Multivariate analysis reveals shared genetic architecture of brain morphology and human behavior

R. de Vlaming[1,*], Eric A.W. Slob[2,3,4,*], Philip R. Jansen[5,6], Alain Dagher[7], Philipp D. Koellinger[1,8], Patrick J.F. Groenen[9], Cornelius A. Rietveld[2,3,**]

[1] School of Business and Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2] Department of Applied Economics, Erasmus School of Economics, Rotterdam, The Netherlands

[3] Erasmus University Rotterdam Institute for Behavior and Biology, Erasmus School of Economics, Rotterdam, The Netherlands

[4] MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge, Cambridge, UK

[5] Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[6] Department of Clinical Genetics, VU Medical Center, Amsterdam UMC, Amsterdam, The Netherlands

[7] Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada

[8] La Follette School of Public Affairs, University of Wisconsin-Madison, WI, USA

[9] Econometric Institute, Erasmus School of Economics, Rotterdam, The Netherlands

[*]These authors contributed equally

[**]Corresponding author: Cornelius A. Rietveld, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 2061 PA, Rotterdam, The Netherlands, Phone: +31(0)10-408-1401, E-mail: nrietveld@ese.eur.nl

1

Abstract

Human variation in brain morphology and behavior are related and highly heritable. Yet, it is largely unknown to what extent specific features of brain morphology and behavior are genetically related. Here, we introduce multivariate genomic-relatedness restricted maximum likelihood (MGREML) and provide estimates of the heritability of grey-matter volume in 74 regions of interest (ROIs) in the brain. We map genetic correlations between these ROIs and health-relevant behavioral outcomes including intelligence. We find four genetically distinct clusters in the brain that are aligned with standard anatomical subdivision in neuroscience. Behavioral traits have distinct genetic correlations with brain morphology which suggests trait-specific relevance of ROIs.

Introduction

Global and regional grey matter volumes are known to be linked to differences in human behavior and mental health[1]. For example, reduced grey matter density has been implicated in a wide range of neurodegenerative diseases and mental illnesses[2,3,4,5]. In addition, differences in grey matter volume have been related to cognitive and behavioral phenotypic traits such as fluid intelligence and personality, although results have not always been replicable[6,7].

Variation in brain morphology can be measured non-invasively using magnetic resonance imaging (MRI). Large-scale data collection efforts such as the UK Biobank that included both MRI scans and genetic data enabled recent studies to discover the genetic architecture of human variation in brain morphology and to explore the genetic correlations of brain morphology with behavior and health[8,9,10,11,12]. These studies have demonstrated that all features of brain morphology are genetically highly complex traits and their heritable component is mostly due to the combined influence of many common genetic variants, each with a small effect.

A corollary of this insight is that even the largest currently possible genome-wide association studies (GWASs) identify only a small part of the genetic variants underlying the heritable components of brain morphology: the vast majority of their heritability remains missing[8,9,10,11,12,13]. As a consequence, the genetic correlations of regional brain volumes with each other as well as with human behavior and health have remained largely elusive. However, such estimates could advance our understanding of the genetic architecture of the brain for example regarding its structure and plasticity. Similarly, a strong genetic overlap of specific features of brain morphology with mental health would provide clues about the neural mechanisms behind the genesis of disease[14,15,16].

2

We developed multivariate genome-based restricted maximum likelihood method (MGREML) to provide a comprehensive map of the genetic architecture of brain morphology. MGREML overcomes several limitations of existing approaches to estimate heritability and genetic correlations from molecular genetic data. In contrast to existing pairwise bivariate approaches, MGREML guarantees internally consistent (i.e., semi-positive definite) genetic correlation matrices and it yields standard errors that reflect the multivariate structure of the data correctly. The software implementation of MGREML is computationally substantially more efficient than traditional bivariate GREML[17,18]. Moreover, we show that MGREML allows for stronger statistical inference than methods that are based on GWAS summary statistics such as bivariate LD score regression (LDSC)[19,20]. In sum, MGREML yields precise estimates of genetic correlations across a large number of traits when existing approaches applied to the same data are either inaccurate, computationally unfeasible, or underpowered.

We leverage the advantages of MGREML by analyzing brain morphology based on MRI-derived grey-matter volumes in 74 regions of interest (ROIs). We also estimate the genetic correlations of these ROIs with global measures of brain volume and eight human behavioral traits that have well-known associations with mental and physical health. The anthropometric measures height and body-mass index (BMI) are also analyzed, because of their relationships with brain size[6,12]. Our analyses are based on data from the UK Biobank brain imaging study[21].

Results

Estimating genetic correlations

Several methods allow the estimation of heritability and genetic correlations from molecular genetic data. One class of these methods is based on GWAS summary statistics[19,20,22]. Another class of methods is based on individual-level data, such as genome-based restricted maximum likelihood (GREML) and variations of this approach[23,24,25,26,27,28]. Methods based on GWAS summary statistics such as LDSC[19,20] and variants thereof[29] can leverage the ever increasing sample sizes of GWAS meta- or mega-analyses[30] and they are computationally efficient once GWAS results have been obtained. These methods benefit from the fact that GWAS summary statistics are often publicly shared[31,32]. However, the computationally more intensive methods based on individual-level data such as GREML are statistically more powerful[33].

Where GWAS meta-analysis sample sizes for genetically complex traits such as height[34] or educational attainment[35] currently exceed 1 million, most datasets contributing to such a meta-analysis are considerably smaller, which put a constraint on the statistical inferences one can

3

obtain using methods based on individual-level data. Due to the high costs of MRI brain scans, GWAS samples for brain imaging genetics are still relatively small compared to GWAS samples for traits that can be measured at low cost (e.g., height and BMI). The UK Biobank brain imaging study (**Online Methods**) is currently by far the largest available sample that includes both MRI scans and genetic data, often surpassing the sample size of most previous studies in neuroscience by an order of magnitude or more[8,9,12]. This UK Biobank data led to recent breakthroughs in imaging genetics, however even the sample size of the currently largest GWAS of brain morphology is too small to yield precise estimates of genetic correlations with LDSC[10,11].

Irrespective of whether one uses GWAS summary statistics or individual-level data, the use of bivariate methods poses another challenge when computing genetic correlation across more than two traits. When genetic correlations are estimated across more than two traits, the correlation estimates of pairwise combinations of traits are often aggregated into >2 by >2 genetic correlation matrices[19,20,36]. However, this 'pairwise bivariate' approach can result in genetic correlation matrices that are not internally consistent (i.e., they describe interrelationships across traits that cannot exist simultaneously). In mathematical terms, the resulting matrices are not positive semi-definite. Although the correlation between two traits can vary between −1 and +1, their correlation with a third trait is naturally bounded. For a set of three traits, positive semi-definiteness holds when the correlation coefficients satisfy the condition $r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1$. This condition does for instance not hold when pairwise correlations are estimated to be $r_{12} = 0.9$, $r_{13} = 0.9$, and $r_{23} = 0.2$. For example, the genetic correlation matrix in the well-known atlas of genetic correlations is not positive semi-definite[19]. A second consequence of the pairwise bivariate approach is that the standard errors of the resulting genetic correlation matrix do not adequately reflect the multivariate structure of the data.

MGREML

Our multivariate extension of the GREML method[17,27] guarantees the internal consistency of the estimated genetic correlation matrix by adopting a factor model for the covariance matrices (**Supplementary Information S1**). This parametrization also ensures that the standard errors of the estimated genetic correlations reflect the multivariate structure of the data correctly. Therefore, methods such as genomic-SEM[37] that use multivariate genetic correlation matrices as input information may benefit from using results obtained with MGREML. To deal with the computational burden and to make MGREML applicable to large data sets in terms of

individuals and traits, we derived efficient expressions for the likelihood function and developed a new optimization algorithm (**Supplementary Information S1**). Runtime analyses described in **Supplementary Information S3** show that MGREML is computationally faster than pairwise bivariate GREML. Increasing the number of individuals N in the analyses increases the running time of both pairwise bivariate GREML and MGREML at approximately the same rate, but, when increasing the number of traits T, the running time of pairwise bivariate GREML increases disproportionally as compared to MGREML. Moreover, comparison of results obtained with MGREML with results obtained using LDSC shows that standard errors obtained with MGREML are 38.9%-46.0% smaller, illustrating the substantial gains in statistical power afforded by MGREML.

Analysis of brain morphology

We used MGREML to analyze the heritability of and genetic correlations across 86 traits in 20,190 unrelated 'white British' individuals from the UK Biobank (**Fig. 1**, **Online Methods**). The subset of 76 brain morphology traits includes total brain volume (grey and white matter), total grey matter volume, and grey matter volumes in 74 regions of interest (ROIs) in the brain. Relative volumes were obtained by dividing ROI grey matter volumes by total grey matter volume. The full set of heritability estimates is available in **Extended Data Table 1**. **Fig. 2a** and **Fig. 2b** show that SNP-based heritability (h_{SNP}^2) (i.e., the proportion of phenotypic variance which can be explained by autosomal SNPs) is on average highest in the insula, and in the cerebellar and subcortical structures of the brain (average h_{SNP}^2 is 33.1%, 32.4%, and 29.5%, respectively) and lowest in the parietal, frontal and temporal lobes of the cortex (average h_{SNPS}^2 is 21.2%, 21.4%, and 25.2%, respectively). Grouping of the h_{SNPS}^2 estimates in networks of intrinsic functional connectivity[38] reveals that ROIs in the heteromodal cortex (frontoparietal, dorsal attention) are less heritable than primary (visual, somatomotor), subcortical and cerebellar regions (**Fig. 3a**).

The full set of estimated genetic correlations is available in **Extended Data Table 1**. Using spatial mapping, **Fig. 2c** visualizes the estimated genetic correlations across the relative volumes of the cortical and subcortical brain areas. The largest positive genetic correlations were found between the insular and frontal regions (average $r_g = 0.17$) and between the cerebellar and subcortical areas (average $r_g = 0.15$). The largest negative correlations were present between the cerebellar and insular regions (average $r_g = -0.18$) and between the cerebellar and frontal regions (average $r_g = -0.15$) (**Fig. 2d**). **Fig. 3b** shows that the genetic

5

correlations are particularly strong within intrinsic connectivity networks, especially the visual, somatomotor, subcortical, and cerebellum networks, possibly because of lower experience-dependent plasticity in these brain regions compared to heteromodal and associative areas[39]. Using Ward's method for hierarchical clustering[40], we identify four clusters within the estimated genetic correlations for the 74 ROIs in the brain (**Fig. 4**). The first cluster (*n*=18) includes most of the frontal cortical areas of the brain, the second (*n*=18) the cerebellar cortex, the third (*n*=18) subcortical structures including the brain stem, and the last cluster (*n*=20) contains a mixture of temporal and occipital brain areas.

We also used MGREML to estimate the genetic correlations between brain morphology and eight human behavioral traits that are known to be related to health and that have previously been studied in large-scale GWASs, as well as the anthropometric measures height and BMI. Statistically significant correlations are highlighted in **Extended Data Table 1** (Panel C). Spatial maps of the genetic correlation between brain morphology and the behavioral traits are shown in **Fig. 5**. For subjective well-being, we find the strongest genetic correlation with the Middle Frontal Gyrus (**Fig. 5a**, $r_g = 0.21$), a region that has been linked before to emotion regulation[41]. The genetic correlations of the ROIs with neuroticism (**Fig. 5b**) and depression (**Fig. 5c**) are generally weak and insignificant, perhaps reflecting the coarseness of these phenotypic measures in the UK Biobank data. The strongest genetic correlation with the number of alcoholic drinks consumed per week is with the Lateral Occipital Cortex, superior and inferior divisions (**Fig. 5d**, $r_g = 0.23$ and $r_g = 0.18$, respectively). Although the phenotypic correlations between the analyzed ROIs and alcohol consumption are generally negative[42], these particular brain regions are among those implicated in the affective response to drug cues based on the perception-valuation-action model[43]. For educational attainment and intelligence, the strongest correlations are found in the frontal lobe region ($r_g = -0.13$ between educational attainment and the Superior Frontal Gyrus, and $r_g = 0.16$ between intelligence and the Frontal Medial Cortex). **Fig. 5e** and **Fig. 5f** show that the genetic correlation structures estimated for educational attainment and intelligence are largely similar, in line with earlier studies showing the strong genetic overlap between these two traits[44]. Genetic correlations of the ROIs with visual memory (**Fig. 5g**) are insignificant, and the strongest genetic correlation of reaction time is with the Middle Temporal Gyrus, temporooccipital part (**Fig. 5h**, $r_g = 0.20$). Activity within the middle temporal gyrus has been linked before with reaction time[45].

Earlier studies suggest that the size of the brain is positively associated with traits such as intelligence[6], and our use of relative measures may have attenuated the estimation of these

6

relationships in our analyses. When analyzing absolute brain volumes of the ROIs rather than relative brain volumes (i.e., relative to total grey matter volume in the brain), we indeed observe robust positive relationships between the absolute volumes of the ROIs on the one hand and height and intelligence on the other hand (**Extended Data Table 2**). The main differences we observe in the set of estimated correlations across the ROIs are that the genetic correlations within the cerebellum clusters are slightly smaller and that the positive correlations within the subcortical structures are somewhat larger.

Discussion

We designed MGREML to estimate high dimensional genetic correlation matrices from large-scale individual-level genetic data in a computationally efficient manner while guaranteeing the internal consistency of the estimated genetic correlation matrix. For comparison, we used GWAS and bivariate LDSC[20] to obtain a genetic correlation matrix with the pairwise bivariate approach using the exact same set of individuals (N=20,190) and traits (T=86) as in our main analysis (**Extended Data Table 3**). The correlation between the heritability estimates obtained by MGREML and LDSC is 0.95, but the 95% confidence interval of h^2 estimates obtained with MGREML are on average 38.9% smaller than those obtained from LDSC. The 95% confidence intervals of the genetic correlations obtained using MGREML are on average 46.0% smaller compared to those obtained with LDSC, illustrating the advantages of MGREML in terms of statistical power (1,519 versus 1,044 significant correlations at the 5% level). This gain in statistical efficiency is slightly larger than the efficiency gain a recently developed variant of bivariate LDSC was able to achieve[29]. Importantly, the genetic correlation matrix obtained using bivariate LDSC is not positive semi-definite and thus the estimated genetic correlations across traits are not internally coherent. The use of this correlation matrix would pose challenges for multivariate methods such as Genomic-SEM[37] that are currently based on genetic correlation matrices estimated by LDSC.

Our results show marked variation in the estimated heritability across cortical grey matter volumes, with on average higher heritability estimates in subcortical and cerebellar areas than in cortical areas (**Fig. 2b**). Grouping of by h^2_{SNPs} estimates by networks of intrinsic functional connectivity suggests that heritability is particularly low in brain areas with presumed stronger experience-dependent plasticity (**Fig. 3a**). These results suggest that neocortical areas of the brain are under weaker genetic control perhaps reflecting greater environmentally-determined plasticity[39,46]. Furthermore, the estimated genetic correlations suggest the presence of four genetically distinct clusters in the brain (**Fig. 4**). These clusters largely correspond with the

7

conventional subdivision of the brain in different lobes based on anatomical borders[47]. The estimated genetic correlations also provide evidence for a shared genetic architecture of traits between which an association has been observed before in phenotypic studies such as between intelligence and educational attainment[44]. In addition, new genetic correlations were identified between alcohol consumption and cerebellar volume, and between subjective well-being and the temporo-occipital part of the Middle Temporal Gyrus (**Extended Data Table 1**).

To verify that our results are not merely a reflection of the physical proximity of brain regions, we regressed the estimated genetic correlations on the physical distance between the different brain regions. Although this correction procedure decreased the estimated genetic correlations by approximately 25%, the main patterns are still observed. For the same reason, we recreated the dendogram (**Fig. 3**) after aggregating the results for sub regions into an average for the larger region because the optimization procedure of MGREML puts equal weight on each trait and does not account for physical proximity. The results of this robustness check show that the four identified clusters do not merely reflect the number of analyzed measures for a specific brain region.

Estimates of heritability increase our understanding of the relative impact of genetic and environmental variation on traits[13,27], and estimates of genetic correlation lead to a better understanding of the shared biological pathways between traits[48]. MGREML has been designed to estimate both SNP-based heritability and genetic correlations in a computational efficient and internally consistent manner using individual-level genetic data. Its newly developed optimization algorithm makes it possible to apply MGREML to estimate high dimensional genetic correlation matrices in large datasets such as the UK Biobank.

Data and code availability

Individual-level genotype and phenotype data are available by application from the UKB Biobank (https://www.ukbiobank.ac.uk/). MGREML is available on https://github.com/devlaming/mgreml as a ready-to-use command-line tool. The GitHub page comes with a full tutorial on the usage of this tool. A MGREML analysis of 86 traits, observed in a sample of 20,190 unrelated individuals (i.e., the dataset we exploit in our empirical application), takes around four hours on a four-core laptop with 16GB of RAM.

Acknowledgements

Online Methods

Statistical framework

In a genome-wide association study (GWAS) of a quantitative trait y, the effect of SNP m on is modelled as:

$$y_j = g_{jm}\alpha_m + \mathbf{x}_j'\boldsymbol{\beta} + \varepsilon_j, \tag{1}$$

where y_j is the phenotypic value for individual j, and g_{jm} is the genotype for individual j for SNP m (i.e., a value equal to zero, one, or two, indicating the number of copies of the coded allele the individual carries) with effect α_m on the phenotype, $\mathbf{x}_j'$ is a $1 \times k$ vector of control variables with effects $\boldsymbol{\beta}$, and ε_j is a normally distributed error term. Instead of one equation per SNP, we can also consider the contribution of all SNPs jointly:

$$y_j = \mathbf{g}_j'\boldsymbol{\alpha} + \mathbf{x}_j'\boldsymbol{\beta} + \varepsilon_j, \text{ with } \mathbf{g}_j'\boldsymbol{\alpha} = \sum_{m=1}^{M} g_{jm}\alpha_m. \tag{2}$$

Here, $\mathbf{g}_j'$ is the $1 \times M$ vector of genotypes for individual j with standardized effects $\boldsymbol{\alpha}$. Assuming the phenotype is either mean-centered and/or an intercept is included in the set of control variables, we can assume, without loss of generality, that SNPs are standardized in accordance with their distribution under Hardy–Weinberg equilibrium. Equation 2 can then be rewritten in matrix notation for a sample comprising N individuals as:

$$\mathbf{y} = \mathbf{G}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{3}$$

where $\mathbf{G}$ is the $N \times M$ matrix of standardized genotypes with effects $\boldsymbol{\alpha}$, $\mathbf{X}$ is the $N \times k$ matrix of control variables with effects $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$ is the error term. Following Yang et al. (2010) in their original development of genome-based genomic-relatedness restricted maximum likelihood (GREML)[27], we assume $\boldsymbol{\beta}$ to be fixed, $\boldsymbol{\alpha} \sim N(\mathbf{0}, \mathbf{I}_M \sigma_\alpha^2)$ and $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}_N \sigma_E^2)$. Here, σ_α^2 is the variance in the effects of the genetic variants and σ_E^2 is the environmental variance. In the resulting linear mixed model (LMM), the total genetic contribution $\mathbf{G}\boldsymbol{\alpha} \sim N(0, \mathbf{G}\mathbf{G}'\sigma_\alpha^2)$. Moreover, the phenotypic variance-covariance matrix across individuals can be decomposed as:

$$\text{Var}(\mathbf{y}) = \mathbf{G}\mathbf{G}'\sigma_\alpha^2 + \mathbf{I}_N \sigma_E^2 = \left(\frac{1}{M}\mathbf{G}\mathbf{G}'\right)(M\sigma_\alpha^2) + \mathbf{I}_N \sigma_E^2 = \mathbf{A}\sigma_G^2 + \mathbf{I}_N \sigma_E^2, \tag{4}$$

where $\mathbf{A} = M^{-1}\mathbf{G}\mathbf{G}'$ is the genomic-relatedness matrix (GRM) capturing genetic similarity between individuals based on all SNPs. In Equation 4, $\sigma_G^2 (= M\sigma_\alpha^2)$ can be interpreted as the

10

contribution of additive genetic effects to the phenotypic variance. Therefore, the SNP-based heritability of the trait of interest y can be defined as:

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_E^2}. \tag{5}$$

By combining Equations 3 and 4, we can write the LMM as:

$$\mathbf{y} \sim N(\mathbf{X\beta}, \sigma_G^2 \mathbf{A} + \sigma_E^2 \mathbf{I}_N). \tag{6}$$

For two quantitative traits, observed in the same set of N individuals, this model can be generalized to the following bivariate model[17]:

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \begin{pmatrix} \sigma_{G_{11}}\mathbf{A} & \sigma_{G_{12}}\mathbf{A} \\ \sigma_{G_{12}}\mathbf{A} & \sigma_{G_{22}}\mathbf{A} \end{pmatrix} + \begin{pmatrix} \sigma_{E_{11}}\mathbf{I}_N & \sigma_{E_{12}}\mathbf{I}_N \\ \sigma_{E_{12}}\mathbf{I}_N & \sigma_{E_{22}}\mathbf{I}_N \end{pmatrix} \right), \tag{7}$$

with $\mathbf{X}_1$ (resp. $\mathbf{X}_2$) the N-by-k_1 (N-by-k_2) matrix of control variables for Trait 1 (2) and associated fixed effects $\boldsymbol{\beta}_1$ ($\boldsymbol{\beta}_2$), $\sigma_{G_{st}}$ the genetic covariance between traits s and t, and $\sigma_{E_{st}}$ the environmental covariance between traits s and t, for $s = 1, 2$ and $t = 1, 2$. The Kronecker product (denoted by '$\otimes$') can be used to extend the bivariate model in Equation 7 to a multivariate model for T different traits (i.e., $\mathbf{y}_t$ for $t = 1, \ldots, T$), as follows[49]:

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_T \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_T \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_T \end{pmatrix}, \mathbf{V}_G \otimes \mathbf{A} + \mathbf{V}_E \otimes \mathbf{I}_N \right), \text{ where} \tag{8}$$

$$\mathbf{V}_G = \begin{pmatrix} \sigma_{G_{11}} & \cdots & \sigma_{G_{1T}} \\ \vdots & \ddots & \vdots \\ \sigma_{G_{1T}} & \cdots & \sigma_{G_{TT}} \end{pmatrix} \text{ and } \mathbf{V}_E = \begin{pmatrix} \sigma_{E_{11}} & \cdots & \sigma_{E_{1T}} \\ \vdots & \ddots & \vdots \\ \sigma_{E_{1T}} & \cdots & \sigma_{E_{TT}} \end{pmatrix}. \tag{9}$$

In the multivariate model, SNP-based heritability of trait t, denoted by h_t^2, and the genetic correlation between traits t and s, denoted by $\rho_{G_{st}}$, are defined as:

$$h_t^2 = \frac{\sigma_{G_{tt}}}{\sigma_{G_{tt}} + \sigma_{E_{tt}}} \text{ and } \rho_{G_{st}} = \frac{\sigma_{G_{st}}}{\sqrt{\sigma_{G_{tt}}\sigma_{G_{ss}}}}, \tag{10}$$

for $t = 1, \ldots, T$ and $s = 1, \ldots, T$.

Estimation procedure

To estimate the genetic and environmental covariance matrices $\mathbf{V}_G$ and $\mathbf{V}_E$ in Equations 8 and 9, we use restricted maximum likelihood (REML) estimation. As optimization method for our REML function, we employ a quasi-Newton approach, using a Broyden–Fletcher–Goldfarb–

Shanno (BFGS) algorithm[50]. **Supplementary Information S1** provides efficient expressions for the log-likelihood and gradient to optimize the model with a time complexity that scales linearly with the number of observations and quadratically with the number of traits. The optimization procedure guarantees that the estimated matrices $\mathbf{V}_G$ and $\mathbf{V}_E$ are positive semi-definite, by imposing an underlying factor model for both components. After optimization, standard errors can be calculated with a computational complexity that scales linearly with the number of observations and quadratically with the number of parameters in the model (which in turn scales quadratically with the number of traits).

MGREML is available as command-line tool via https://github.com/devlaming/mgreml. Runtime analyses reported in **Supplementary Information S3** show that MGREML is computationally faster than pairwise bivariate GREML. Moreover, comparison with results obtained using LD-score regression[20] shows that MGREML provides relatively tight confidence intervals for the heritability estimates and genetic correlation estimates.

Sample and data

UK Biobank is a prospective cohort study in the UK that collects physical, health and cognitive measures, and biological samples (including genotype data) in about 500,000 individuals[51]. In 2016, UK Biobank started to collect brain imaging data with the aim to scan 100,000 subjects by 2022[21,52]. We selected 43,691 individuals with available genotype data from the UK Biobank brain imaging study who self-identified as 'white British' and with similar genetic ancestry based on a principal component analysis. After stringent quality control (**Supplementary Information S4**), we estimated pairwise genetic relationships using 1,384,830 autosomal common (Minor Allele Frequency ≥ 0.01) SNPs and retained 37,392 individuals whose pairwise relationship was estimated to be less than 0.025 (approximately corresponding to second- or third-degree cousins or more distant shared ancestry). From these unrelated individuals, we retained the 20,190 individuals with complete information on all 86 traits in our analyses. A description of all the variables used in the empirical analyses is available in **Supplementary Information S2**. Mapping of each cortical region to a network of intrinsic functional connectivity (**Fig. 3**) is based on the assignment of each brain parcel in the Harvard-Oxford atlas[53] to the intrinsic functional connectivity network[38] with the highest overlap.

12

References

1. Kanai, R., & Rees, G. The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience* **12**, 231-242 (2011).
2. Crossley, N. A., et al. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain* **137**, 2382-2395 (2014).
3. Hwang, J., et al. Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **2**, 58-67 (2016).
4. Thompson, P. M., et al. ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Translational Psychiatry* **10**, 1-28 (2020).
5. Seidlitz, J., et al. Transcriptomic and cellular decoding of regional brain vulnerability to neurogenetic disorders. *Nature Communications* **11**, 1-14 (2020).
6. Nave, G., Jung, W. H., Karlsson Linnér, R., Kable, J. W., & Koellinger, P. D. Are bigger brains smarter? Evidence from a large-scale preregistered study. *Psychological Science* **30**, 43-54 (2019).
7. Avinun, R., Israel, S., Knodt, A. R., & Hariri, A. R. Little evidence for associations between the big five personality traits and variability in brain gray or white matter. *NeuroImage* **220**, 117092 (2020).
8. Elliott, L. T., et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216 (2018).
9. Grasby, K. L., et al. The genetic architecture of the human cerebral cortex. *Science* **367**, eaay6690 (2020).
10. Hofer, E., et al. Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. *Nature Communications* **11**, 1-16 (2020).
11. Smith, S. M., et al. Enhanced brain imaging genetics in UK Biobank. *BioRxiv*, 10.1101/2020.07.27.223545 (2020).
12. Zhao, B., et al. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nature Genetics* **51**, 1637-1644 (2019).
13. Witte, J. S., Visscher, P. M., & Wray, N. R. The contribution of genetic variants to disease depends on the ruler. *Nature Reviews Genetics* **15**, 765–776 (2014).
14. Posthuma, D., et al. The association between brain volume and intelligence is of genetic origin. *Nature Neuroscience* **5**, 83-84 (2002).
15. Liu, S., Smit, D. J., Abdellaoui, A., van Wingen, G., & Verweij, K. J. Brain structure and function show distinct relations with genetic predispositions to mental health and cognition. *MedRxiv*, 10.1101/2021.03.07.21252728 (2021).
16. Van der Schot, A. C., et al. Influence of genes and environment on brain volumes in twin pairs concordant and discordant for bipolar disorder. *Archives of General Psychiatry* **66**, 142-151 (2009).
17. Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M., & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540-2542 (2012).
18. Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**, 76-82 (2011).
19. Bulik-Sullivan, B., et al. An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236-1241 (2015a).

20. Bulik-Sullivan, B., et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics* **47**, 291-295 (2015b).

21. Miller, K. L., et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience* **19**, 1523-1536 (2016).

22. Shi, H., Kichaev, G., & Pasaniuc, B. Contrasting the genetic architecture of 30 complex traits from summary association data. *American Journal of Human Genetics* **99**, 139-153 (2016).

23. Evans, L. M., et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nature Genetics* **50**, 737-745 (2018).

24. Loh, P. R., et al. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature Genetics* **47**, 1385–1392 (2015).

25. Speed, D., et al. Reevaluation of SNP heritability in complex human traits. *Nature Genetics* **49**, 986–992 (2017).

26. Speed, D., Hemani, G., Johnson, M. R., & Balding, D. J. Improved heritability estimation from genome-wide SNPs. *American Journal of Human Genetics* **91**, 1011–1021 (2012).

27. Yang, J., et al. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565-569 (2010).

28. Young, A. I., et al. Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature Genetics* **50**, 1304–1310 (2018).

29. Ning, Z., Pawitan, Y., & Shen, X. (2020). High-definition likelihood inference of genetic correlations across human complex traits. *Nature Genetics* **52**, 859-864.

30. Mills, M. C., & Rahal, C. A scientometric review of genome-wide association studies. *Communications Biology* **2**, 1-11 (2019).

31. Watanabe, K., et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics* **51**, 1339-1348 (2019).

32. Zheng, J., et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).

33. Ni, G., et al. Estimation of genetic correlation via linkage disequilibrium score regression and genomic restricted maximum likelihood. *American Journal of Human Genetics* **102**, 1185-1194 (2018).

34. Yengo, L., et al. Meta-analysis of genome-wide association studies for height and body mass index in∼ 700000 individuals of European ancestry. *Human Molecular Genetics* **27**, 3641-3649 (2018).

35. Lee, J. J., et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* **50**, 1112-1121 (2018).

36. Power, R. A., & Pluess, M. Heritability estimates of the Big Five personality traits based on common genetic variants. *Translational Psychiatry* **5**, e604-e604 (2015).

37. Grotzinger, A. D., et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour* **3**, 513-525 (2019).

38. Yeo, B. T., et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology* **106**, 1125-1165 (2011).

39. Mesulam, M. M. From sensation to cognition. *Brain* **121**, 1013-1052 (1998).

40. Kaufman, L., & Rousseeuw, P. J. *Finding Groups in Data: An Introduction to Cluster Analysis* (John Wiley & Sons, 1990).

41. Beauregard, M., Lévesque, J., & Bourgouin, P. Neural correlates of conscious self-regulation of emotion. *Journal of Neuroscience* **21**, RC165 (2001).

42. Daviet, R, et al. Multimodal brain imaging study of 36,678 participants reveals adverse effects of moderate drinking. BioRxiv, 10.1101/2020.03.27.011791 (2021).

43. Giuliani, N. R., & Berkman, E. T. Craving is an affective state and its regulation can be understood in terms of the extended process model of emotion regulation. *Psychological Inquiry* **26**, 48-53 (2015).

44. Allegrini, A. G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B., & Plomin, R. Genomic prediction of cognitive traits in childhood and adolescence. *Molecular Psychiatry* **24,** 819-827 (2019).

45. Tam, A., Luedke, A. C., Walsh, J. J., Fernandez-Ruiz, J., & Garcia, A. Effects of reaction time variability and age on brain activity during Stroop task performance. *Brain Imaging and Behavior* **9**, 609-618 (2015).

46. Rakic, P. Evolution of the neocortex: A perspective from developmental biology. *Nature Reviews Neuroscience* **10**, 724-735 (2009).

47. Standring, S. *Gray's Anatomy E-book: The Anatomical Basis of Clinical Practice* (Elsevier Health Sciences. 2015).

48. Van Rheenen, W., Peyrot, W. J., Schork, A. J., Lee, S. H., & Wray, N. R. Genetic correlations of polygenic disease traits: from theory to practice. *Nature Reviews Genetics* **20**, 567-581 (2019).

49. Lynch, M., & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer, 1998).

50. Nocedal, J. and Wright, S.J. *Numerical Optimization* (Springer, 2006).

51. Sudlow, C., et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine* **12**, e1001779 (2015).

52. Alfaro-Almagro, F., et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400-424 (2018).

53. Desikan, R. S., et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968-980 (2006).

Fig. 1. Visualization of multivariate genomic-relatedness restricted maximum likelihood (MGREML). Common genetic variants (single-nucleotide polymorphisms) in the human genome (Panel a) are used to construct a Genomic-Relatedness Matrix (GRM) capturing pairwise genetic similarity between individuals in the sample (Panel b). MGREML uses this GRM to jointly estimate heritabilities of phenotypes and genetic correlations (r_g) across multiple phenotypes (Panel c), by quantifying the degree to which genetic similarity maps to phenotypic similarity (across all individuals and phenotypes in the sample). In our empirical application, 1,384,830 common SNPs are used to analyze the genetic correlations across T=86 phenotypes in a sample of N=20,190 unrelated individuals.
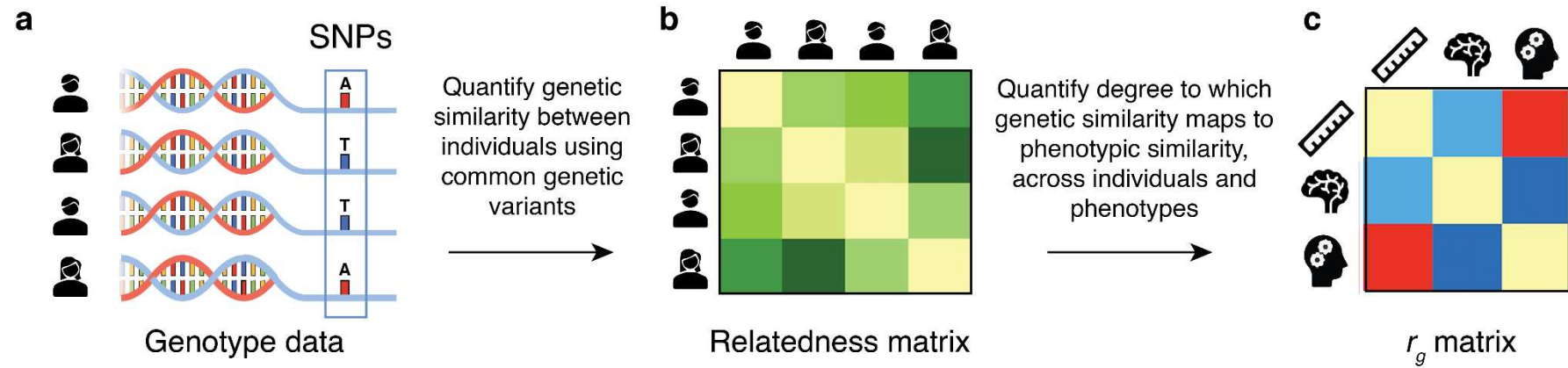
Fig. 2. Spatial mapping of SNP-based heritability and genetic correlation estimates obtained using MGREML (*N*=20,190) of relative grey matter volumes in different cortical and subcortical brain areas. a. SNP-based heritability of relative grey matter volume mapped to the respective brain region in three dimensions. Each dot represents an area, the color and size represent the heritability of that area. **b.** SNP-based heritability and standard error of relative grey matter volume of each brain region grouped by global anatomical area. **c.** Genetic correlations between cortical and subcortical relative grey matter volumes. The opacity and color represent the strength of the genetic overlap between these two areas (blue vertices represent a negative correlation, red vertices a positive correlation). Only genetic correlations larger than |0.25| are shown. **d.** Average genetic correlations in broad anatomical areas of the brain.
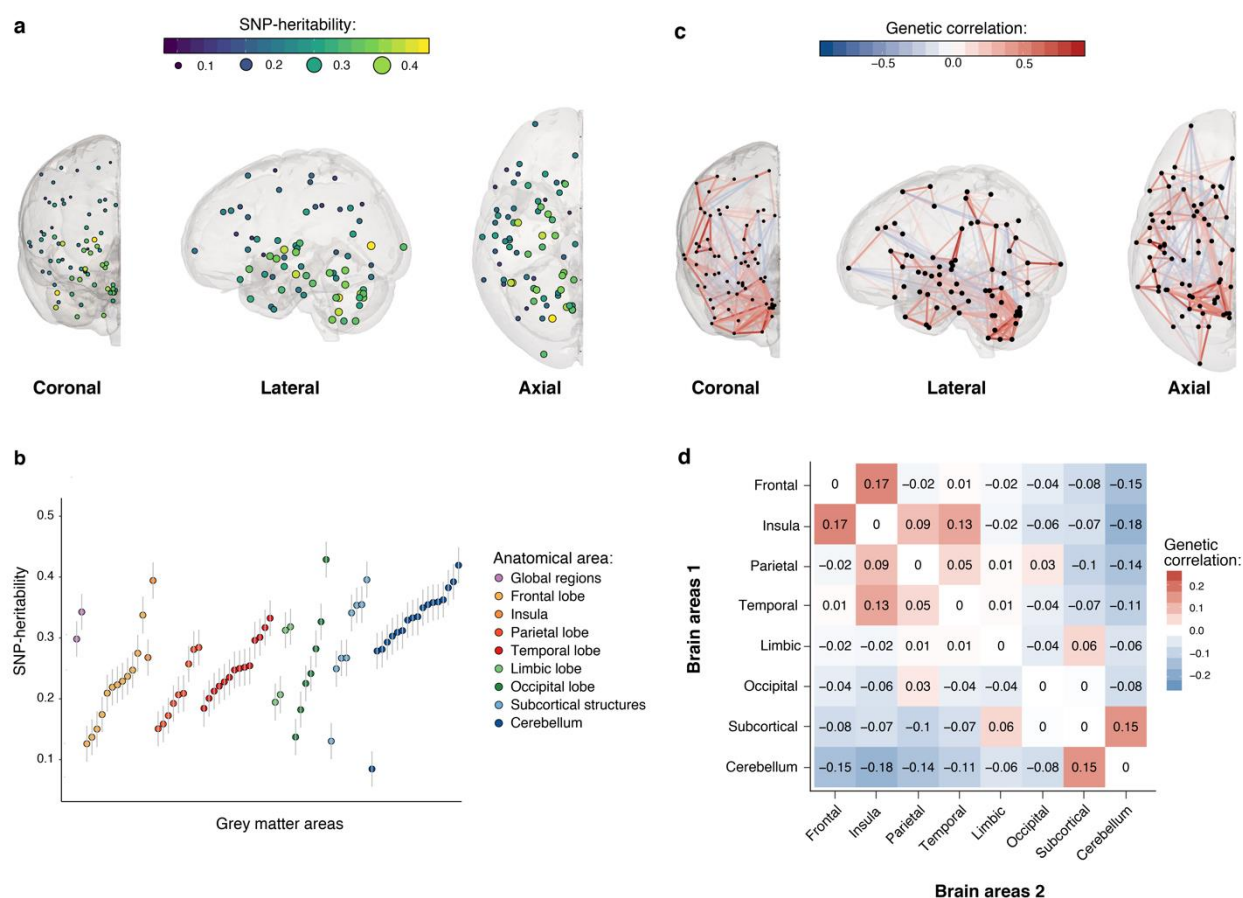
Fig. 3. Mapping of SNP-based heritability and genetic correlation estimates obtained using MGREML (*N*=20,190) of relative grey matter volumes in networks of intrinsic functional connectivity. a. Average SNP-based heritability of relative grey matter volume in networks of intrinsic functional connectivity. **b.** Genetic correlations in the brain in networks of intrinsic functional connectivity (blue vertices represent a negative correlation, red vertices a positive correlation).
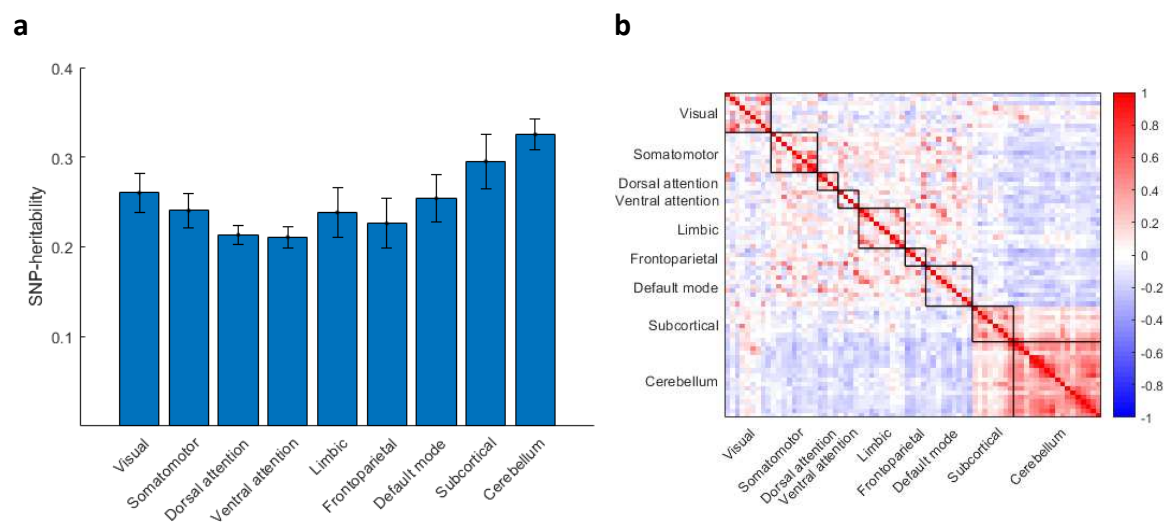
Fig. 4. Dendogram of the estimated genetic correlations for the relative grey matter volumes of the 74 regions of interest in the brain. Genetic correlations are estimated using MGREML, and clusters are identified using Ward's method with a D2 ward for hierarchical clustering. Each color represents a different cluster.
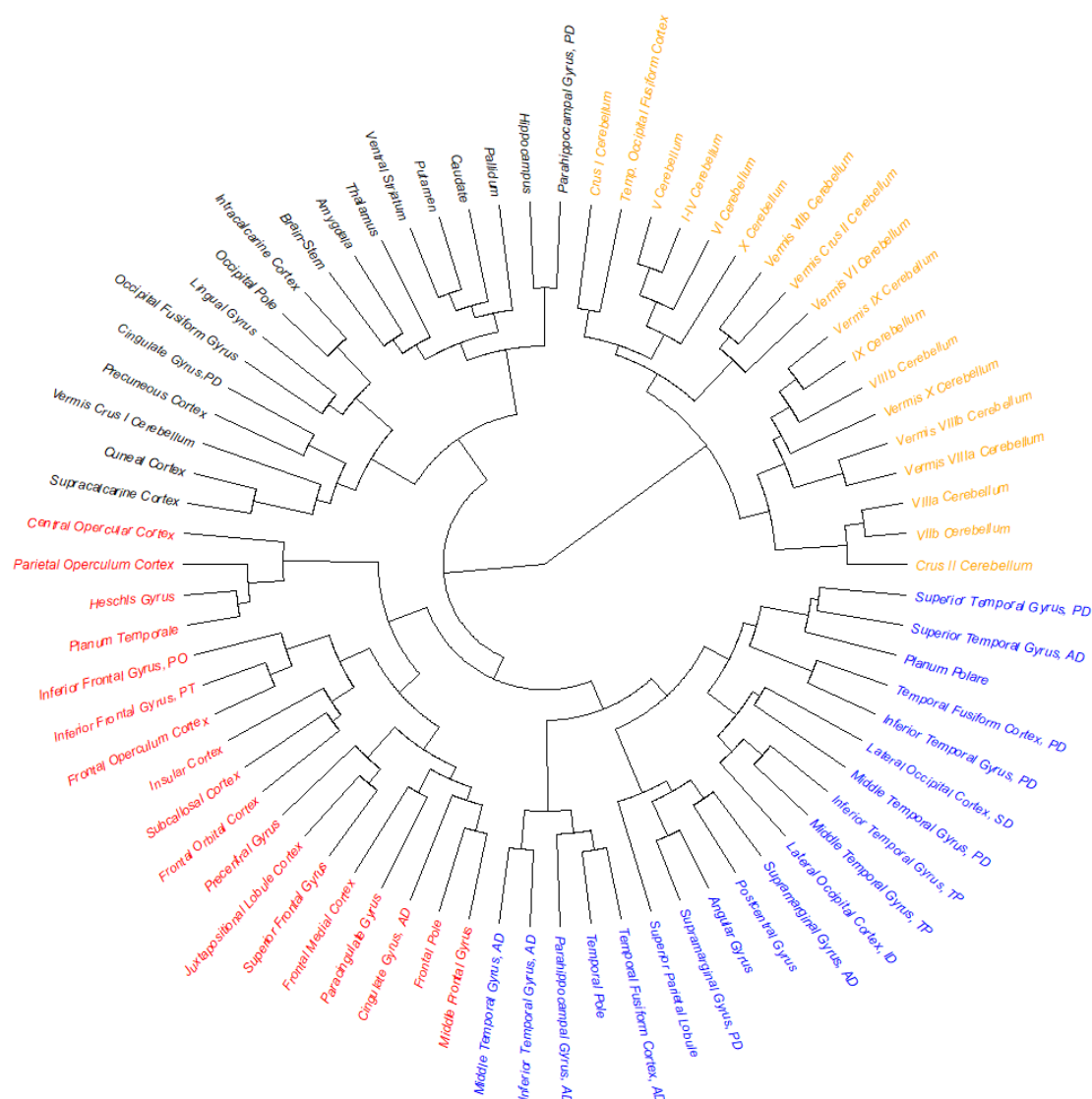
Fig. 5. Spatial mapping of genetic correlation estimates obtained using MGREML (N=20,190) of relative grey matter volumes of the 74 regions of interest in the brain and 8 behavioral traits. Blue and red points represent negative and positive genetic correlations, respectively. **a**, Subjective well-being. **b**, Neuroticism. **c,** Depression. **d**, Alcoholic drinks per week. **e**, Educational attainment. **f,** Intelligence. **g**, Visual spatial memory. **h**, Reaction time.