of 13,086 subjects captured a statistically significant although very small proportion of brain region volume variance. The genome-wide polygenic scores were computed for ~6,000 additional participants from the UK Biobank who were not used in the GWAS. The predictions captured an amount of phenotypic variance ranging from 0.5% ($p < 10^{-22}$) for the amygdala to 2.2% ($p < 10^{-31}$) for brain volume (Fig. S15, Fig. S16, and Supplemental Table S8). For height, a genome-wide polygenic score obtained based on GWAS summary statistics from the 13,086 UK Biobank subjects with MRI captured ~3% of the variance. To evaluate the impact of the number of samples used in the GWAS on the amount of variance captured by the genome-wide polygenic scores, we also computed the genome-wide polygenic scores for height using the GWAS summary statistics from ~277k unrelated UK Biobank subjects not included in the validation dataset. This allowed us to capture ~27% of the variance of height (an >8-times increase) (Fig. S17).

Discussion

Our results suggest that neuroanatomical diversity is the product of a highly polygenic architecture, with SNPs capturing from 40% to 54% of regional brain volume variance, confirming our original findings (Toro et al., 2015) as well as those of others (Elliott et al. 2018, Zhao et al. 2018). At a global scale, causal variants were distributed across the genome: for different brain regions, chromosomes containing a larger number of SNPs captured a proportionally larger amount of variance than smaller chromosomes, with a correlation of r~0.64 on average. At a local scale, however, SNPs within genes (~51%) captured ~1.5 times more genetic variance than the rest; and SNPs with low minor allele frequency (MAF) captured significantly less variance than those with higher MAF.

When partitioning genetic variance into MAF bins, the lowest MAF partition, going from 0.1% to 5% (10% of the total MAF range) contained ~40% of all SNPs, but captured only ~16% of the total genetic variance: SNPs with low MAF captured significantly less variance than those with higher MAF. However, they captured more variance than expected under a neutral evolution model, where a MAF bin is expected to capture an amount of variance proportional to its size (10% size but 16% of the variance). This result suggests a negative selection model, where loci of large effect are being removed from the population. This apparent contradiction between high effect size and low captured variance for low MAF variants can be explained by the fact that the relationship between captured heritability h^2 and the allele effect b for a given SNP is dependent of its allele frequency p: $h^2 = 2\,b^2\,p\,(1-p)$ (Schoech et al. 2017, Zeng et al. 2018). Given this heterogeneity in effect sizes, using more flexible models that do not make strong assumptions on the relationships between effect size, LD, and MAF, such as GREML-LDMS (Yang et al. 2015), should improve the accuracy of the heritability estimates. Our study of rare variants is limited because of the use of datasets based on SNP arrays. The availability of whole-genome sequencing data for large cohorts is starting to allow the study of more refined partitions of rare variants and is showing

that rare causal variants might be a main source of the variance remaining to be explained (Wainschtein et al. 2019).

In addition to showing the heritability of regional volume diversity, our analyses also show an extensive pleiotropy across brain regions. The computation of genetic covariance for pairs of brain regions allowed us to estimate their genetic correlation. We observed an average genetic correlation of r_G~0.45. Interestingly, we observed that although genetic correlations were similar to phenotypic correlations across brain regions, if we compared the left and right aspects of the same brain region, their genetic correlations were close to 1 and systematically larger than phenotypic correlations. This could be an indication that the observed phenotypic asymmetries in regional brain volume are of environmental origin. To confirm this result, we used an alternative way of analysing the genetic/environmental nature of regional asymmetry, by looking at the heritability of the differences in left/right volumes. These heritabilities were generally not statistically significantly >0, which again supports the idea of the environmental nature of regional brain asymmetries.

The analysis of genetic correlations also allows us to explore the link between neuroanatomical diversity and other anatomical parametres or even cognitive functions. Brain volume is correlated with body size, and in the recent years the polygenic architecture of height has been well described (Yengo et al. 2018). It could be argued that neuroanatomical diversity is simply determined by the same genetic factors that produce body size variability in general. The genetic correlation between brain volume and height was, however, relatively small (r_G~0.16), suggesting little overlap between their genetic causes (in Toro et al. 2015 we did not find a significant genetic correlation, most likely due to lack of statistical power). By contrast, the genetic correlation between brain volume and intelligence scores was of medium strength (r_G~0.34), suggesting a larger overlap. The difference between these two genetic correlations was not significant and the question deserves further study. The genetic correlation between height and

intelligence scores was the smallest of all those we studied, r_G<0.05, suggesting that their relationship with brain volume may be due to different genetic factors.

One important source of bias in our estimates of heritability and genetic correlation could be related to measurement error. In MRI data, for example, some brain regions are more clearly delimited than others, which makes them easier to segment accurately. Furthermore, there is an important variability in volume across brain regions. Segmentation errors would then be comparatively more important for small, poorly delimited regions such as the amygdala, than for large regions like the thalamus. We sought to take into account errors in automatic segmentation of brain regions by analysing data from the CoRR project (Zuo et al. 2014), where the same subjects were scanned several times. For intelligence scores, we used the subset of UK Biobank where subjects passed the fluid intelligence test on multiple occasions. Our results remained for the most part unchanged after adjusting the estimates for phenotypic measurement errors. A limitation of this approach, however, is that we used MRI scans from a different project (although the processing pipeline was the same). Ideally, one would have preferred to have repeated scans for as subset of UK Biobank subjects, because segmentation quality depends on MRI quality which varies between datasets. The availability of repeated measurements in the same datasets from which genetic variance is estimated may allow to more precisely distinguish between the different sources of phenotypic variations.

Another possible source of bias could be the multi-centric nature of our data. The results presented here come from six different projects. Rather than combining our genetic raw data into a single dataset, we chose to estimate the part of phenotypic variance captured by SNPs independently in each dataset and then combine estimates in a meta-analysis. While our chosen solution helps to handle heterogeneity, it trades on statistical power. The reason for this is that the standard error of the GCTA GREML heritability estimate is approximately inversely proportional to the number of subjects (Visscher et al. 2014) (Fig. S14). In the present analysis, the UK Biobank project accounted for ~94% of the estimates, largely driving the results. Indeed, the inclusion of the other projects reduced the standard error of the estimates based

on the UK Biobank project only by a factor of ~1.03. If the raw genotyping data had been combined in a single mega-analysis instead of in a meta-analysis, the decrease in standard error would have been ~1.36 smaller than what we reported. The meta-analytical approach may however prove interesting in the future when large imaging genetics datasets other than UK Biobank will be available or in cases where raw genotyping data cannot be easily shared.

Our analyses are also limited in their ability to provide information at the individual level. With the advent of large GWAS studies, genome-wide polygenic scores (GPS) have become increasingly used to predict phenotypes from whole-genome genotyping data. GPSs are based on effect sizes estimated through GWAS in a large population. These estimates can then be used to predict the phenotype in an independent sample. We aimed at evaluating to which extent GWAS of the N~13k UK Biobank subjects used for our heritability analyses allowed us to predict the phenotypes of additional N~6k subjects. GPSs captured a very small, although statistically significant, proportion of the variance. For brain volume, for example, GPS captured ~2.5% of the variance. This is the expected result in the presence of a strongly polygenic phenotype (Wray et al. 2013). It is also expected that prediction accuracy will improve as the number of subjects used for effect size estimation increases (the UK Biobank project alone should provide data for N~100k in the years to come). We aimed at testing the potential increase in predictive power by computing a GPS for height. When effect sizes were estimated from N~13k subjects, GPS captured ~3% of height variance, however, when effect sizes were estimated from N~277k subjects, the amount of variance explained increased to ~27%. We expect that effect sizes for regional brain volume estimated from N~100k will allow us to compute GPSs capturing ~12% of the variance (based on Wray et al. 2013 and Daetwyler, Villanueva, and Woolliams 2008).

The detection of candidate genes of large effect is an appealing tool for gaining mechanistic insight on normal and pathological phenotypes. However, this approach is ill adapted to strongly polygenic architectures, where not only a few large-effect alleles are involved, but potentially hundreds of thousands of alleles of almost infinitesimal effect (Wray et al. 2018). Neuroimaging endophenotypes such as those

obtained using structural and functional MRI could provide an alternative source of mechanistic insight. The brain imaging literature is rich in examples of associations between different brain regions and networks with normal and pathological cognitive phenotypes. The automatic mining of these associations could provide a layer of annotation for brain regions and networks similar to those available today for genome annotation. Further investigation of the genomic architecture of neuroimaging endophenotypes should prove an important tool to better understand the biological basis of brain diversity and evolution in humans, as well as the biological basis of the susceptibility to psychiatric disorders.

Funding

This work was supported by the Institut Pasteur; Center for Research and Interdisciplinarity (CRI), Centre National de la Recherche Scientifique (CNRS); the University Paris Diderot; the Fondation pour la Recherche Médicale [DBI20141231310]; the European Commission Horizon 2020 [COSYN]; The Human Brain Project; the European Commission Innovative Medicines Initiative [AIMS2-TRIALS; No. 777394]; the Cognacq-Jay foundation; the Bettencourt-Schueller foundation; the Orange foundation; the FondaMental foundation; the Conny-Maeva foundation; and the Agence Nationale de la Recherche (ANR) [SynPathy]; the Laboratory of Excellence GENMED (Medical Genomics) grant no. ANR-10-LABX-0013, Bio-Psy; by the INCEPTION program ANR-16-CONV-0005, managed by the ANR part of the Investment for the Future program.

J.-B.P. was partially funded by NIH-NIBIB P41 EB019936 (ReproNim) NIH-NIMH R01 MH083320 (CANDIShare) and NIH 5U24 DA039832 (NIF), as well as the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative.

The Study of Health in Pomerania (SHIP) is part of the Community Medicine Research net (CMR) (http://www.medizin.uni-greifswald.de/icm) of the University Medicine Greifswald, which is supported by the German Federal State of Mecklenburg- West Pomerania. MRI scans in SHIP and SHIP-TREND have been supported by a joint grant from Siemens Healthineers, Erlangen, Germany and the Federal State of

Mecklenburg-West Pomerania. This study was further supported by the EU-JPND Funding for BRIDGET (FKZ:01ED1615).

References

Alfaro-Almagro, Fidel, Mark Jenkinson, Neal K. Bangerter, Jesper L. R. Andersson, Ludovica Griffanti, Gwenaëlle Douaud, Stamatios N. Sotiropoulos, et al. 2018. “Image Processing and Quality Control for the First 10,000 Brain Imaging Datasets from UK Biobank.” *NeuroImage* 166 (February): 400–424. https://doi.org/10.1016/j.neuroimage.2017.10.034.

Browning, Sharon R., and Brian L. Browning. 2011. “Population Structure Can Inflate SNP-Based Heritability Estimates.” *The American Journal of Human Genetics* 89 (1): 191–93. https://doi.org/10.1016/j.ajhg.2011.05.025.

Daetwyler, Hans D., Beatriz Villanueva, and John A. Woolliams. 2008. “Accuracy of Predicting the Genetic Risk of Disease Using a Genome-Wide Approach.” *PLOS ONE* 3 (10): e3395. https://doi.org/10.1371/journal.pone.0003395.

Elliott, Lloyd T., Kevin Sharp, Fidel Alfaro-Almagro, Sinan Shi, Karla L. Miller, Gwenaëlle Douaud, Jonathan Marchini, and Stephen M. Smith. 2018. “Genome-Wide Association Studies of Brain Imaging Phenotypes in UK Biobank.” *Nature* 562 (7726): 210. https://doi.org/10.1038/s41586-018-0571-7.

Euesden, Jack, Cathryn M. Lewis, and Paul F. O’Reilly. 2015. “PRSice: Polygenic Risk Score Software.” *Bioinformatics* 31 (9): 1466–68. https://doi.org/10.1093/bioinformatics/btu848.

Fisher, Ronald Aylmer. 1970. *Statistical Methods for Research Workers*. 14th ed., revised and enlarged. Edinburgh: Oliver and Boyd.

Hibar, Derrek P., Jason L. Stein, Miguel E. Renteria, Alejandro Arias-Vasquez, Sylvane Desrivières, Neda Jahanshad, Roberto Toro, et al. 2015. “Common Genetic Variants

Influence Human Subcortical Brain Structures." *Nature* 520 (7546): 224–29. https://doi.org/10.1038/nature14101.

Lee, S Hong, Teresa R DeCandia, Stephan Ripke, Jian Yang, Patrick F Sullivan, Michael E Goddard, Matthew C Keller, Peter M Visscher, and Naomi R Wray. 2012. "Estimating the Proportion of Variation in Susceptibility to Schizophrenia Captured by Common SNPs." *Nature Genetics* 44 (3): 247–50. https://doi.org/10.1038/ng.1108.

Lee, S Hong, Jian Yang, Michael E Goddard, Peter M Visscher, and Naomi R Wray. 2012. "Estimation of Pleiotropy between Complex Diseases Using Single-Nucleotide Polymorphism-Derived Genomic Relationships and Restricted Maximum Likelihood." *Bioinformatics (Oxford, England)* 28 (19): 2540–42. https://doi.org/10.1093/bioinformatics/bts474.

Li, Mingfeng, Gabriel Santpere, Yuka Imamura Kawasawa, Oleg V. Evgrafov, Forrest O. Gulden, Sirisha Pochareddy, Susan M. Sunkin, et al. 2018. "Integrative Functional Genomic Analysis of Human Brain Development and Neuropsychiatric Risks." *Science* 362 (6420): eaat7615. https://doi.org/10.1126/science.aat7615.

Plomin, R., and I. J. Deary. 2015. "Genetics and Intelligence Differences: Five Special Findings." *Molecular Psychiatry* 20 (1): 98–108. https://doi.org/10.1038/mp.2014.105.

Ponzi, Erica, Lukas F. Keller, Timothée Bonnet, and Stefanie Muff. 2018. "Heritability, Selection, and the Response to Selection in the Presence of Phenotypic Measurement Error: Effects, Cures, and the Role of Repeated Measurements." *Evolution* 72 (10): 1992–2004. https://doi.org/10.1111/evo.13573.

Price, Alkes L., Michael E. Weale, Nick Patterson, Simon R. Myers, Anna C. Need, Kevin V. Shianna, Dongliang Ge, et al. 2008. "Long-Range LD Can Confound Genome Scans in Admixed Populations." *American Journal of Human Genetics* 83 (1): 132–35.

https://doi.org/10.1016/j.ajhg.2008.06.005.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.” *American Journal of Human Genetics* 81 (3): 559–75. https://doi.org/10.1086/519795.

Raychaudhuri, Soumya, Joshua M. Korn, Steven A. McCarroll, The International Schizophrenia Consortium, David Altshuler, Pamela Sklar, Shaun Purcell, and Mark J. Daly. 2010. “Accurately Assessing the Risk of Schizophrenia Conferred by Rare Copy-Number Variation Affecting Genes with Brain Function.” *PLOS Genetics* 6 (9): e1001097. https://doi.org/10.1371/journal.pgen.1001097.

Schoech, Armin, Daniel Jordan, Po-Ru Loh, Steven Gazal, Luke O’Connor, Daniel J. Balick, Pier F. Palamara, Hilary Finucane, Shamil R. Sunyaev, and Alkes L. Price. 2017. “Quantification of Frequency-Dependent Genetic Architectures and Action of Negative Selection in 25 UK Biobank Traits.” *BioRxiv*, September, 188086. https://doi.org/10.1101/188086.

Skene, Nathan G, Julien Bryois, Trygve E Bakken, Gerome Breen, James J Crowley, Héléna A Gaspar, Paola Giusti-Rodriguez, et al. 2018. “Genetic Identification of Brain Cell Types Underlying Schizophrenia.” *Nature Genetics* 50 (6): 825–33. https://doi.org/10.1038/s41588-018-0129-5.

Stein, Jason L., Sarah E. Medland, Alejandro Arias Vasquez, Derrek P. Hibar, Rudy E. Senstad, Anderson M. Winkler, Roberto Toro, et al. 2012. “Identification of Common Variants Associated with Human Hippocampal and Intracranial Volumes.” *Nature Genetics* 44 (5): 552–61. https://doi.org/10.1038/ng.2250.

Taki, Yasuyuki, Hiroshi Hashizume, Yuko Sassa, Hikaru Takeuchi, Michiko Asano, Kohei

Asano, Yuka Kotozaki, et al. 2012. “Correlation among Body Height, Intelligence, and Brain Gray Matter Volume in Healthy Children.” *NeuroImage* 59 (2): 1023–27. https://doi.org/10.1016/j.neuroimage.2011.08.092.

Toro, R., J.-B. Poline, G. Huguet, E. Loth, V. Frouin, T. Banaschewski, G. J. Barker, et al. 2015. “Genomic Architecture of Human Neuroanatomical Diversity.” *Molecular Psychiatry* 20 (8): 1011–16. https://doi.org/10.1038/mp.2014.99.

Visscher, Peter M., Gibran Hemani, Anna A. E. Vinkhuyzen, Guo-Bo Chen, Sang Hong Lee, Naomi R. Wray, Michael E. Goddard, Jian Yang, and Gregory S. Barsh. 2014. “Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples.” *PLoS Genetics* 10 (4): e1004269. https://doi.org/10.1371/journal.pgen.1004269.

Wainschtein, Pierrick, Deepti P. Jain, Loic Yengo, Zhili Zheng, Trans-Omics for Precision Medicine Consortium TOPMed Anthropometry Working Group, L. Adrienne Cupples, Aladdin H. Shadyab, et al. 2019. “Recovery of Trait Heritability from Whole Genome Sequence Data.” *BioRxiv*, March, 588020. https://doi.org/10.1101/588020.

Wray, Naomi R., Cisca Wijmenga, Patrick F. Sullivan, Jian Yang, and Peter M. Visscher. 2018. “Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model.” *Cell* 173 (7): 1573–80. https://doi.org/10.1016/j.cell.2018.05.051.

Wray, Naomi R., Jian Yang, Ben J. Hayes, Alkes L. Price, Michael E. Goddard, and Peter M. Visscher. 2013. “Pitfalls of Predicting Complex Traits from SNPs.” *Nature Reviews Genetics* 14 (7): 507–15. https://doi.org/10.1038/nrg3457.

Yang, Jian, Andrew Bakshi, Zhihong Zhu, Gibran Hemani, Anna A. E. Vinkhuyzen, Ilja M. Nolte, Van Vliet-Ostaptchouk, et al. 2015. “Genome-Wide Genetic Homogeneity between Sexes and Populations for Human Height and Body Mass Index.” *Human Molecular Genetics* 24 (25): 7445–49. https://doi.org/10.1093/hmg/ddv443.

Yang, Jian, Beben Benyamin, Brian P McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, et al. 2010. “Common SNPs Explain a Large Proportion of the Heritability for Human Height.” *Nature Genetics* 42 (7): 565–69. https://doi.org/10.1038/ng.608.

Yang, Jian, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. 2011. “GCTA: A Tool for Genome-Wide Complex Trait Analysis.” *The American Journal of Human Genetics* 88 (1): 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.

Yang, Jian, Teri A. Manolio, Louis R. Pasquale, Eric Boerwinkle, Neil Caporaso, Julie M. Cunningham, Mariza de Andrade, et al. 2011. “Genome Partitioning of Genetic Variation for Complex Traits Using Common SNPs.” *Nature Genetics* 43 (6): 519–25. https://doi.org/10.1038/ng.823.

Yengo, Loic, Julia Sidorenko, Kathryn E. Kemper, Zhili Zheng, Andrew R. Wood, Michael N. Weedon, Timothy M. Frayling, Joel Hirschhorn, Jian Yang, and Peter M. Visscher. 2018. “Meta-Analysis of Genome-Wide Association Studies for Height and Body Mass Index in ~700000 Individuals of European Ancestry.” *Human Molecular Genetics* 27 (20): 3641–49. https://doi.org/10.1093/hmg/ddy271.

Zeng, Jian, Ronald de Vlaming, Yang Wu, Matthew R. Robinson, Luke R. Lloyd-Jones, Loic Yengo, Chloe X. Yap, et al. 2018. “Signatures of Negative Selection in the Genetic Architecture of Human Complex Traits.” *Nature Genetics* 50 (5): 746. https://doi.org/10.1038/s41588-018-0101-4.

Zhao, Bingxin, Joseph G. Ibrahim, Yun Li, Tengfei Li, Yue Wang, Yue Shan, Ziliang Zhu, et al. 2018. “Heritability of Regional Brain Volumes in Large-Scale Neuroimaging and Genetic Studies.” *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhy157.

Zuo, Xi-Nian, Jeffrey S. Anderson, Pierre Bellec, Rasmus M. Birn, Bharat B. Biswal, Janusch

Blautzik, John C. S. Breitner, et al. 2014. “An Open Science Resource for Establishing Reliability and Reproducibility in Functional Connectomics.” *Scientific Data* 1 (December): 140049. https://doi.org/10.1038/sdata.2014.49.

Supplementary Material

1. Comparison of methods for the estimation of enrichment in genetic variance partitioning.

We can estimate the significance of the difference in variance captured by an individual partition against a model where each SNP captures exactly the same amount of variance. In this latter case, the amount of variance captured by each partition should be directly proportional to the number of SNPs it contains.

We first aimed to assess the statistical significance of the partition enrichment using a permutation test as in Toro et al. (2015) and a derivation of statistical significance obtained from the covariance matrix of variance estimates reported by GCTA.

In the permutation test, SNPs were randomly selected to build the partitions, keeping the same number of SNPs as in the original partition. For each set of random partitions, the same linear mixed-effects model as before was fitted (including age, sex, centre and 10 PCs). We tested whether any of these partitions captured more variance than what could be expected given its number of SNPs. Briefly, a Z-score was computed by comparing the SNP-set genetic estimated variance V_{G_i} of partition i to the SNP-set genetic variance $f_i \cdot V_{G_{tot}}$ expected under no enrichment:

$$Z = \frac{V_{G_i} - f_i \cdot V_{G_{tot}}}{\sqrt{Var(V_{G_i} - f_i \cdot V_{tot})}} \quad \text{(Eq. 1)}$$

where f_i is the fraction of the SNPs included in partition i. The p-value of enrichment was computed by comparing the observed Z score to those obtained from 1,000 permutations.

One limitation of this approach is that it does not preserve the local LD relationships among SNPs. We observed that the standard errors of the genetic variance estimates reported by the permutation approach were systematically larger than those computed from the covariance matrix of variance estimates. To better preserve the original LD, we tried an alternative permutation method in which we permuted blocks of contiguous SNPs. For the partitioning based on genic status, the standard errors of the simulated partitions estimates were compatible with the standard error of the original partitions (Fig. S1). However, this alternative permutation method did not reduce the gap with the theoretical values for the partitioning by MAF since they depend only on the frequency of the individual SNPs, and not on their contiguity over the genome (Fig. S2). Because of this, and because the permutation approach was much more demanding in terms of computation, we decided to use only the theoretical derivation of enrichment test: We evaluated whether $V_{G_i} - f_i \cdot V_{G_{tot}}$ was statistically significantly positive with a one-sided Z-test, considering that in the null hypothesis of no enrichment, each genomic partition should carry an amount of variance proportional to its number of SNPs. We estimated the variance of the observed enrichment using this equation:

$$Var(V_{G_i} - f_i \cdot V_{G_{tot}}) = Var(V_{G_i}) + f_i^2 \cdot Var(V_{G_{tot}}) - 2 \cdot f_i \cdot Cov(V_{G_i}, V_{G_{tot}}). \quad \text{(Eq. 2)}$$

Here, V_{G_i} represents the genetic variance of partition i and $V_{G_{tot}} = \sum_i V_{G_i}$.

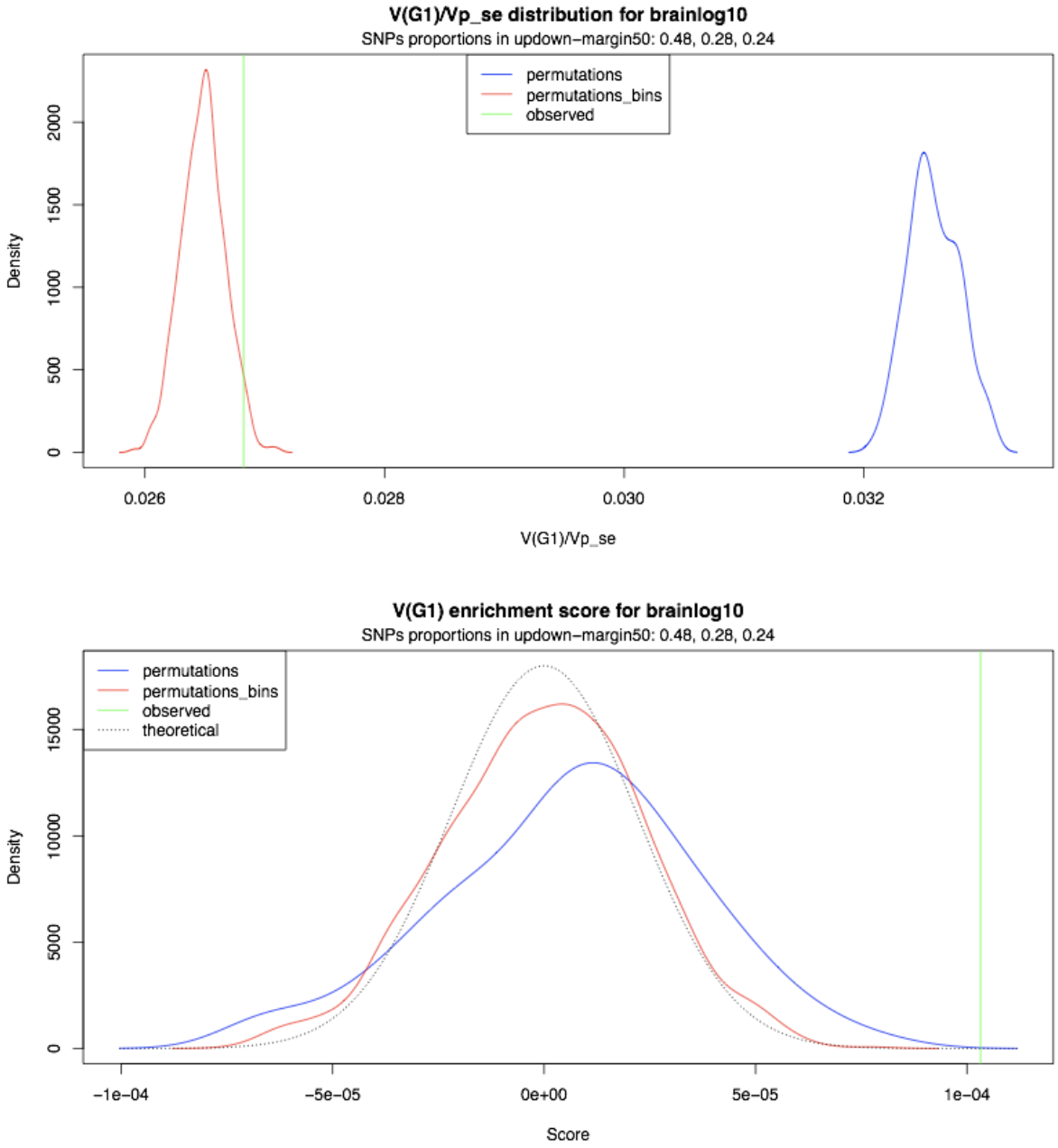


Figure S1. Simulation of partitioning the genetic variance between genic, near genic and non genic regions. **Upper panel:** Density of standard errors reported by GCTA in permutation vs standard error reported by GCTA in the original partition. SNPs. **Lower panel:** Density of genic region enrichment for permutations in the null hypothesis of no enrichment computed as the difference between V_{G_i} and $E(V_{G_i}) = V_{G_{tot}} \cdot f_i$. The observed value is the reported value for the real partition. Permutations were made for UK biobank dataset with 1000 iterations. permutations: permutations by SNPs. permutations_bins: permutations by blocks of contiguous SNPs.

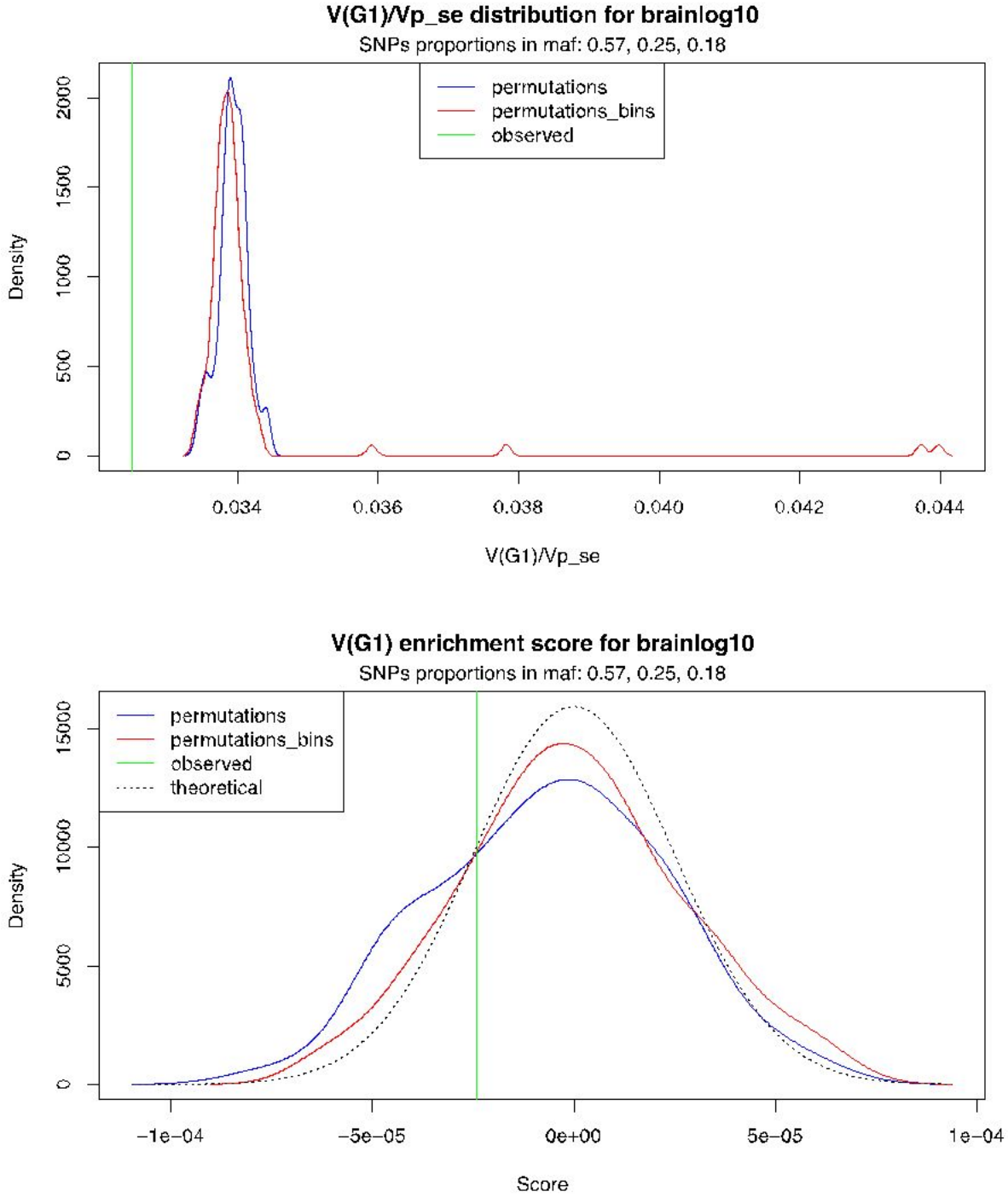


Figure S2. Simulation of partitioning the genetic variance in function of MAF (0.05-0.20, 0.20-0.35, 0.35-0.5). **Upper panel:** Density of standard errors reported by GCTA in permutation vs standard error reported by GCTA in the original partition. SNPs. **Lower panel:** Density of low MAF enrichment for permutations in the null hypothesis of no enrichment computed as the difference between V_{G_i} and $E(V_{G_i}) = V_{G_{tot}} \cdot f_i$. The observed value is the reported value for the real partition. Permutations were made for UK biobank dataset with 1000 iterations. permutations: permutations by SNPs. permutations_bins: permutations by blocks of contiguous SNPs.

2. Genetic variance estimates in UK Biobank with different set of covariates

We compared the estimations of genetic variance for volumes measured using Freesurfer and FIRST in the UK Biobank project obtained by including or not the 10 first principal components of the genetic relationship matrix, and by including or not brain volume (BV) as a covariate in the linear model. The results with an without the top 10 PCs are included in supplemental table S2.

We observed that V_G/V_P estimates diminished when no covariate was included, meaning that covariates globally capture more environmental variability. The inclusion of brain volume as a covariate did not seem to have a large impact on the estimations.

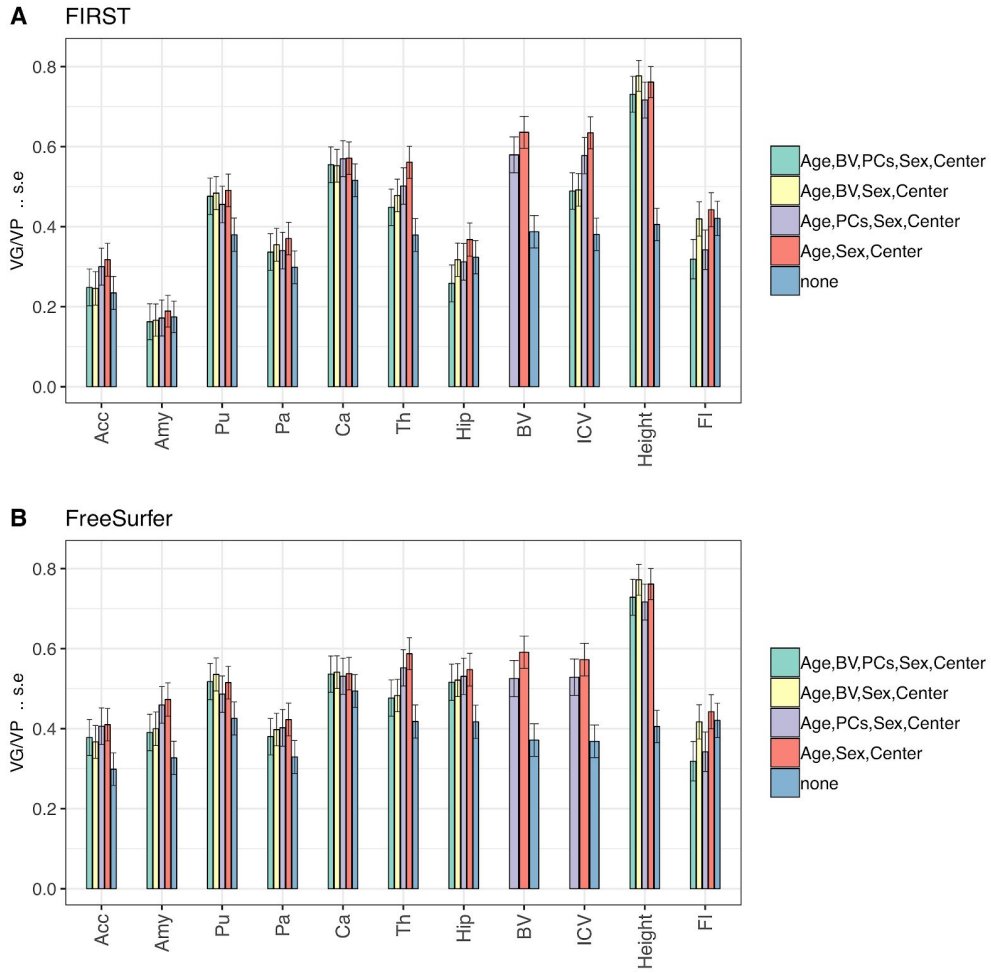


Figure S3. Fraction of phenotypic variance (VP) accounted for by genetic variance(VG), computed from all genotyped SNPs for the UK Biobank dataset. V_G/V_P estimates were computed using GCTA REML unconstrained method, for total volumes estimated using either FIRST (A) or FreeSurfer (B). The bar plots also include estimates of V_G/V_P for height and fluid intelligence. The colors represent results obtained with four different sets of covariates (BV denotes the brain volume, PCs the top 10 principal components).

3. Correlation among the phenotypes studied, and comparison between volume measurement tools.

We computed the raw correlation matrix of all the phenotypes studied. Figure S3 shows regional volumes computed using Freesurfer, with left and right volumes of each region plot separately. Figure S4 shows the correlation of the volume measurements obtained with Freesurfer and FIRST, with left and right volumes of each structure combined. We observed that patterns of correlations differed across structures, with accumbens and amygdala displaying less correlation than the other structures between their left and right parts. Correlation between volumes measured by either FIRST or FreeSurfer were different across structures, the lowest consistency being found for amygdala (r = 0.621).

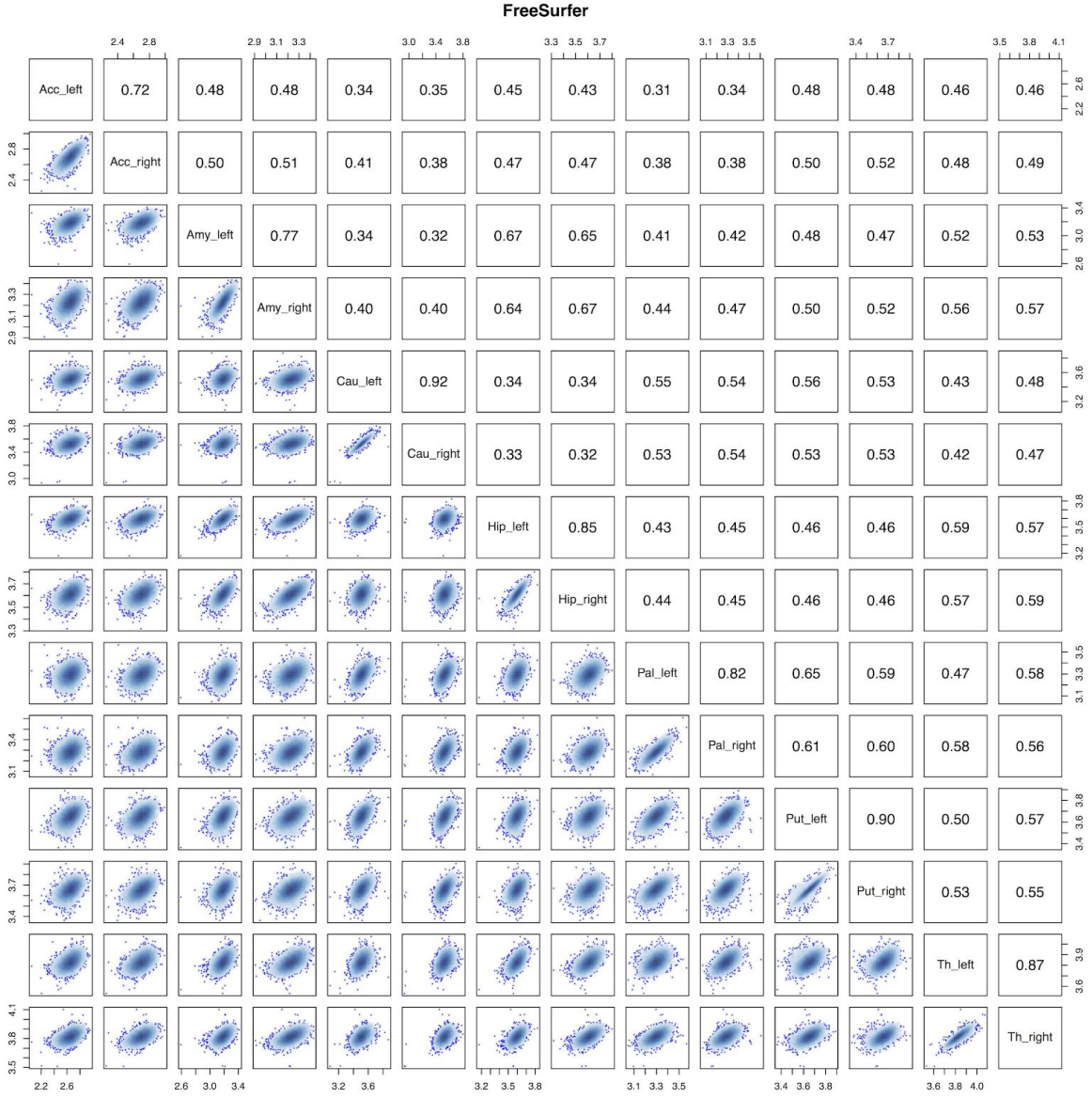


Figure S4. Pairs plot of the volumes of the left and right hemispheres of the subcortical structures measured using FreeSurfer (Acc : Accumbens, Amy : Amygdala, Cau : Caudate, Hip : Hippocampus, Pal : Pallidum, Put : Putamen, Th : Thalamus).

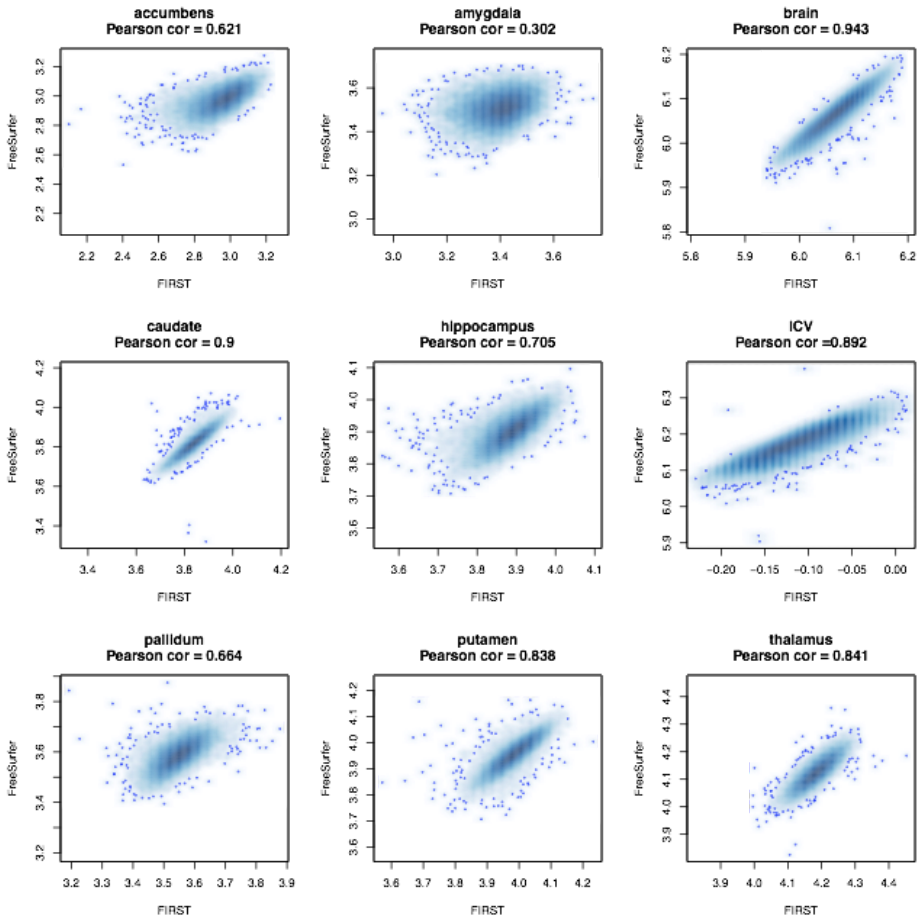


Figure S5. Scatter plots of total volumes (sum of left and right hemispheres) measured by FIRST (x-axis) and FreeSurfer (y-axis).

4. Comparison of genetic variance estimates with those of previous studies of the UK Biobank and ADNI projects by Elliott et al (2018) and Zhao et al (2018).

Several recent studies have used the UK Biobank data to estimate genetic variance of regional brain volumes. The reported estimates vary across studies, with subcortical volumes, in particular the volume of the amygdala, showing the largest differences. We compared our genetic variance estimates with those obtained by two recent studies (Elliott et al. 2018; Zhao et al. 2018). Zhao et al. (2018) used GCTA to compute V_G/V_P estimates of the left and right volumes of various subcortical structures on a smaller sample of 9,031 subjects from the UK Biobank project. Regional brain volumes were quantified using ANTs (Advanced Normalization Tools software (Avants et al. 2014) together with the MindBoggle-101 atlas (Klein and Tourville 2012). Elliott et al. (2018) reported genetic variance estimates for thousands of phenotypes on 8,411 subjects, using SBAT (Sparse Bayesian Association Test (Elliott et al. 2018)). These phenotypes included the volumes of subcortical structures that we report here. Elliott et al. (2018) provide V_G/V_P estimates obtained from measurements of subcortical structures performed using FSL's FIRST (FMRIB's Integrated Registration and Segmentation Tool (Patenaude et al. 2011)) and FreeSurfer. In previous study of the IMAGEN data we reported only total volumes (left plus right) computed using FIRST. For the sake of comparison, we computed V_G/V_P estimates in the UK Biobank data for the left and right volumes separately, using both FIRST and FreeSurfer (Table S2).

The largest difference of hemispheric variance estimates was observed for the amygdala and the nucleus accumbens (Fig. S6a), which were the smallest structures of those we studied. These structures also displayed the lowest correlation between right and left volumes (Fig. S4) and between FIRST and FreeSurfer measurements (Fig. S5), which suggests that their measurement may be less precise than for larger structures. A similar discrepancy between left and right amygdala volume was observed by both Zhao et al. (2018) and Elliott et al. (2018) (Fig. S6a). Discrepancies between V_G/V_P estimates were more important when comparing different methods to estimate subcortical volumes (FreeSurfer, FIRST, or MindBoggle/ANTs) than when comparing different methods to estimate genetic variance (GCTA or SBAT). The inclusion of the brain volume as a covariate did not change noticeably our variance estimates and thus could not fully explain the differences observed with the study of Zhao et al. (2018) (Fig. S6b). In all cases, however, 95% CIs built from the standard errors of each study contained the estimates of the other studies.

We also compared our V_G/V_P estimates for the ADNI project with those obtained by Zhao et al. (2018). After filtering and merging of the non-imputed genotypes across the two ADNI sub-projects, we included 982 subjects and 210,543 SNPs; Zhao et al. (2018) included 1,023 subjects and 7,368,446 SNPs (after imputation). Despite both analyses using the same covariates, the V_G/V_P estimates showed differences ranging from 7% to 99%. These differences are not surprising, however, given the small sample size, and thus the large standard errors of the estimates (Fig. S6b).

In conclusion, when comparing heritability estimates reported by different studies, the method used to measure volumes had a larger impact on the estimation of genetic variance than the method used for the estimation of heritability. The smallest subcortical structures such as the nucleus accumbens, the amygdala and the pallidum seemed to be the most difficult to segment accurately. Large volume differences both between manual and automated segmentation and between FreeSurfer and FIRST segmentations have been previously reported, in particular for the amygdala (Morey et al. 2009; Schoemaker et al. 2016).

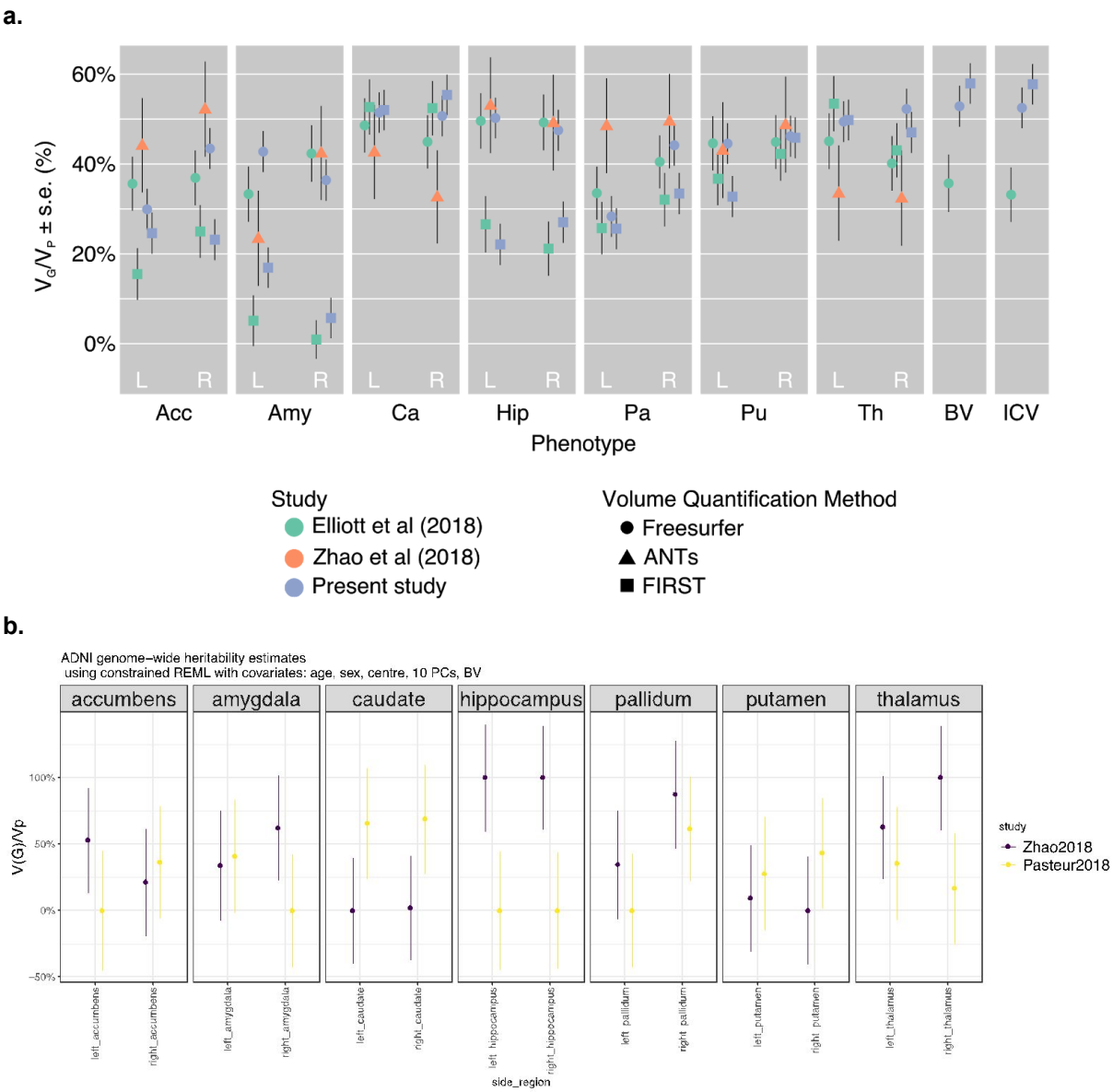


Figure S6. (a) Comparison of our estimates of genetic variance (V_G/V_P) in the UK Biobank project with two recent studies: Elliot et al. (2018) and Zhao et al. (2018). V_G/V_P estimates were computed for brain, intracranial volume (ICV), and subcortical volumes of the right and left hemispheres measured using different methods (FreeSurfer, or MindBoggle/ANTs, FIRST) represented by the shapes of the dots. In yellow are represented the estimates of V_G/V_P we computed using GCTA REML unconstrained method, the other V_G/V_P estimates were published by Elliott et al. 2018 (green) and Zhao et al. 2018 (orange). Age, sex, center, and the top 10 principal components were included as covariates. The error bars show the standard errors of the V_G/V_P estimates. (b) Fraction of phenotypic variance (V_P) accounted for by genetic variance (V_G), computed from all genotyped SNPs for the volumes of six subcortical structures in the ADNI dataset. We observed that the sample size was too small to get reliable estimations. Estimates published by Zhao et al., 2018 are shown in violet, estimates we obtained using GCTA REML constrained method with covariates similar to the ones used by Zhao et al. (age, sex, center, 10 top principal components, brain volume, ADNI phase) are shown in yellow.

5. Comparison of phenotypic and genetic correlations

We computed matrices of phenotypic correlations (r_P in Fig. S7, left panel) and genetic correlations (r_G in Fig. S7, right panel). The correlations among pairs of phenotypes were computed using GCTA. We used Ward hierarchical clustering based on euclidean distance to cluster the phenotypes.

Additionally, we computed environmental correlations which are the environmental counterpart of genetic correlations (Fig. S8, left panel). They represent the correlation between the environmental effects of two traits. They were computed using the environmental variances of each trait and the environmental covariance between the two traits output by GCTA ($r_E = \frac{Cov(E_{tr1}, E_{tr2})}{\sqrt{Var(E_{tr1}) \cdot Var(E_{tr2})}}$). Standard errors were computed with the delta method. The concordance between genetic and environmental correlation was medium (R^2 = 0.47) (Fig S8, right panel).

Correlations were overall strong and significant, especially between the left and right parts of the same structure. Figure S9 shows the values of the differences between genetic and environmental correlations (r_G - r_E, left panel) and genetic and phenotypic (r_G - r_P, right panel) are shown in the lower triangular part of the matrix. Differences between genetic and environmental correlations were larger than the differences between genetic and phenotypic correlations. Statistical significance was assessed using a Z-test to determine whether the differences were non null (Benjamini Yekutieli FDR adjusted *P<0.05, **P<0.01, ***P<0.001). Table S5 provides the values of the genetic, phenotypic, and environmental correlations and the results of their comparison.

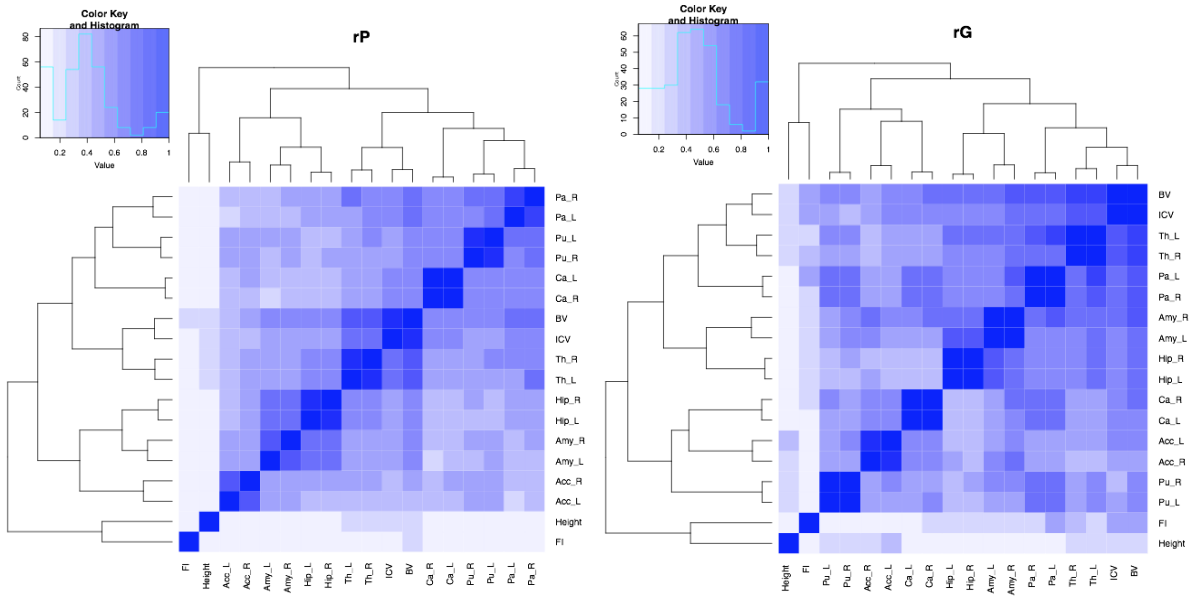


Figure S7. Heat maps of phenotypic correlations (r_P, left panel) and genetic correlation (r_G, right panel) values computed with GCTA. The following phenotypes are represented: the left and right hemisphere volumes of accumbens (Acc_L and Acc_R), amygdala (Amy_L and Amy_R), caudate (Cau_L and Cau_R), hippocampus (Hip_L and Hip_R), pallidum (Pa_L and Pa_R), putamen (Pu_L and Pu_R), and thalamus (Th_L and Th_R) ; brain volume (BV), intracranial volume (ICV), height and fluid intelligence (FI) are also represented. Ward hierarchical clustering based on euclidean distance was used to cluster the phenotypes.

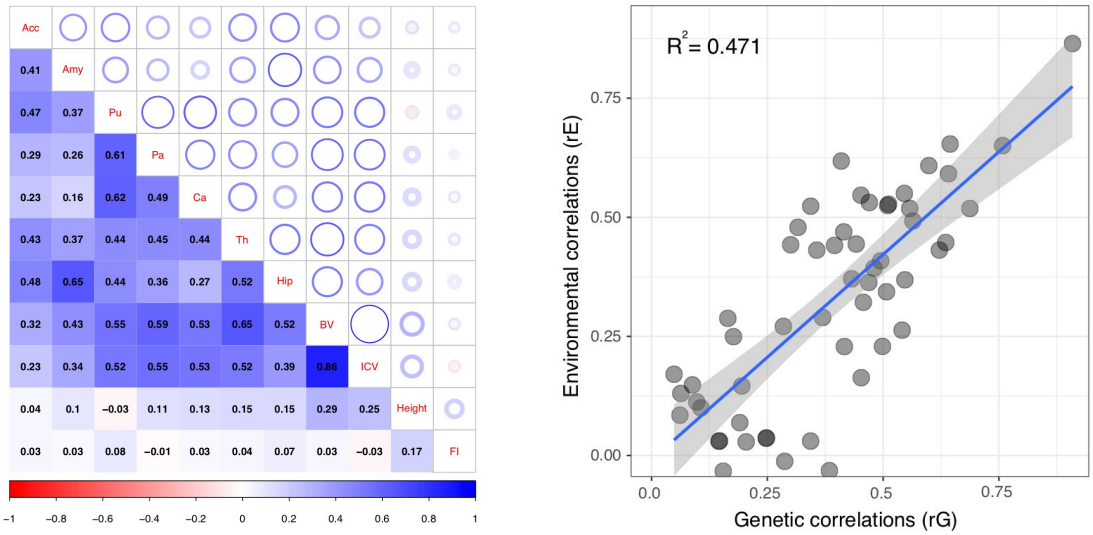


Figure S8. Left: values of environmental correlations are shown in the lower triangular part of the matrix. Circles's diameter is proportional to the strength of the correlation, and a thickness proportional to the size of the confidence interval. Right: scatter plot of environmental versus genetic correlations. Genetic and environmental correlations were adjusted for age, center, sex, and the top 10 principal components of the genetic relationship matrix.

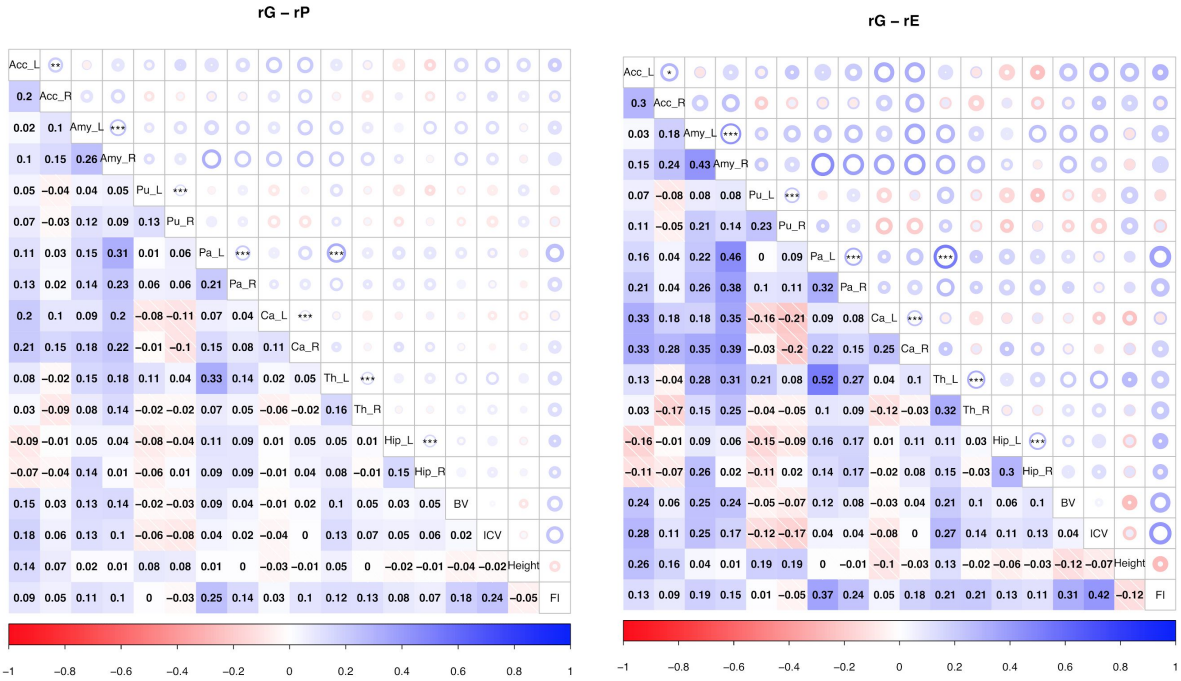


Figure S9. Values of the differences between genetic and phenotypic correlations (r_G - r_P, left) and genetic and environmental correlations (r_G - r_E, right) are shown in the lower triangular part of the matrix. Circles' diameter is proportional to the difference and a thickness proportional to the size of the confidence interval. Stars indicate statistical significance from a *Z*-test testing if the differences are non null (Benjamini Yekutieli FDR adjusted **P*<0.05, ***P*<0.01, ****P*<0.001). The following phenotypes are represented: the left and right hemisphere volumes of accumbens (Acc_L and Acc_R), amygdala (Amy_L and Amy_R), caudate (Cau_L and Cau_R), hippocampus (Hip_L and Hip_R), pallidum (Pa_L and Pa_R), putamen (Pu_L and Pu_R), and thalamus (Th_L and Th_R) ; brain volume (BV), intracranial volume (ICV), height and fluid intelligence (FI) are also represented.

Accounting for the effect of phenotypic measurement error

Let x and y be two measured phenotypes, they can be decomposed into their real values x' and y' and their measurement errors ε_x and ε_y : $x = x' + \varepsilon_x$ and $y = y' + \varepsilon_y$.

The correlation between the two measured phenotypes can be expressed like this:

$$r_{x,y} = \frac{cov(x, y)}{\sqrt{var(x)\ var(y)}} = \frac{cov(x'+\varepsilon_x, y'+\varepsilon_y)}{\sqrt{var(x'+\varepsilon_x)\ var(y'+\varepsilon_y)}} \quad \text{(Eq. 3)}$$

Considering that there is no correlation between measurement errors and the real values, one can develop the following formula:

$$r_{x,y} = \frac{cov(x', y')+cov(\varepsilon_x, \varepsilon_y)}{\sqrt{(var(x')+var(\varepsilon_x))(var(y')+var(\varepsilon_y))}}$$

$$= \frac{cov(x', y')}{\sqrt{var(x')var(y')}} \cdot \frac{cov(x', y')+cov(\varepsilon_x, \varepsilon_y)}{cov(x', y')} \cdot \sqrt{\frac{var(x')}{var(x')+var(\varepsilon_x)}} \cdot \sqrt{\frac{var(y')}{(var(y')+var(\varepsilon_y))}} \quad \text{(Eq. 4)}$$

$\frac{var(x')}{var(x')+var(\varepsilon_x)}$ and $\frac{var(y')}{var(y')+var(\varepsilon_y)}$ can be estimated as being the intraclass correlation coefficient (ICC) respectively for *x* and *y* (called r_x and r_y).

We used a generalised version of the intraclass correlation coefficient to estimate $\frac{cov(x', y')}{cov(x', y')+cov(\varepsilon_x, \varepsilon_y)}$ (called r_c): When *x* and *y* are measured for *N* different subjects with *k* repeated measurements, $r_c = \frac{1}{k(k-1)Nc} \sum_{i=1}^{k} \sum_{j=1, j\neq i}^{k} \sum_{n=1}^{N} (x_{n,i} - \bar{x})(y_{n,j} - \bar{y})$ where $\bar{x}$ and $\bar{y}$ are the respective means for all *x* and *y* measurements, and $c = \frac{1}{kN} \sum_{i=1}^{k} \sum_{n=1}^{N} (x_{n,i} - \bar{x})(y_{n,i} - \bar{y})$.

We were thus able to obtain an estimation of the phenotypic correlation adjusted for measurement errors:

$$r_{x', y'} = r_{x, y} \frac{r_c}{\sqrt{r_x\ r_y}} \quad \text{(Eq. 5).}$$

Values of $r_{x', y'}$ are available in the second sheet of Table S5, values of r_x , r_y , and r_c are available in the third sheet of Table S5. The estimates of V_G/V_P adjusted for measurement errors using the ICC values of each phenotype are shown in Figure S10.

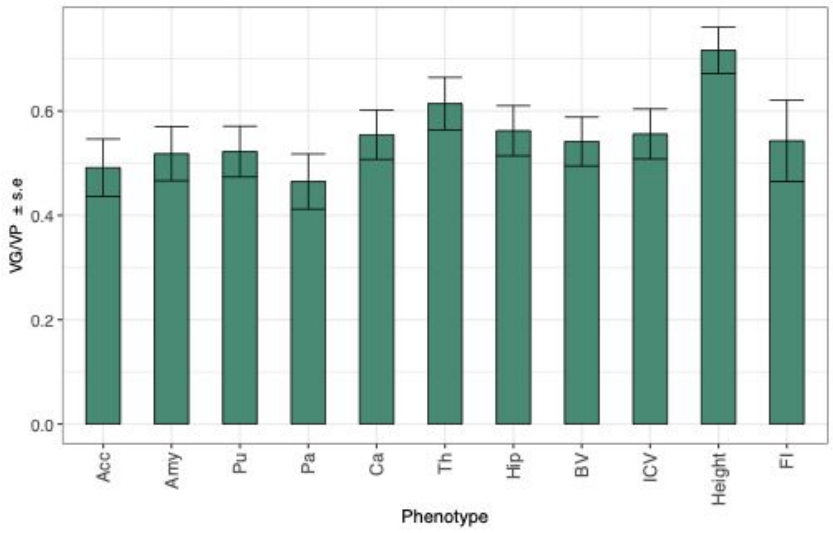


Figure S10. Adjusted proportion of variance captured by common genotyped variants (V_G/V_P) for brain regions, height and intelligence score.

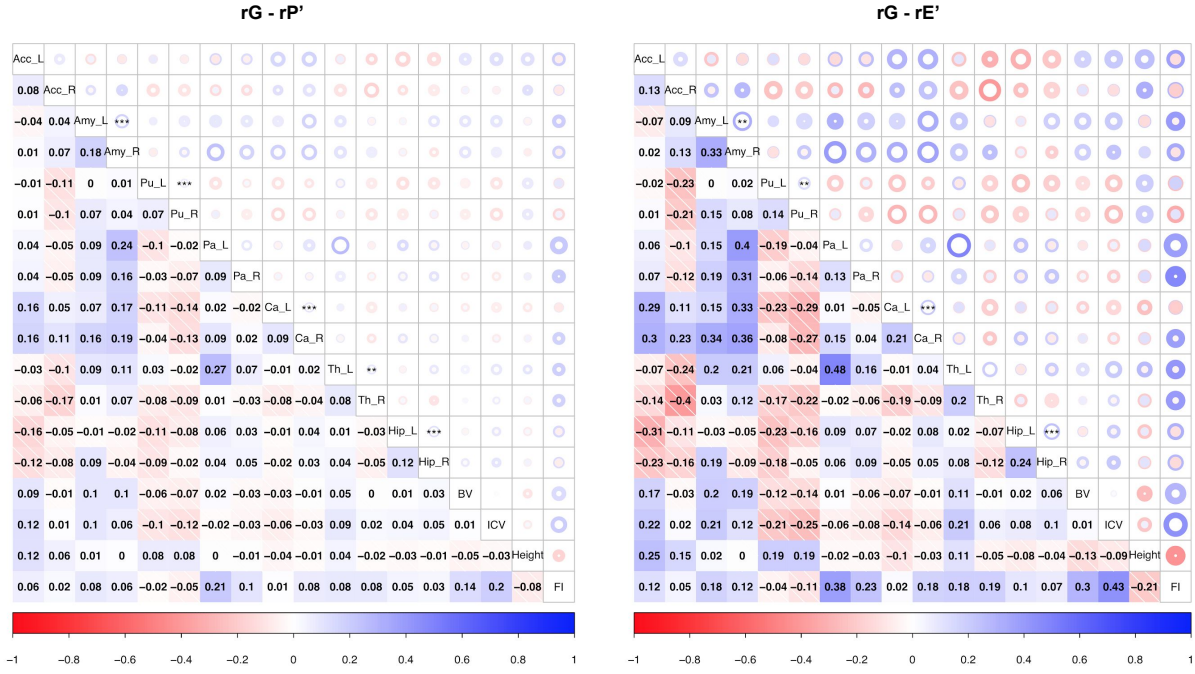


Figure S11. Values of the differences between genetic and corrected phenotypic correlations (r_G - r_P', left panel), and between genetic and corrected environmental correlations (r_G - r_E', right panel) are shown in the lower triangular part of the matrix. Circles' diameter is proportional to the difference and a thickness proportional to the size of the confidence interval. Stars indicate statistical significance from a Z-test testing if the differences are non null (Benjamini Yekutieli FDR adjusted *P<0.05, **P<0.01, ***P<0.001). The following phenotypes are represented: the left and right hemisphere volumes of accumbens (Acc_L and Acc_R), amygdala (Amy_L and Amy_R), caudate (Cau_L and Cau_R), hippocampus (Hip_L and Hip_R), pallidum (Pa_L and Pa_R), putamen (Pu_L and Pu_R), and thalamus (Th_L and Th_R) ; brain volume (BV), intracranial volume (ICV), height and fluid intelligence (FI) are also represented.

6. Distribution of V_G/V_P estimates for simulated phenotypes for different sample sizes

We studied how the distribution of heritability estimates differed from their theoretical distribution as a function of sample size. V_G/V_P estimates were obtained from 1,000 simulated phenotypes with a heritability of V_G/V_P = 50% for sub-samples of N = 800, N = 400, N = 200 and N = 100 subjects.
Figure S12 show simulations based on the ADNI project and Figure S13 show simulations based on the UK Biobank project.

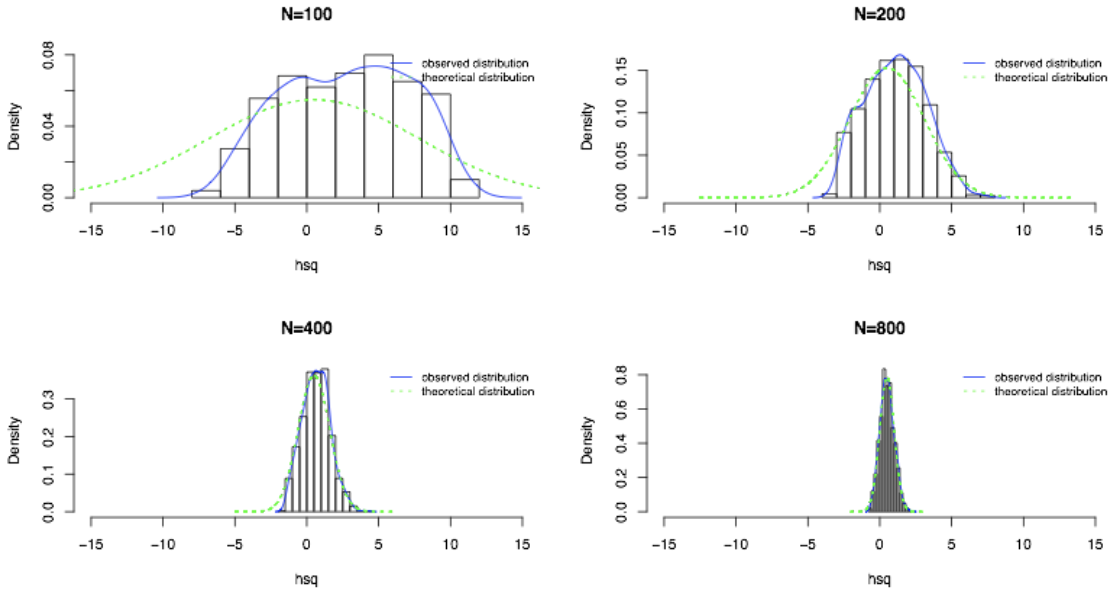


Figure S12. Distribution of estimated values for V_G/V_P of simulated phenotypes for different sub-sample sizes in ADNI dataset, compared with the expected (theoretical) distribution.

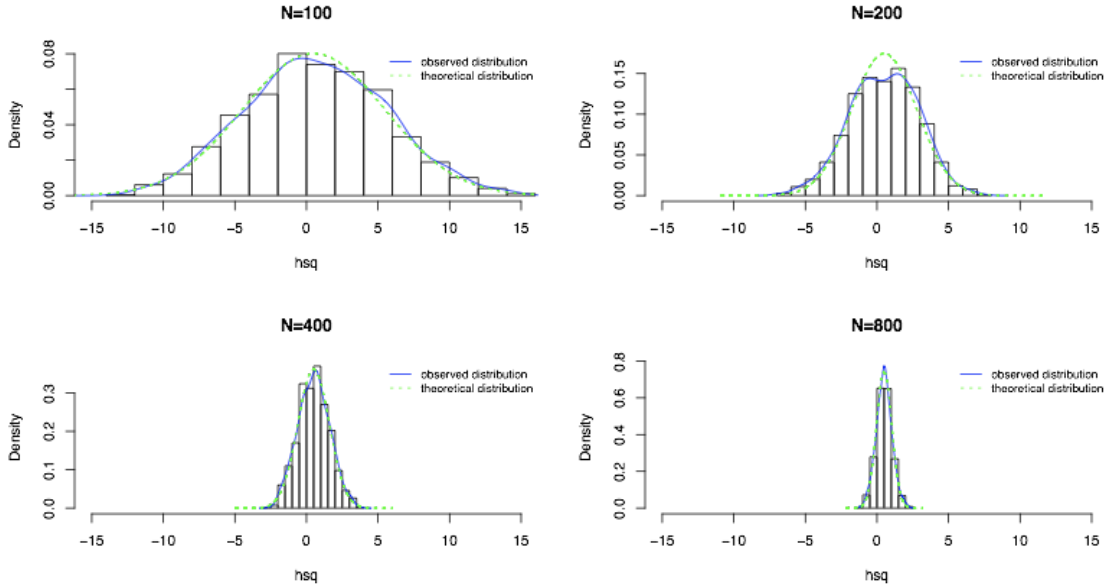


Figure S13. Distribution of estimated values for V_G/V_P of simulated phenotypes for different sub-sample sizes in UK Biobank dataset, compared with the expected (theoretical) distribution.

7. Comparison between meta-analysis and mega-analysis

Fig. S14 shows the difference in standard errors obtained when a single group is analysed (labelled as "merging") versus a meta-analytic approach combining several small groups (labelled as "meta-analysis"). The plot represents the change in standard error after including an additional dataset for GREML analysis.

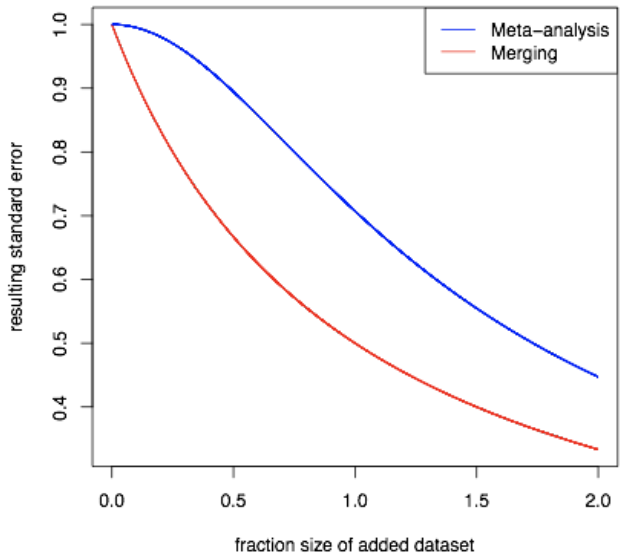


Figure S14. Resulting standard error after addition of another dataset for GREML analysis. The x-axis represents relative size of the added dataset compared with the initial dataset. The y-axis represents relative standard error of the combinated estimate compared with the standard error obtained only with the first dataset. Meta-analysis: resulting standard error computed as $\frac{1}{\sqrt{1+fraction^2}}$. Merging: resulting standard error computed as $\frac{1}{1+fraction}$.

8. Genome-wide polygenic scores

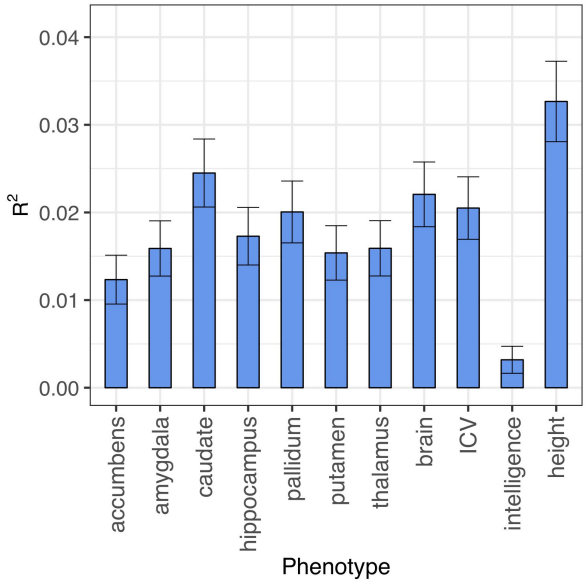


Figure S15. Variance captured by genome-wide polygenic scores (coefficient of determination R^2) for each phenotype. The analyses were performed on the residuals of the linear regression including the effects of age, sex, imaging centre, and the 10 first principal components of the GRM as covariates. Error bars represent the 95% confidence intervals.

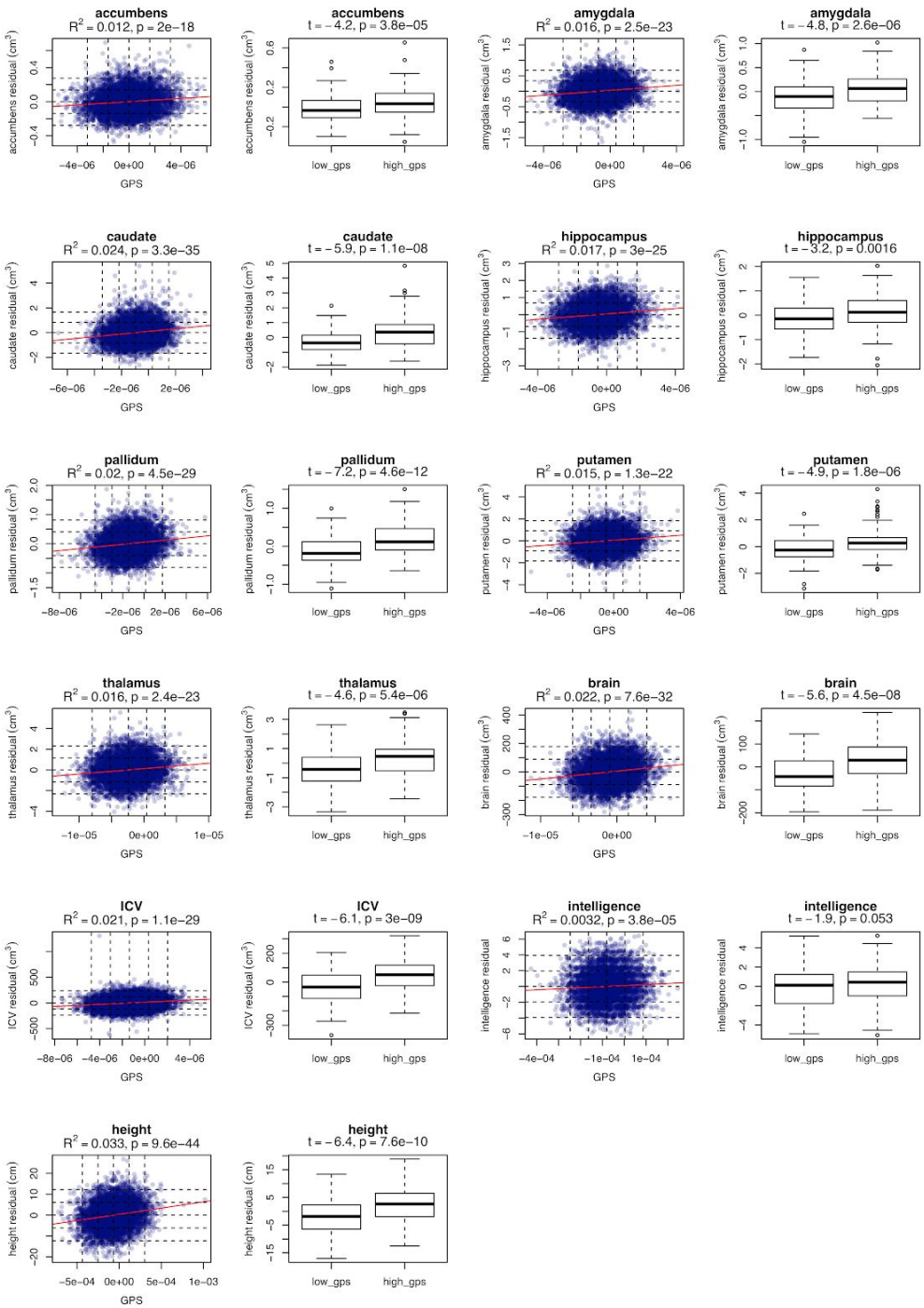


Figure S16. For each phenotype**,** two plots are shown: a scatter plot of the phenotype residuals (after covarying by age, sex, imaging centre and genetic principal components) as a function of the genome-wide polygenic scores (GPS), and boxplots showing the distribution of the phenotype residuals of the individuals located more than minus (low_gps) or plus (high_gps) two standard deviations away from the average GPS. Units are cm^3 for volumes and cm for height. The variance captured by genome-wide polygenic scores (coefficient of determination R^2) and the p-value of the Fisher test testing whether R^2 is different from zero are given for the scatter plots. The t-value and p-value of the Student test comparing the individuals with low and high GPS are given for the boxplots.

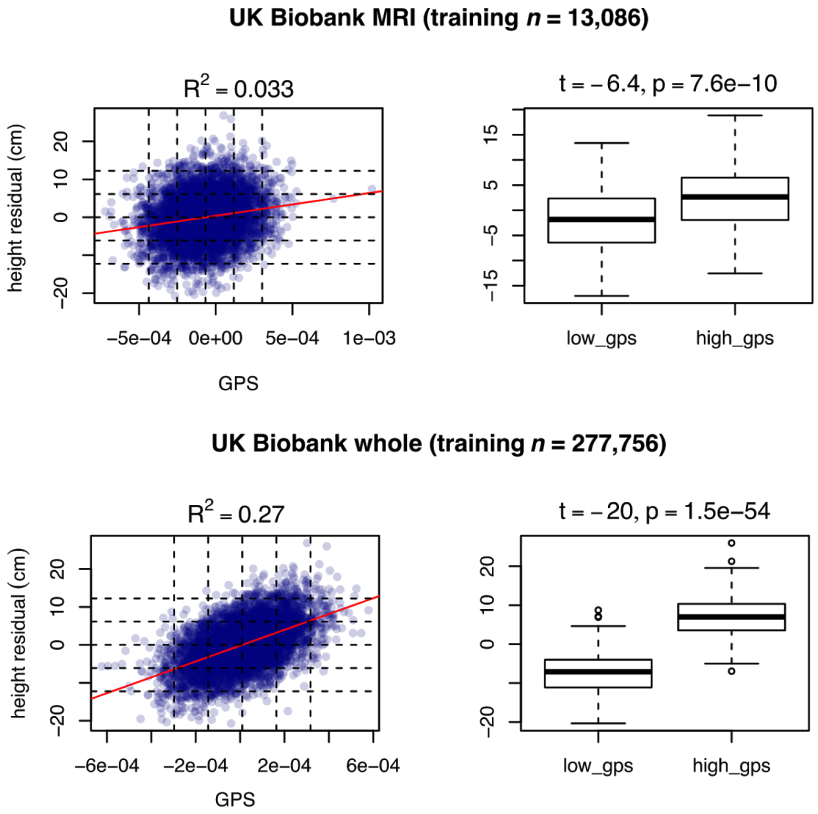


Figure S17. Comparison of the prediction of height with genome-wide polygenic scores using as base dataset the MRI cohort of UK Biobank (n = 13,086, top panel) and the whole UK Biobank cohort without subjects included in the target dataset (n = 277,756, bottom panel). The additional MRI cohort of UK Biobank (n = 6,189) was used as target dataset.

For each base dataset**,** two plots are shown: a scatter plot of the phenotype residuals (after covarying by age, sex, imaging centre and genetic principal components) as a function of the genome-wide polygenic scores (GPS), and boxplots showing the distribution of the phenotype residuals of the individuals located more than minus (low_gps) or plus (high_gps) two standard deviations away from the average GPS. Units are cm^3 for volumes and cm for height. The variance captured by genome-wide polygenic scores (coefficient of determination R^2) are given for the scatter plots. The t-value and p-value of the Student test comparing the individuals with low and high GPS are given for the boxplots.

9. Supplementary Table Legends

Table S1: SNP-heritability results.
Sheet1: SNP-heritability computed with the GCTA GREML method (including age, sex, centre and the top 10 PCs of the GRM as covariates) for each of the studied phenotypes in UK Biobank and five ENIGMA cohorts.
Sheet2: SNP-heritability estimates obtained by averaging across the six cohorts using the inverse variance-weighted average method.

Table S2: SNP-heritability computed with GCTA GREML for UK Biobank (including age, sex, centre), with (Table S2.1) and without (Table S2.2) including the top 10 principal components of the GRM matrix as covariates.

Table S3: SNP-heritability and enrichment results for several partitions using the UK Biobank project data.
Table S3.1 provides data on four partitions based on genic status. The first set includes all SNPs within the 66,632 RefSeq gene boundaries of the hg19 assembly ("genic-margin0"), the two further sets include also SNPs 0 to 20 kbp ("updown-margin20") and 20kbp to 50 kbp upstream and downstream of each gene ("updown-margin20-50"), and a last set includes the SNPs not located in regions less than 50 kbp upstream or downstream of genes (nongenic-margin50).
Table S3.2 provides data on four partitions based on minimum allele frequency (MAF): from 0.1 to 5%, from 5 to 20%, from 20 to 35% and from 35 to 50%.
Table S3.3 provides data on three partitions (genic in gene set region, genic not in gene set region, non-genic) based on preferential central nervous system (CNS) gene expression (Raychaudhuri et al. 2010; Lee, DeCandia, et al. 2012) using ± 50 kbp as gene boundaries.
Table S3.4 provides data on three partitions (genic in gene set region, genic not in gene set region, non-genic) based on markers of brain cell types from the scRNA-Seq study of Skene et al. (2018) (PMID: 29785013). We used the specificity metric, provided by the authors in R datafiles at http://www.hjerling-leffler-lab.org/data/scz_singlecell/, for each gene across 24 brain cell types**.** In order to obtain a partition with a sufficient number of SNPs, we built a gene set (using ± 20 kbp as gene boundaries) of brain cell types markers by taking the union of the genes with the top 1% specificity scores for each cell type.
Table S3.5 provides data on three partitions (genic in gene set region, genic not in gene set region, non-genic) based on markers of brain cell types from the scRNA-Seq study of Li et al. (2018) (PMID: 30545854). The union of the genes in the Table S6 of the manuscript were used to build the gene set (using ± 20 kbp as gene boundaries).
Table S3.6 provides data on three partitions (genic in gene set region, genic not in gene set region, non-genic) based on markers of different time periods of human brain development from the scRNA-Seq study of Li et al. (2018) (PMID: 30545854). The union of the genes in the Table S7 of the manuscript were used to build the gene set (using ± 20 kbp as gene boundaries).
For each partition i, enrichment was computed using a Z-score comparing the estimated genetic variance VGi of the SNP set i to the SNP-set genetic variance fi*VGi expected under no enrichment, the corresponding enrichment p-value comparing this Z-score to zero with a Z-test, the number of individuals, the p-value of the likelihood ratio test output by GCTA.

Table S4. Meta-Analysis of the genetic partitioning of SNP-heritability across the six datasets.
The partitioning of SNP-heritability estimates were averaged across the six datasets using the inverse variance-weighted average method.
Table S4.1 provides four partitions based on genic status.
Table S4.2 provides four partitions based on minimum allele frequency (MAF).
Table S4.3 provides two partitions based on preferential central nervous system (CNS) gene expression (Raychaudhuri et al. 2010; Lee, DeCandia, et al. 2012) using ± 50 kbp as gene boundaries.
For each partition i, the following results are provided: f_i the average fraction of SNPs belonging to the SNP set i, the average values of V_{Gi}/V_P and $V_{Gi}/V_G/f_i$ and the associated standard errors, the enrichment p-value of the Z-test comparing $V_{Gi}/V_G/f_i$ to 1.

Table S5: Genetic, phenotypic, and environmental correlations in UK Biobank.
Table S5.1 provides the values of the genetic, phenotypic, and environmental correlations, together with their standard errors. Differences between r_G and r_E (column dr_r_Gr_E), and r_G and r_P (column dr_r_Gr_P) were computed and their associated standard errors (column dr_r_Gr_E_SE and column dr_r_Gr_P_SE) was estimated using the delta method using the output of the GCTA REML bivariate method. A Z-test was used to compare these differences to zero, the p-values are available in columns p.dr_r_Gr_E and p.dr_r_Gr_P. The column nind contains the number of individuals available with both annotated phenotypes.
Table S5.2 provides the same values as for Table S5.1 but for phenotypic and environmental correlations values that were corrected for measurement error (as described in the 5. of this supplementary document).
Table S5.3 provides the values used to correct the phenotypic correlations for measurement errors (see equation Eq. 5 of this supplementary document).

Table S6: Distribution of estimated values for V_G/V_P of simulated phenotypes for different sub-sample sizes in ADNI dataset. For four sample sizes, 1000 sub-sampling were performed without replacement. nb_ind: number of sub-sampled individuals, nb_conv: number of times GREML converged and an estimate was given, th_mean: theoretical mean for estimates (equal to the simulated V_G/V_P), obs_mean, obs_mean_se: observed mean of estimates and standard error, p_mean: p value for difference between theoretical and observed mean (z-test), th_sd: theoretical standard deviation of estimates (computed as the mean of standard errors reported by GCTA), obs_sd, obs_sd_se: observed standard deviation of estimates and standard error, p_sd: p value for difference between theoretical and observed standard deviation (z-test), D_ks, p_ks: effect size and p-value of Kolmogorov-Smirnov test of concordance between theoretical normal distribution and observed distribution.

Table S7: Distribution of estimated values for V_G/V_P of simulated phenotypes for different sub-sample sizes in UK Biobank dataset.

Table S8: Proportion of captured phenotypic variance by genome-wide polygenic scores and covariates. mean_pheno: mean value of the raw phenotype, sd_pheno: standard deviation of the raw phenotype, Rsq_age_only: proportion of the raw phenotype captured by age, Rsq_pca_only: proportion of the raw phenotype captured by genetic principal components, Rsq_sex_only: proportion of the raw phenotype captured by sex, Rsq_centre_only: proportion of the raw phenotype captured by imaging centre, Rsq_gps_only: proportion of the raw phenotype captured by genome-wide polygenic score, sd_resid: standard deviation of residual phenotype after covarying by age, pca, sex and centre, Rsq: proportion of residual phenotype captured by polygenic score, SErsq: standard error of R^2, LCL: lower 95% confidence interval limit, UCL: upper 95% confidence interval limit.