Supplementary Information for: A unified framework for association and prediction from vertex-wise grey-matter structure

Baptiste Couvy-Duchesne, Lachlan T. Strike, Futao Zhang, Yan Holtz, Zhili Zheng, Kathryn E. Kemper, Loic Yengo, Margaret J. Wright, Naomi R. Wray, Jian Yang, Peter M. Visscher

Correspondence: BCD (b.couvyduchesne@uq.edu.au), JY (jian.yang@uq.edu.au) or PMV (peter.visscher@uq.edu.au).

This PDF file includes:

Supplementary text Appendix S1 to S9
Figure S1 to S12
Tables S1 to S3
Captions for databases S1 to S13
References for SI reference citations

Other supplementary materials for this manuscript include the following:

Datasets S1 to S13

Appendix S1: MRI acquisition parameters

UKB samples

MRI images were collected using a 3T Siemens Skyra machine (software platform VD13) and a 32-channel head coil (1). The T1 weighted (T1w) images were acquired over 4:54 minutes, voxel size 1.0x1.0x1.0mm, matrix of 208x256x256mm, using a 3D MPRAGE sequence (2), sagittal orientation of slice acquisition, R=2 (in plane acceleration factor), TI/TR=880/2000 ms (1). The T2 FLAIR acquisition lasted 5:52 minutes, voxel size 1.05x1.0x1.0 mm, matrix of 192x256x256 voxels, 3D SPACE sequence (3), sagittal orientation, R=2, partial Fourier 7/8, fat saturated, TI/TR=1800/5000ms, elliptical (1).

HCP sample

T1w and T2 weighted (T2w) images were collected at the Washington University (St Louis, Missouri) on a 3T Siemens Skyra scanner using a standard 32-channel head coil (4, 5). Two T1w images were acquired, each over 7 minutes and 40 seconds with a voxel size of 0.7x0.7x0.7mm, matrix/FOV of 224x224x224mm using a 3D MPRAGE sequence (2), TR/TE/TI=2400/2.14/1000ms, flip angle 8degrees, R=2, sagittal orientation of slice acquisition (6). Similarly, two T2w images were acquired over 8:24 min each, voxel size 0.7x0.7x0.7mm, matrix of 224x224x224mm, 3DSPACE sequence (3), sagittal orientation, R=2, TR/TE=3200/565, no fat suppression pulse.

Appendix S2: Summary of image processing and QC

Exclusion due to MRI processing in the UKB

At the time of download (July 2017), T1w images were available for 10,102 participants of the UK Biobank (UKB) project. None of the participants had withdrawn consent after the data was collected. We excluded 175 participants with T1w images labelled as unusable by the UKB, leaving 9,928 MRI scans to process. T2 FLAIR images were available for 9,755 of those. The FreeSurfer processing failed or did not complete within 48 hours for a handful of participants: 37 for cortical processing, 19 for subcortical, including 17 for whom both processing failed. For simplicity, we chose not to re-run image processing on these participants as their exclusion should have a minimal impact on the results obtained from the full sample. Excluded individuals are described in **Dataset S1**. Our final sample comprised 9,890 participants with usable cortical data, 9,908 with subcortical data and 9,888 with both cortical and subcortical data. This sample consisted of 9,888 adults aged 62.5 on average (SD=7.5, range 44.6–79.6) and comprised 52.4% of female participants. We further excluded 391 participants with extreme brains (outliers) or likely to have a large effect on the analyses (**see below for details about QC**).

Replication data set was downloaded in May 2018 and consisted in an additional 4,942 participants with a T1w image. Image processing and phenotype selection were identical to that of the discovery sample. This led to the exclusion of 381 participants whose processing failed and 238 excluded from QC (**details below**). See **Dataset S1** for a full description of replication participants (final, QCed and failed processing) in addition to a comparison of the discovery and replication samples.

Automated quality control based on the BRM

The standards in imaging are to perform a visual QC of the processed images following a (mostly) automated pipeline. For example, the ENIGMA protocol recommends checking participants with outlying measurements but also requires a visual QC of each scan to control the cortical and subcortical parcellation (http://enigma.ini.usc.edu/protocols/imaging-protocols/). This may prove extremely time consuming, especially on large samples such as the UKB that were not available when the ENIGMA pipeline was created.

Here, we propose to utilise the information contained in the BRMs to perform QC. We excluded participants showing extreme values on the diagonals of the BRMs (diagonal>2.5, we did not observe any heavy left, see **Appendix S4**). In addition, we excluded individuals with outlying off-diagonal elements as they could confound our variance component analyses. We took the average of the BRM elements (in absolute value) for each individual (i.e. average of ith row of the BRM for the ith individual) and excluded participants with a statistic more that 4SD away from the mean. We reported the histograms of BRM diagonals and off-diagonals before and after QC (**Appendix S4**). The arbitrary cut-off for the BRM diagonal (>2.5) was determined from the HCP sample on which we had performed visual QC as per ENIGMA protocols (http://enigma.ini.usc.edu/protocols/imaging-protocols/).

We applied the same level of QC in the UKB sample but could not compare our approach to visual QC exclusion due to the size of the dataset. Instead, we describe the participants excluded due to failed processing or QC, to check if their exclusion may impact the results presented (**Dataset S1**).

Comparison of visual vs. BRM-based QC approaches in the HCP sample

A total of twenty-four participants were excluded in our QC step based on the diagonal values (>2.5) or outlying off-diagonal elements in the BRM. Twenty-two participants were flagged using each of the BRM QC criteria. More importantly, 20 outlying individuals were flagged by both BRM criteria. Thus, participants with outlying brains, as indicated by large BRM diagonal values tended to exhibit outlying off-diagonal values (i.e. brain structure similarities) with other individuals, potentially causing unstable estimates in variance component analyses.

Out of the 24 individuals excluded in our data driven QC, 14 had also been flagged using the ENIGMA visual QC protocol: 3 were fully excluded for incorrect cortical reconstruction, 7 had an incorrectly segmented hippocampus and 7 others failed visual QC for 3+ cortical regions. Finally, our data driven QC did not identify some individuals flagged using the ENIGMA visual QC: 4 with incorrect hippocampal reconstruction and 108 with incorrect parcellation of the cingulate cortex. The case of the cingulate parcellation is highlighted in the ENIGMA QC protocols as its boundary with regions in the frontal cortex are often misplaced in FreeSurfer. However, this should not be a problem when working at a vertex level as the cortical ribbon remains well segmented, and this may be why these individuals are not identified by our QC approach.

Description of excluded participants in the UKB

We report the mean (SD) or % of each answer (for qualitative variables) for all the phenotypes considered from the UKB (**Dataset S1**) and compare the mean and variances between included and excluded participants. We used a conservative Bonferroni significance threshold of 1e-4 to account for the number of tests.

The participants we excluded (either for unusable T1, or QC) were on average more than 2.7 years older than the individuals used in the analysis (p-value<3.3e-7) and men were over-represented (62% of excluded were men vs. 47%, p-value<1.8e-5). In addition, excluded individuals were more variable in term of digit matching reaction time, dried fruit intake and exposure to passive smoking. They were less variable than included participants in regard to their basophil percentages (**Dataset S1)**.

Individuals with unusable T1 reported a smaller amount of passive smoking at home (0.001 days a week vs. 0.2, p-value=4.7e-9). They were also less variable than individuals included in the analysis in term of their depression scores.

On the other hand, individuals excluded from QC were 10% less performant as the digit matching task than included participants (smaller number of correct matches or attempted matches, greater reaction time, p-value<4.8e-7). They also were more likely to be diabetics (10% in QCed participants vs. 5%, p-value=1.e-5) and had a reduced acceleration force (-2.4m/s-1, p-value=3.5e-5) as well as greater waist circumference (+3.1cm, p-value=7.0e-7). In addition, the individuals QCed out of the analyses had a greater ICV,

smaller grey matter volume, hippocampus volumes or cortical thickness. More importantly excluded individuals exhibited much greater variances in all brain measurements which suggests imperfect/failed processing.

Sample description HCP

Similar to the results in the UKB, HCP participants excluded by QC showed a significantly greater variance in brain measurements than included participants (**Dataset S2**). This further validates our QC approach, suggesting that the participants QCed out exhibit outlying brains, some due to failure of the MRI processing pipeline.

In addition, excluded participants differed (p-value<1e-4) on some aspects of cognition: delay discounting $200 at 5 years (smaller mean and variance), spatial orientation (total positions; greater variance) and sustained attention (longest run non-response; smaller variance), depression scores (smaller mean and variance).

The similarities between excluded participants in the UKB and HCP (e.g. depression scores or cognition) are intriguing. We hypothesise that these phenotypes may be associated with greater level of movement in the scanner leading to lower image quality and failed processing. Us and others previously showed that inattention and hyperactivity are associated with greater movement in resting-state fMRI(7, 8), and a subsequent study in the HCP found multiple factors also associated with motion during rs-fMRI (for example: some cognition domains, antisocial or somatic scores, weight and BMI as well as tobacco use)(9).

Note that when the variance in excluded and included participants differs, the sample participants may not capture the full phenotypic variance and the results of variance component analyses should be interpreted with caution. In other words, we are estimating the proportion of in sample variance accounted for by brain features which may differ from the proportion of total phenotypic variance accounted for in the population.

QC when comparing the different cortical processing (mesh coarseness and smoothing)

In this section, we performed a slightly different QC which consisted of excluding individuals with extreme off-diagonal BRM elements (+-5SD from mean, option previously not available in OSCA). We note that this QC is slightly more stringent but did not change the morphometricity estimates reported in the main text (**Dataset S3, S12**).

It is important to have a QC that adapts to the distribution of the BRM off-diagonals, which may differ depending on the processing options. To note, most of the excluded participants were the same across the different processing scenarios.

Appendix S3: HCP image processing

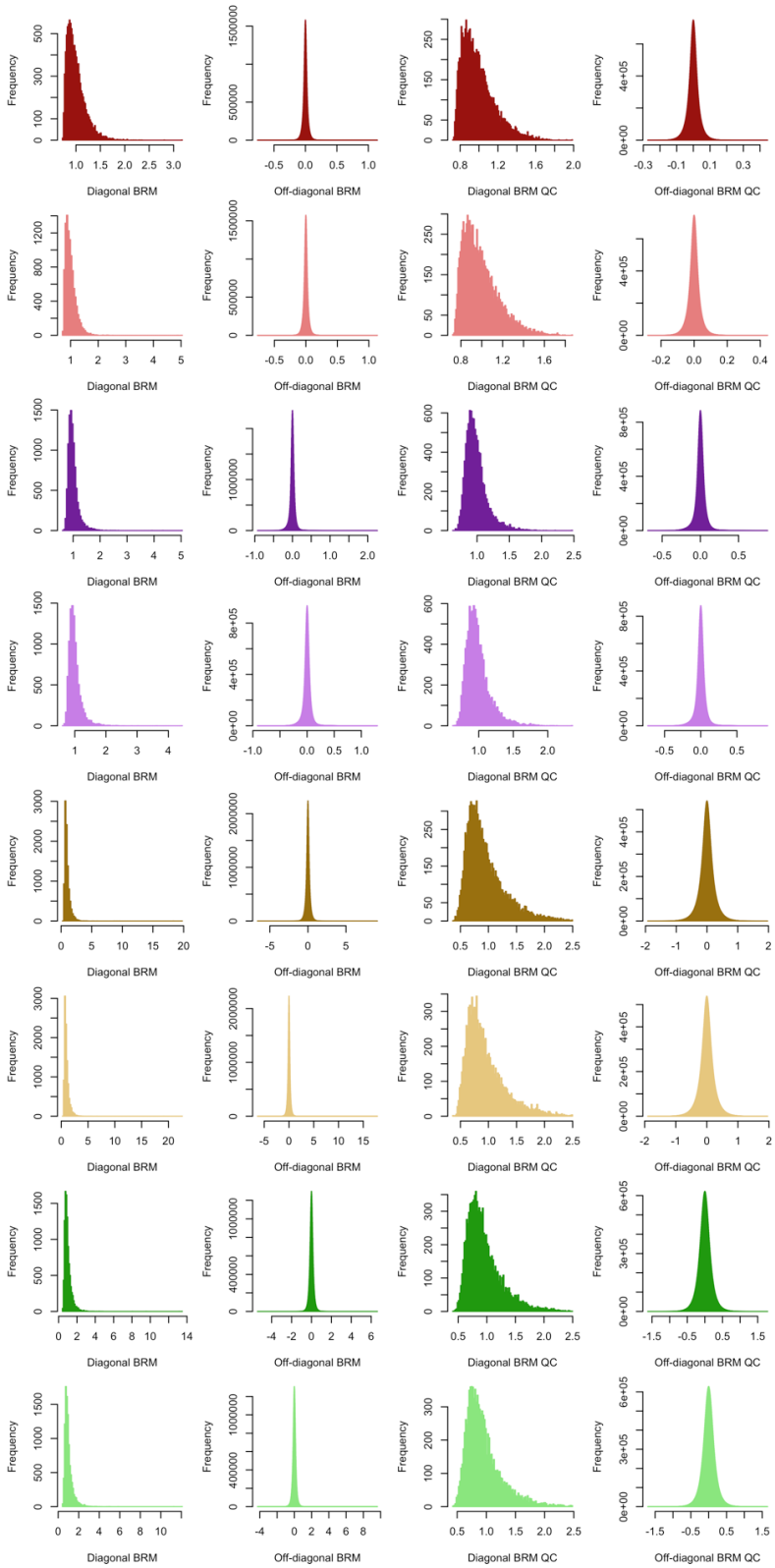
The HCP team (6, 10, 11) pre-processed the structural scans to facilitate scan comparison across individuals, removing spatial artefacts and improving T1w and T2w alignment using FSL (12, 13) and FreeSurfer (14). When both passed HCP quality control (QC), T1w and T2w images they processed them together in FreeSurfer 6.0 (14), otherwise data extraction relied on a single scan (6). Participants with poor quality T1w and T2w scans were re-imaged (6). Cortical processing (recon-all procedure in FreeSurfer) was also performed by the HCP team and included down sampling to 1mm size voxels and 256x256x256 matrix, aided registration using customised brain mask, and two manual steps performed outside of the recon-all procedure to enhance white matter and pial reconstruction (6).

Appendix S4: BRM interpretation

Diagonals of the BRM consist of the mean square of the participants' vertex-wise measurements. Since the vertex-wise data are centred, larger diagonal elements reflect a greater proportion of extreme phenotypes, may they be small or large vertex measurements. Thus, we can interpret large diagonal values as "outstanding brains" in term of size/shape or due to failure in processing (e.g. incorrect cortical or subcortical parcellation). On the other hand, small diagonal values indicate small absolute values across the brain measurements (mean centred), hence brains close to the average brain. Across all the samples and BRMs, we observed that the diagonal elements were centred around 1, and skewed to the right (**Appendix S4 Figure 1-2**).

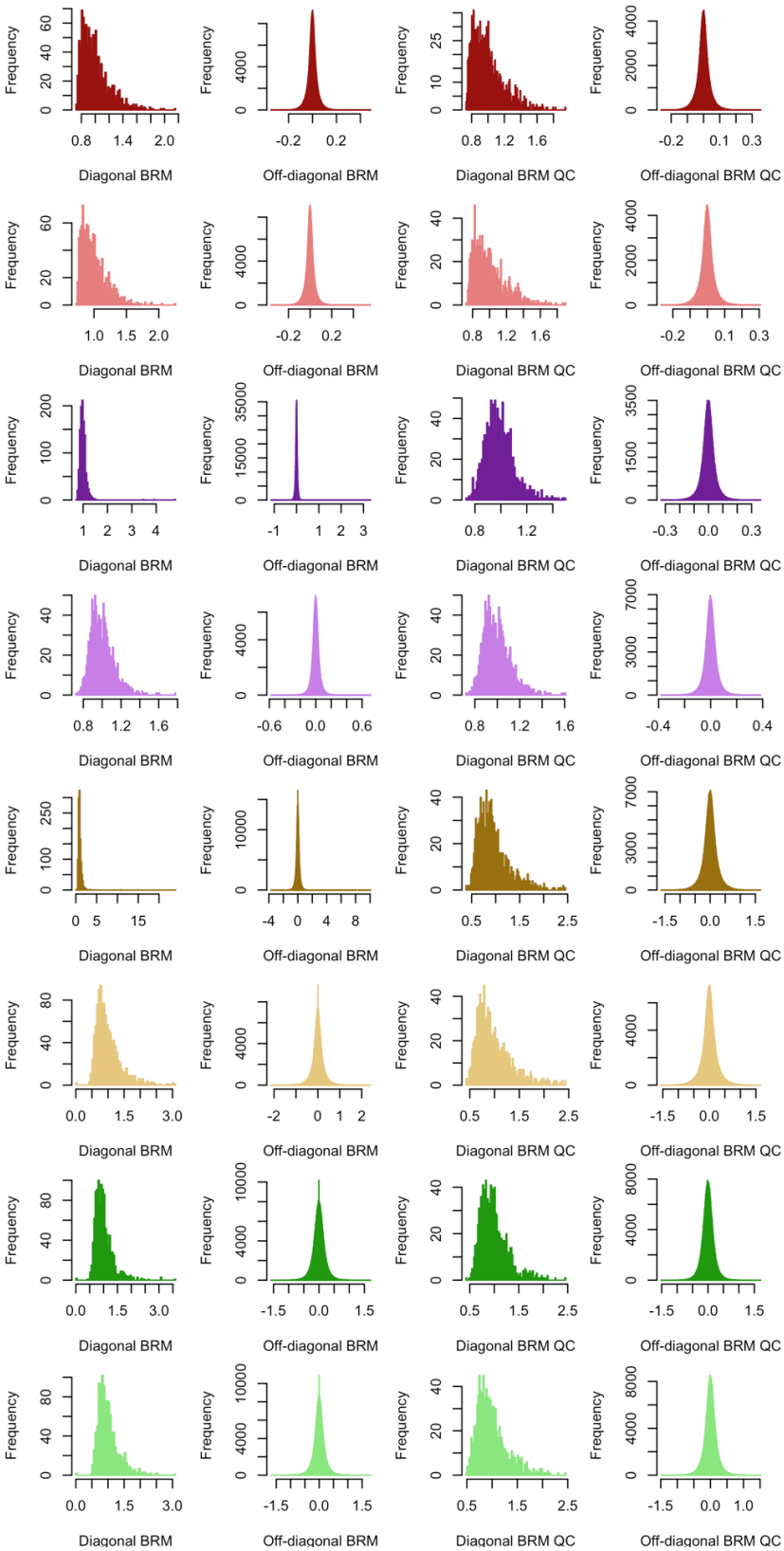
Off-diagonal elements of the BRM are the covariances between two individuals' measurements, thus greater values indicate greater similarities between the pair of participants. Off-diagonal elements are normally distributed with a mean of 0. Their dispersion varies upon the brain modality considered and the degree of correlation between the vertices (**Appendix S4 Figure 1-2**).

Participant's brain similarities are thought to reflect (some of) the participants' similarities in genetics and environment, in other words we expect part of the variance accounted for by the BRM to be genetic. Indeed, we observed a positive correlation between elements of the BRM and of the Genetic Relatedness Matrix (GRM) in the UKB or the pedigree matrix in the HCP, which suggests that participants more alike genetically also exhibit more similar brains (**Appendix S4 Figure 3-4**). The GRM was calculated in GCTA(15) from the hap-map 3 variants further filtered for MAF>0.01, pHWE<10-6 and missingness<0.05 for a total of 1,123,943 variants. The GRM calculation was restricted to participants of European ancestry (N=456,426) defined by a >0.9 posterior probability of belonging to the 1000G reference ancestry cluster. Overall, the pairs of twins and especially monozygotic twins did not exhibit outlying BRM values compared to the unrelated pairs (**Appendix S4 Figure 4**), suggesting that they should not confound the results of a variance component analysis. Quite the opposite, they increase the variance of the off-diagonal BRM elements which results in an increased power **Appendix S6**).



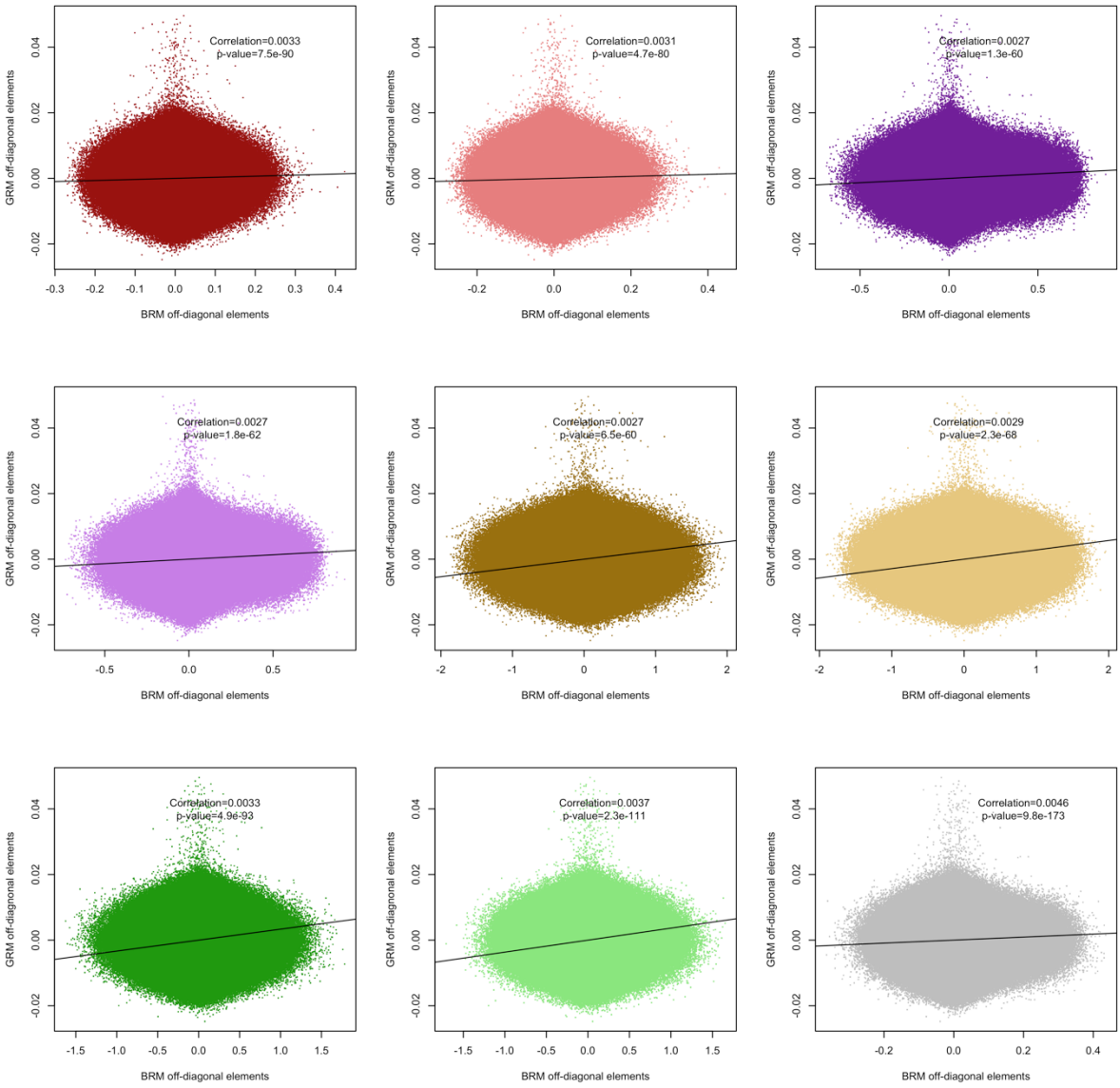
Appendix S4 Figure 1: Histograms of diagonal and off-diagonal elements of the UKB brain relatedness matrix.

Plots are presented before (left panels) and after participants' QC exclusion (right panels). Colors correspond to the different brain modalities: dark red – left cortical thickness, light red – right cortical thickness, dark purple – left cortical area, light purple – left cortical area, dark yellow – left subcortical curvature, light yellow – right subcortical curvature, dark green – left subcortical thickness, light green – right subcortical thickness.



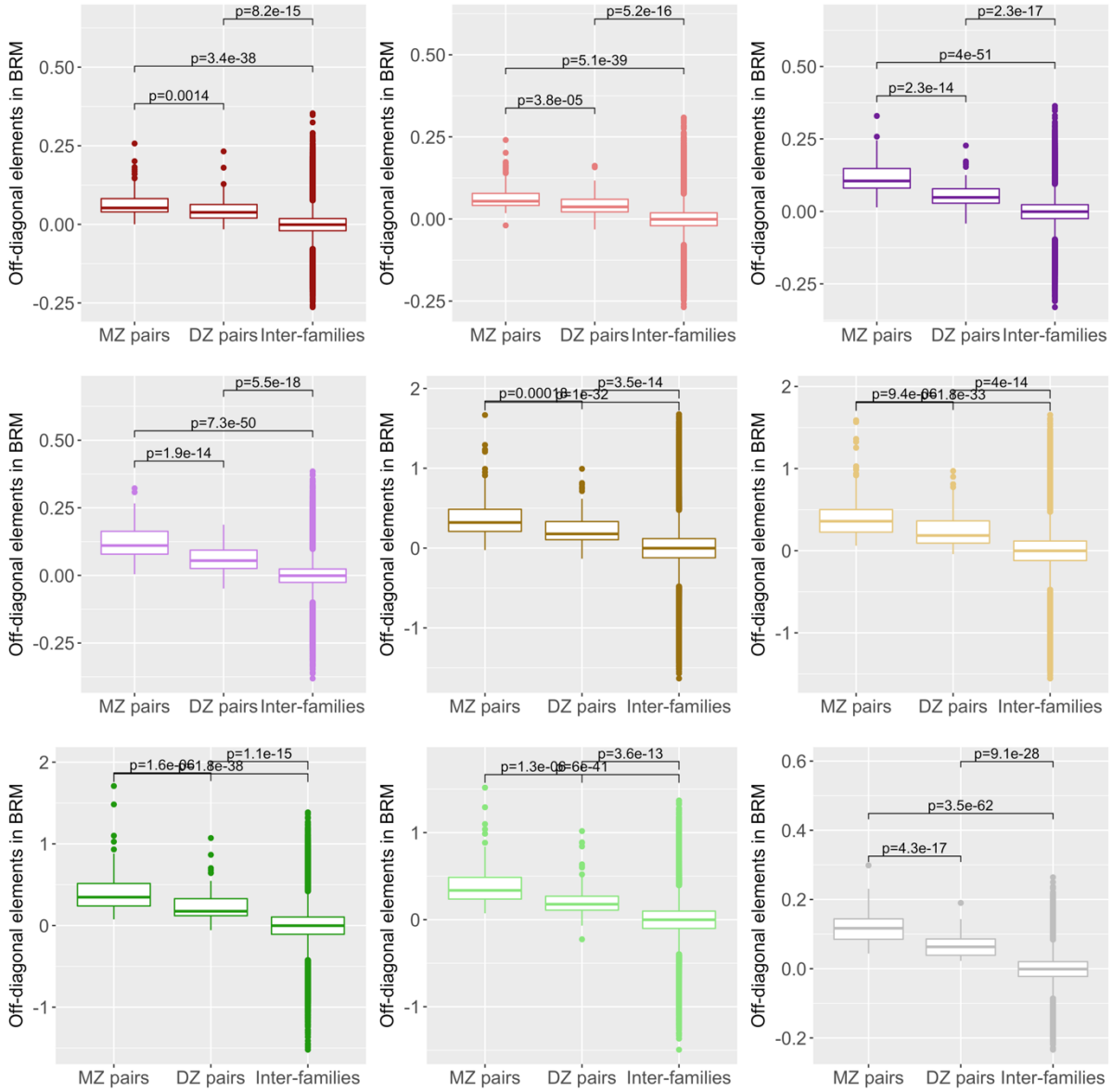
Appendix S4 Figure 2: Histograms of diagonal and off-diagonal elements of the HCP brain relatedness matrix.

Plots are presented before and after participants' QC exclusion. Colors correspond to the different brain modalities: dark red – left cortical thickness, light red – right cortical thickness, dark purple – left cortical area, light purple – left cortical area, dark yellow – left subcortical curvature, light yellow – right subcortical curvature, dark green – left subcortical thickness, light green – right subcortical thickness.



Appendix S4 Figure 3: Scatterplot showing the correlation between GRM and BRM pairwise elements in the UKB sample.

Correlations and p-values are shown. Colors correspond to the different brain modalities: dark red – left cortical thickness, light red – right cortical thickness, dark purple – left cortical area, light purple – left cortical area, dark yellow – left subcortical curvature, light yellow – right subcortical curvature, dark green – left subcortical thickness, light green – right subcortical thickness. Grey corresponds to all brain vertices.



Appendix S4 Figure 4: Boxplot showing the BRM pairwise elements per zygosity group in the HCP. Tests of mean differences are shown. Colors correspond to the different brain modalities: dark red – left cortical thickness, light red – right cortical thickness, dark purple – left cortical area, light purple – left cortical area, dark yellow – left subcortical curvature, light yellow – right subcortical curvature, dark green – left subcortical thickness, light green – right subcortical thickness. Grey corresponds to all brain vertices.

Appendix S5: assumption testing in linear mixed models

We tested whether the variance accounted for by the brain similarities was significantly different from 0 using a likelihood ratio test on nested models (with and without the random effect). The test statistic follows a chi-square distribution with x degree of freedom (x being the number of variance components tested) for a σ_0^2 value inside the parameter space. However, when testing $H0: \sigma_0^2 = 0 \; vs. \; H1: \sigma_1^2 > 0$, the p-value should be interpreted with caution as the estimator may not be asymptotically normally distributed because 0 is a boundary of the parameter space (16, 17). Some have suggested that the p-value could be better approximated using a mixture of chi-square distributions in the test of significance (16, 17). However, a 50:50 mixture has been shown to be sometimes inappropriate (18, 19) as the test relies on assumptions often not met in LMM (such as i.i.d. observations) (19). Thus, we preferred using a $\chi 2$ (x df.), the only consequence being a less powerful hence conservative test (18-20). Such test is implemented in OSCA (21), as well as in GCTA (15).

Appendix S6: SE of the residual correlation

Residual correlation (r_E) offers insight into factors, shared between the traits, but that do not relate to grey-matter structure (e.g. other brain modalities, non-brain contribution). A weighted sum of r_{GM} and r_E make up the phenotypic correlation. Thus, we calculated rE from the phenotypic and grey-matter correlation using:

$$r_E = \frac{r - r_{GM} * \sqrt{R_1^2 * R_2^2}}{\sqrt{(1 - R_1^2) * (1 - R_2^2)}}$$

with $R_1^2$and R_2^2 the brain-morphometricity of the two traits included in the bivariate model.

To derive the SE, we use the fact that the residual correlation between random variables X and Y can be expressed as a function of the residual covariance between X and Y ($\hat{\sigma}_{XY}$), and the residual association with X or Y ($\hat{\sigma}^2{}_X \; and \; \hat{\sigma}^2{}_Y$).

$$rE = \frac{\hat{\sigma}_{XY}}{\sqrt{\hat{\sigma}^2{}_X * \; \hat{\sigma}^2{}_Y}} = \frac{a}{\sqrt{b * c}}$$

We can derive a first order tri-variate Taylor series approximation around the expected values of $\hat{\sigma}_{XY}, \hat{\sigma}^2{}_X \; and \; \hat{\sigma}^2{}_Y \; (denoted \; \mu_a, \mu_b \; and \; \mu_c \; for \; convenience)$. Doing so, we implicitly assume these numbers are estimated with a reasonable confidence so that the Taylor series approximation around the mean holds. Thus,

$$rE \; \approx \; \mu_{rE} + \frac{(a-\mu_a)}{\sqrt{\mu_b * \mu_c}} - \frac{0.5 * \mu_a * (b-\mu_b)}{\mu_b \sqrt{\mu_b * \mu_c}} \; - \frac{0.5 * \mu_a * (c-\mu_c)}{\mu_c \sqrt{\mu_c * \mu_b}}$$

Taking the variance:

$$V(rE) \approx \; \frac{V(a)}{\mu_b * \; \mu_c} + \frac{0.25 * {\mu_a}^2 * \; V(b)}{{\mu_b}^3 * \; \mu_c} + \frac{0.25 * {\mu_a}^2 * \; V(c)}{{\mu_c}^3 * \; \mu_b} - \frac{\mu_a * \; cov(a,b)}{{\mu_b}^2 * \; \mu_c} - \frac{\mu_a * \; cov(a,c)}{{\mu_c}^2 * \; \mu_b} + \frac{0.5 \; * \; {\mu_a}^2 * cov(b,c)}{{\mu_c}^2 * \; {\mu_b}^2}$$

To conclude, approximating the variance of rE using the formula above, requires the variance components estimates ($\hat{\sigma}_{XY}, \hat{\sigma}^2{}_X \; and \; \hat{\sigma}^2{}_Y$) from the bivariate model as well as their matrix of sampling variance-covariance for the values $V(\hat{\sigma}_{XY}), V(\hat{\sigma}^2{}_X), V(\hat{\sigma}^2{}_y), cov(\hat{\sigma}_{XY}, \hat{\sigma}^2{}_X), \; cov(\hat{\sigma}_{X\square}, \hat{\sigma}^2{}_y) \; and \; cov(\hat{\sigma}^2{}_X, \hat{\sigma}^2{}_y)$.

Such values are estimated in OSCA/GCTA and may be found in the log files outputted during the model fitting. Note that SE of the grey-matter correlation is derived using the same approach. The interested readers may refer to (22-24).

For significance testing, we used a one-sided test based on the test statistic:

$$\left(\frac{rE}{SE(rE)}\right)^2 \sim \chi(1)$$

Appendix S7: Power of linear mixed models

Power of the current analyses:

In the UKB discovery sample (assuming N=9,500), we had 80% power to detect an effect >2.2% of variance accounted for by the random effect, while taking into account multiple testing (pvalue significance threshold p<0.05/175). In the HCP sample (assuming N=1,000), considering the number of tests performed (p<0.05/160), we would need an association effect size of 20% of variance to yield the same power (25). For brain correlations, the calculation of statistical power depends on the sample size (set to 9,500), the variance accounted for in each phenotype (we chose 5%), the phenotypic correlation (set to r=0.2), the significance threshold (p<4.2e-5, based on our number of tests) as well as the variance of off-diagonal elements of the BRM $var(B_{ij})$ (0.00096, for the BRM of all brain features) (25). In this example, we had 80% power to detect a brain correlation greater than 0.35, but only a 7% power for a brain correlation of 0.2. Using a sample of N=1,000, as per the HCP, and selecting phenotypes with >20% variance accounted for (everything else being equal), we have a 1% power to detect a brain correlation of 0.35, and we would need a brain correlation greater than 0.99 to achieve 80% power.

Power derivations for LMMs

Power calculation of variance component analysis may be derived from the sampling variance of the estimate: $var(\hat{\sigma}_b^2)$, which is the square of the standard error (SE) of the estimate. In REML analyses (e.g. GCTA or OSCA) the SE is estimated from diagonal elements of the inverse of the information matrix, and it has not been derived analytically. Visscher et al.,(25) showed that the SE could be approximated using a simpler model formulation known as Haseman Elston (HE) regression(26), which is the ordinal least square equivalent of the REML approach. As such, HE regression should be slightly less powerful that the REML approach, resulting in a marginal underestimation of the power.

Briefly, HE regression(26) performs a linear regression of the phenotype pairwise product $z_{ij}=y_i y_j$ on the off-diagonal elements of the BRM: B_{ij}. For a pair of individual i and j:

$$z_{ij} = \mu + b\, B_{ij} + \varepsilon_{ij}$$

This model includes n=N(N-1)/2 pairwise observations and we can easily show that $b = \sigma_B^2$ (25, 26). In this simple linear regression framework, assuming that the ε_{ij} are i.i.d. we can calculate the $var(\hat{\sigma}_B^2) = var(\hat{b}) = \frac{var(\varepsilon_{ij})}{n\, var(B_{ij})}$.

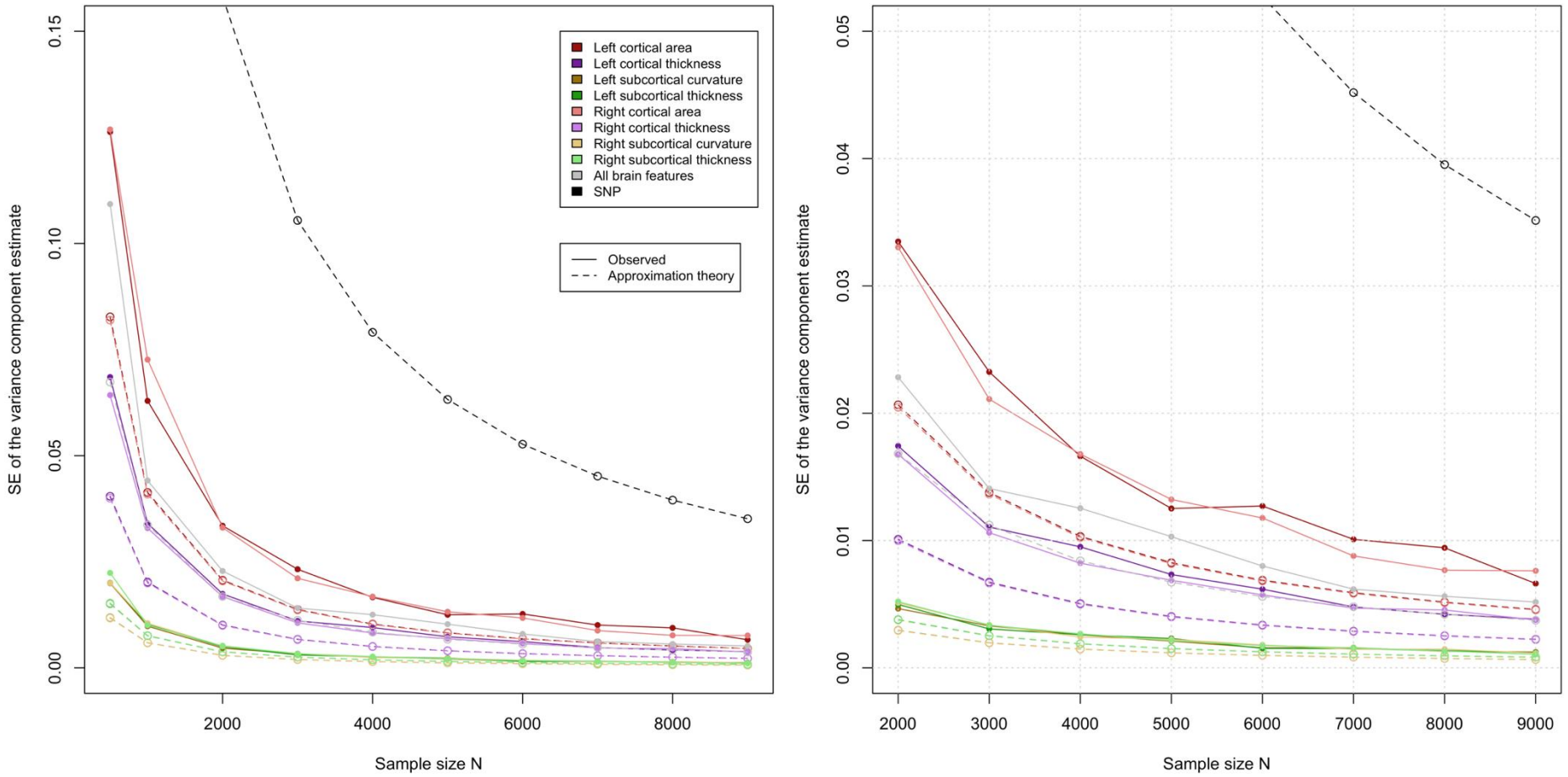
For centred and standardised phenotypes Y, $var(z_{ij}) = 1$ and therefore $var(\varepsilon_{ij}) \leq 1$. Thus, $var(\hat{\sigma}_B^2) \leq \frac{2}{N(N-1)\, var(B_{ij})}$. As a consequence, the power of variance component analysis may be approximated from the sample size N and the variance of the off-diagonal elements of the BRM. The same formula holds for discrete outcome variables (e.g. sex or disease status) and a similar derivations may be performed for a the bivariate case and the power of brain correlation(25). In our UKB and HCP data, we calculated the variance of the off-diagonal BRM in order to get an approximation of the SE of our variance component estimates. The $var(B_{ij})$ were consistent across sample and across the left and right modalities (**Appendix S7 Table 1**). Such stability of the variance of off-diagonal elements has also been observed

for GRM(25, 27). Thus, in our UKB imaging sample, the variance of the off-diagonal GRM was 2.0e-5 (after removing related individuals with GRM elements>0.05), consistent with previous report and analytic derivations from our group(25, 27).
We tried to validate the derivations presented above using simulation. Thus, we simulated 100 normally distributed phenotypes (mean 0, variance 1) and estimated the SE of the variance components using OSCA, varying the sample size (from N=500 to 9,000). We used random subsets of the BRM calculated from the UKB data for the variance-covariance of the random effect. We conducted such analysis for the 8 brain modalities as well as for the global BRM. Results of simulation and approximation theory are presented in **Appendix S7 Table 1, Appendix S7 Figure 1 and 2,** and suggest that the approximation yields realistic, though slightly underestimated, values for the SE of the variance component estimates. Note that HE regression is known to produce underestimated SE at high power as the i.i.d. assumption of errors does not hold anymore(28). Despite being small, the underestimation of SE may lead to overestimate the statistical power using the approximation theory. To circumvent this problem, we provide values for $var(B_{ij})$, derived from our simulation analysis, that result in realistic power calculation using the approximation theory (**Appendix S7 Table 1)**. The statistical power may be calculated (for a hypothesised variance accounted for: σ_B^2 and a selected risk α) from the non-centrality parameter of the chi-square statistic $ncp = \left(\frac{\sigma_B^2}{SE}\right)^2$, using the R formula 1-pchisq(qchisq(1-α, df), df, ncp) or the GCTA-GREML online power calculator (http://cnsgenomics.com/shiny/gctaPower/).
Results presented below also highlight the greater power of the brain variance component analyses compared to estimation of SNP heritability, as indicated by the smaller SE of the estimates. For example, in a sample of N=1000, we would have >60% power to detect an effect $\sigma_B^2 > 0.1$ (with α=0.05), but only a 5.5% power to detect a SNP heritability greater than 0.1 (**Appendix S7 Figure 1 and 2**).

		Var(Bij) UKB	Var(Bij) HCP	Var(Bij) to use in power calculation (corrected based on our simulations)
cortical area	Left	0.0012	0.0015	0.00047
	right	0.0012	0.0016	0.00046
Cortical thickness	Left	0.0049	0.0021	0.0017
	Right	0.0057	0.0024	0.0017
Subcortical curvature	Left	0.057	0.057	0.017
	Right	0.058	0.059	0.017
Subcortical thickness	Left	0.035	0.036	0.018
	right	0.034	0.035	0.020
All modalities		0.0017	0.0014	0.00096

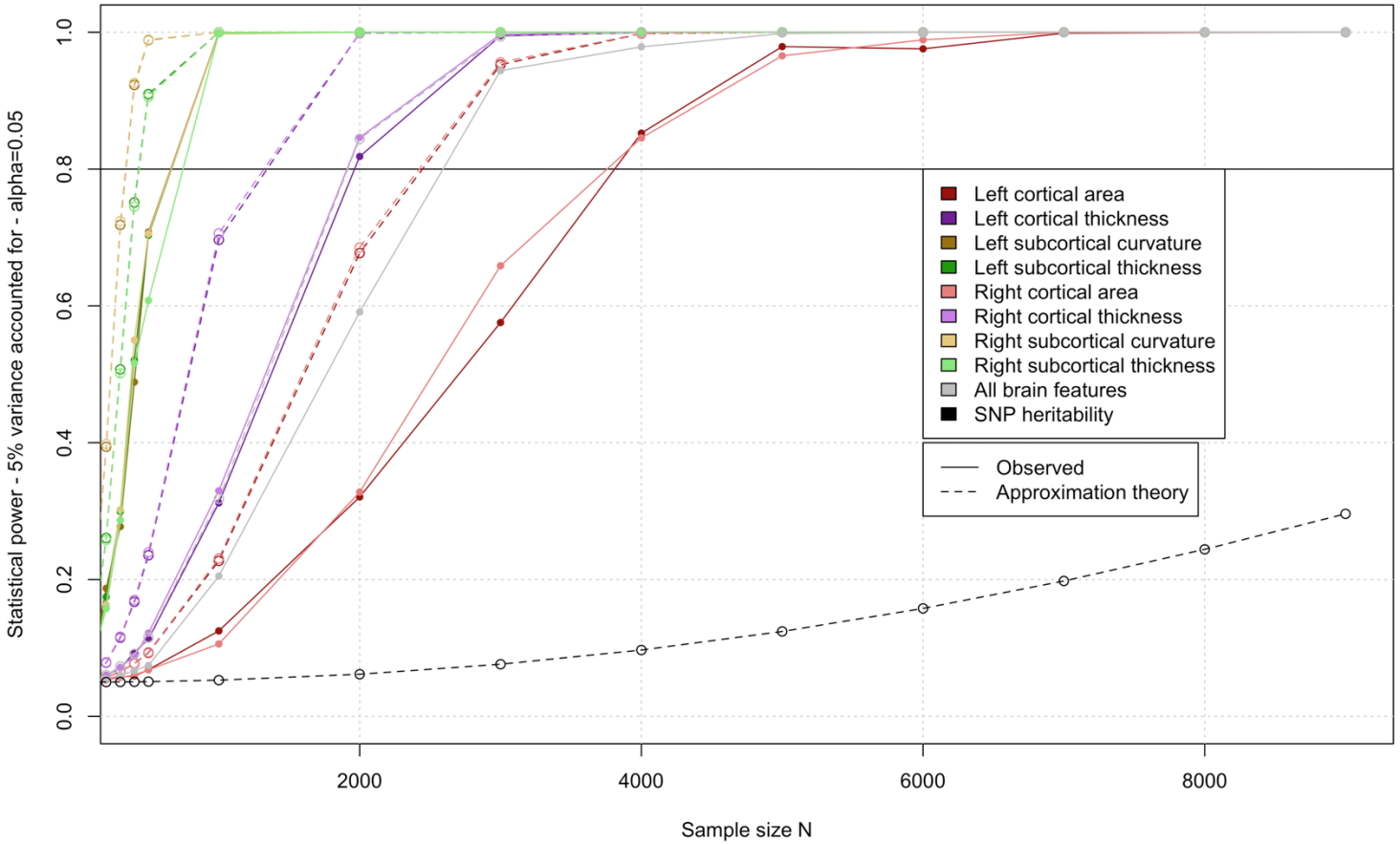
Appendix S7 Table 1: Variance of off-diagonal elements in the UKB and HCP.

Following our simulation results and to avoid overestimating the statistical power of brain variance-component analysis, we recommend using the values in the right-hand side column in the approximation theory formula (see main test for the formula, or http://cnsgenomics.com/shiny/gctaPower/ for online power calculator).



Appendix S7 Figure 1: Empirical and approximated SE of the variance component estimates using BRM and GRM

The right panel is a close up of the left one, restricted to sample sizes above 2000 participants. We did not calculate the empirical power for the genetic case has it has already been shown to match the approximated power (25).

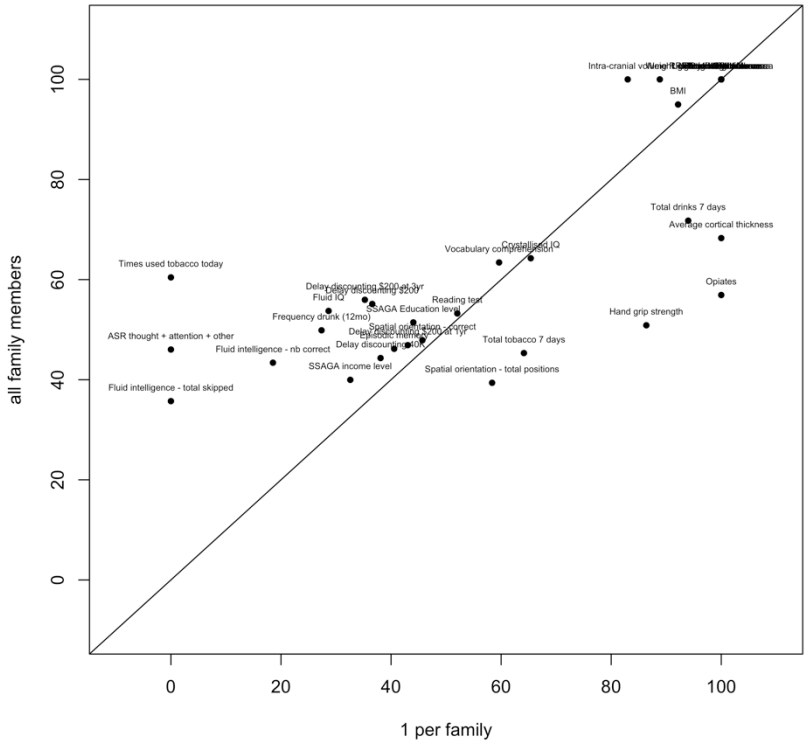


Appendix S7 Figure 2: Statistical power of detecting an association R^2>5%, with a risk alpha=0.05.
Power derived from simulation and from the approximation theory are compared. We did not calculate the empirical power for the genetic case has it has already been shown to match the approximated power (25).

Appendix S8: Including (or not) twins when estimating morphometricity

The HCP comprises many twin pairs (thus, non-independent observations), and we evaluated the impact of their exclusion on the morphometricity estimates. We found a good agreement of the results with and without including related individuals, with no evidence of systematic bias, though the limited power (N=444) results in wide SE (~0.20) (**Appendix S8 Figure1**).

The reason why LMM are robust to the presence of related individuals may be that they explicitly model the grey-matter relatedness in all analyses, which should account for the grey-matter resemblances arising from shared genetics or environment. For another confirmation, our results on the full HCP sample yielded always similar (e.g. Fluid IQ) or lower (e.g. attention) brain-morphometricity estimates than reported by Sabuncu et al., who selected 1 subject per family (29). To note, grey-matter similarities of twin pairs is not as outlying as their genetic relatedness factor (compared to the general population) (**Appendix 4**).



Appendix S8 Figure1: Morphometricity estimates in the HCP sample obtained using the full sample or a reduced set of unrelated individuals.
Only phenotypes with significant morphometricity are included here. Overall correlation between estimates is 0.8 with no evidence of systematic bias when including all family members. The largest differences were found for variables with little variation (opiates) or non-normally distributed (times used tobacco, ASR score, fluid intelligence – total skipped).

Appendix S9: In depth results and discussion of the ROI-traits associations using the vertex-wise resolution

In the UKB, the largest associations were observed between age of the participants and subcortical volumes (R^2 ranging between 0.22 and 0.35 for subcortical thickness, 0.20 0.38 for subcortical area), but most cortical regions were also significantly associated with age, albeit to a lesser extent (R^2 in the 0.0083-0.15 range for cortical thickness, 0.0048-0.15 range for cortical surface area). Next, significant ROI associations included sex, associated with all subcortical volumes (R^2 in the 0.0049-0.024 range for thickness, 0.0058-0.027 for area) and with many cortical regions (R^2 in the 0.0011-0.0076 range for cortical thickness, 0.0019-0.014 for cortical surface area) (**Figure S8** and **Dataset S7**). Maternal smoking around birth was further associated with 28 ROI, mostly located in the occipital and temporal lobes (R^2 in the 0.013-0.026 range with cortical thickness, R^2 in 0.014-0.071 with cortical surface and R^2 in the 0.010-0.039 range with subcortical structure). In addition, we found significant associations between cognition domains and structure of thalamus, putamen, pallidum and hippocampus (R^2 in the 0.0043-0.024 range). Notably, fluid intelligence was associated with all aspects of thalamus anatomy (left and right, thickness and surface area) while the other cognition domains considered were associated with some aspects of thalamus structure. No association between cognition and cortical structure survived multiple testing correction.

Diabetes diagnosis correlated with (left) superior frontal surface area (R^2=0.054), as well as with thalamus, putamen, and pallidum thickness (R^2 ranging between 0.0067 and 0.015), or thalamus and hippocampus surface (R^2 in the 0.0061-0.014 range). Alcohol intake was associated with left thalamus thickness (R^2=0.018) while smoking status and past tobacco use were associated with thalamus, caudate, putamen and pallidum thickness, as well as with thalamus surface area (R^2 in the 0.007-0.020 range). Finally, we also observed small associations between cortical or subcortical regions and overall health rating, time spend watching TV, body fat percentage and physiological measurements (**Figure S8**).

Using the replication UKB sample, we replicated 633 out of the 975 significant ROI-trait associations ($p<0.05/975$). Most associations were found with age, sex and body size variables, though we also replicated associations between subcortical volumes and hand grip strength or time spent watching TV (**Dataset S9**). In addition, the magnitude of the associations with age, and body size were greatly similar between discovery and replication analyses (**Appendix S9 Figure1**). For sex, we observed larger ROI associations in the UKB replication sample (**Appendix S9 Figure 1**), consistent with the larger brain-morphometricity observed in this sample (**Figure S6**).

In the HCP sample, age was associated with thickness (R^2 in the 0.020-0.049 range) and surface area (R^2 ranging between 0.067-0.10) throughout the cortex, as well as with subcortical structure (R^2 in the 0.016-0.087 range). Sex was associated with cortical thickness of the lateral orbitofrontal cortex (R^2 in the 0.059-0.073 range), as well as with subcortical structure (R^2 in the 0.042-0.19 range). In addition, we found large associations

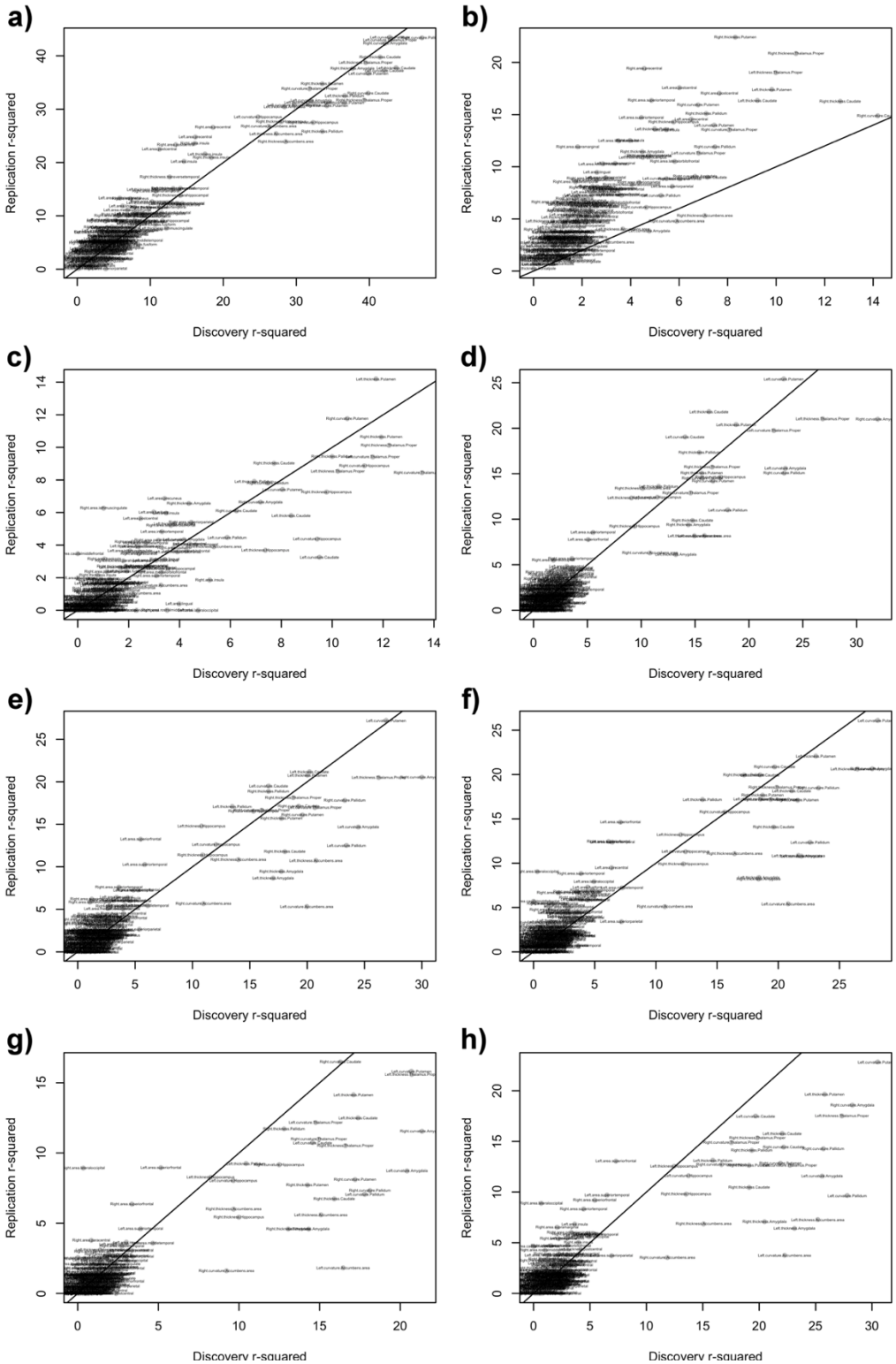
between cocaine, opiate or hallucinogens use and surface area of several cortical regions located in the temporal lobe (fusiform, superior temporal, insula), frontal (pars-triangularis, pars-opercularis, caudal-middle frontal), parietal (supramarginal, superior and inferior-parietal, precuneus) or in the cingulate (R^2 in the 0.25-1.00 range for cocaine test, R^2 in the 0.43-0.46 range for opiates, R^2 in 0.25-0.56 for number of times used hallucinogens). However, the small numbers and possible outliers in the vertex-wise measurements make such associations prone to false positives. Alcohol consumption was also associated with surface are of the frontal cortex (right rostral middle frontal, paracentral and precentral gyri, R^2 in the 0.28-0.36 range). No other association survived multiple testing correction (**Appendix S9 Figure2 and Dataset S10**).

Body size variables were strongly associated with subcortical structure under the baseline model (R^2 ranging between 0.010-0.059 for height, R^2 between 0.048-0.30 for the others) and to a lesser extent with cortical surface area (R^2 between 0.0078-0.026 for height, R^2 between 0.0061-0.060 for the others) and cortical thickness (R^2 in 0.0039 0.016 for height, R^2 in 0.0017 0.045 for the others). The associations between grey-matter structure and body size were pervasive (72/164 significant ROIs associations with height, 109 with waist circumference, 105 with BMI) (**Figure S9, Dataset S8**), suggesting that when acting as confounders height, weight or BMI could lead to false positives in many brain regions.

Discussion

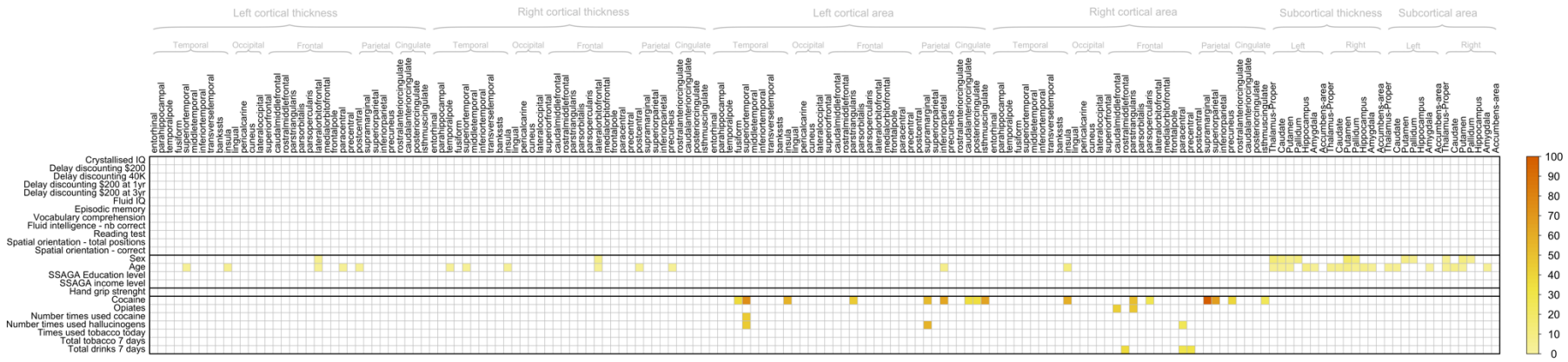
Our results for sex are consistent with results from the UKB first release (N=5,216, using ROI average (30)), while several studies have previously reported associations between BMI and several grey-matter measurements (31-36).

In the UKB, smoking status was associated with thickness and surface of the thalamus (left and right), although we also found associations with the caudate and pallidum. Previous studies have reported association between tobacco usage and volume of left thalamus (37-39), which might be due to faster age related volume loss in smokers (40). We did not replicate other cortical or subcortical associations previously reported (37, 38, 41). Alcohol intake was also associated with left thalamus thickness in the UKB, consistent with the significant grey-matter correlation (**Figure 3**) between the two traits. The thalamus has been implicated in alcohol-related neurological complications (e.g. Korsakoff's syndrome)(42) but may also be associated with regular alcohol usage (42, 43) or alcohol use disorder (44). Maternal smoking around birth was further associated with the thalamus, putamen, hippocampus and pallidum, as well as temporal and occipital ROIs. In addition, diagnosis of diabetes was associated with area of the left superior-frontal cortex (**Dataset S7, Figure S8**). Nervous system complications of diabetes (sometimes labelled diabetic encephalopathy) are widely accepted (45) but little is known about the specific brain regions associated with the condition (46).



Appendix S9 Figure 1: ROI based associations between UKB discovery and replication samples for selected phenotypes.

Each panel correspond to a single phenotype, and displays the association R2 of this phenotype with each ROI of interest considered. The results found in the replication sample (Y-axis) are plotted as a function of the discovery results (X-axis). The labels indicate which ROI corresponds to which point in the figure. Panel a) Age; b) sex; c) height; d) body fat %; e) BMI; f) weight; g) hip circumference; h) waist circumference



Appendix S9 Figure 2: Region Of Interest (ROI) based LMMs in the HCP.
Significant association R^2 between each UKB phenotype associated with grey-matter shape in Figure 1b and the grey-matter vertices from each of the Desikan atlas ROI.

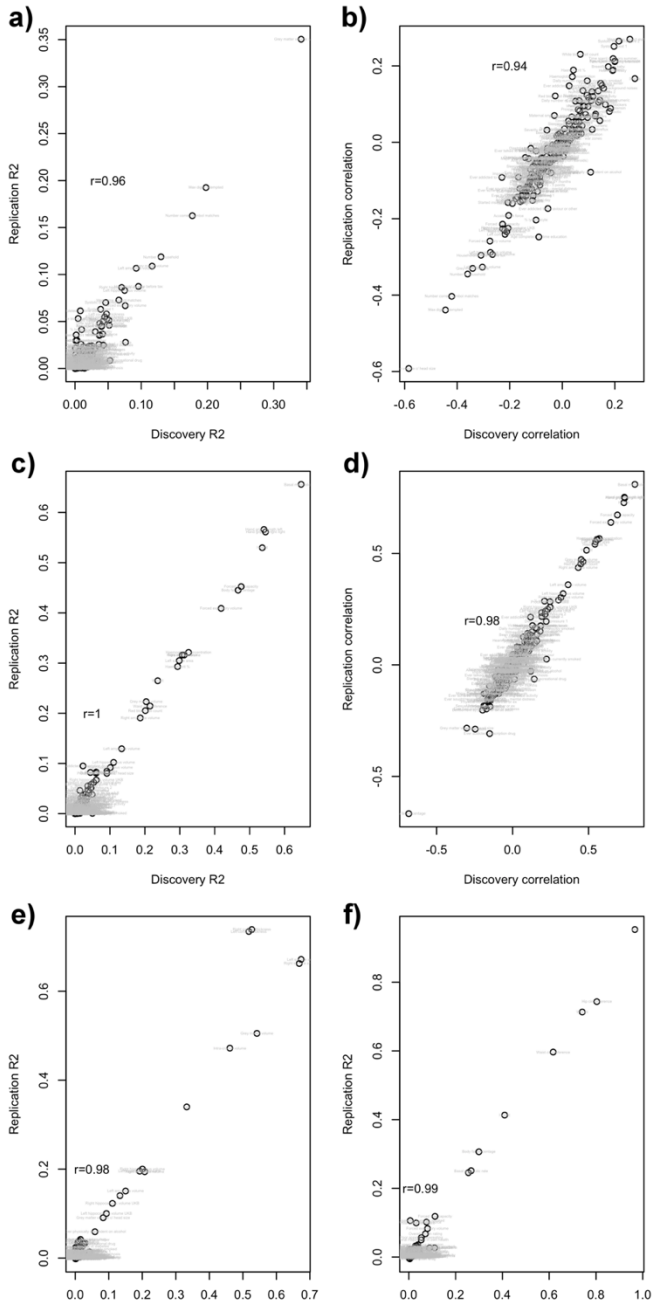


Fig. S1. Replication R2 (or correlations) are presented as a function of the discovery R2 (or correlation).

The R2 or correlations correspond to the fixed effect association between a covariate and all other phenotypes (labelled in plots). For example, panel a) shows the association R2 between age and all phenotypes, in the discovery and replication UKB samples. Panel b) shows the same results but using correlations and not R2 to appreciate the sign. Panels, c) and d) show the R2 and correlations between sex and all other variables. Panel e) presents the R2 between phenotypes and head size (ICV, left and right total thickness and area). Panel f) presents the R2 between phenotypes and body size (height, weight and BMI). As head size or body size are composed of several variables, only the R2 is presented. The correlation between discovery and replication results in shown in each panel as r.

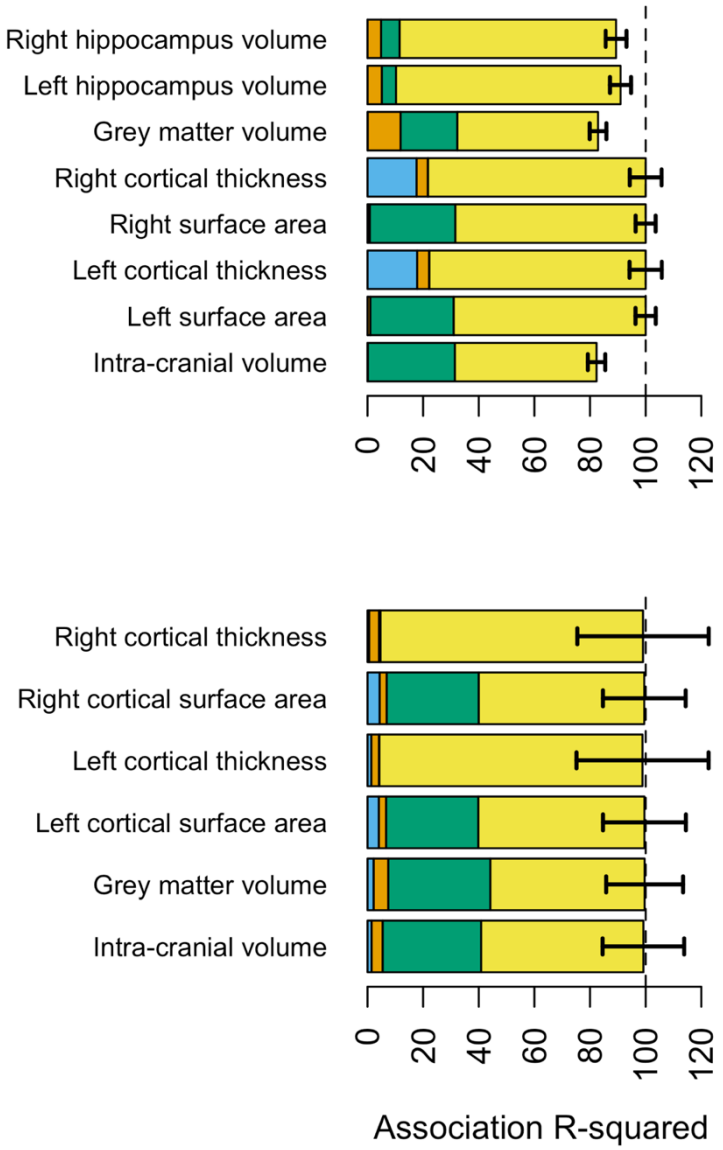


Fig. S2. Morphometricity of brain phenotypes – positive controls

As a positive control, we estimated the association between all grey-matter vertices and global measures of brain size, controlling for acquisition age and sex. Results are shown for the UKB discovery sample (top panel) and the HCP sample (bottom).

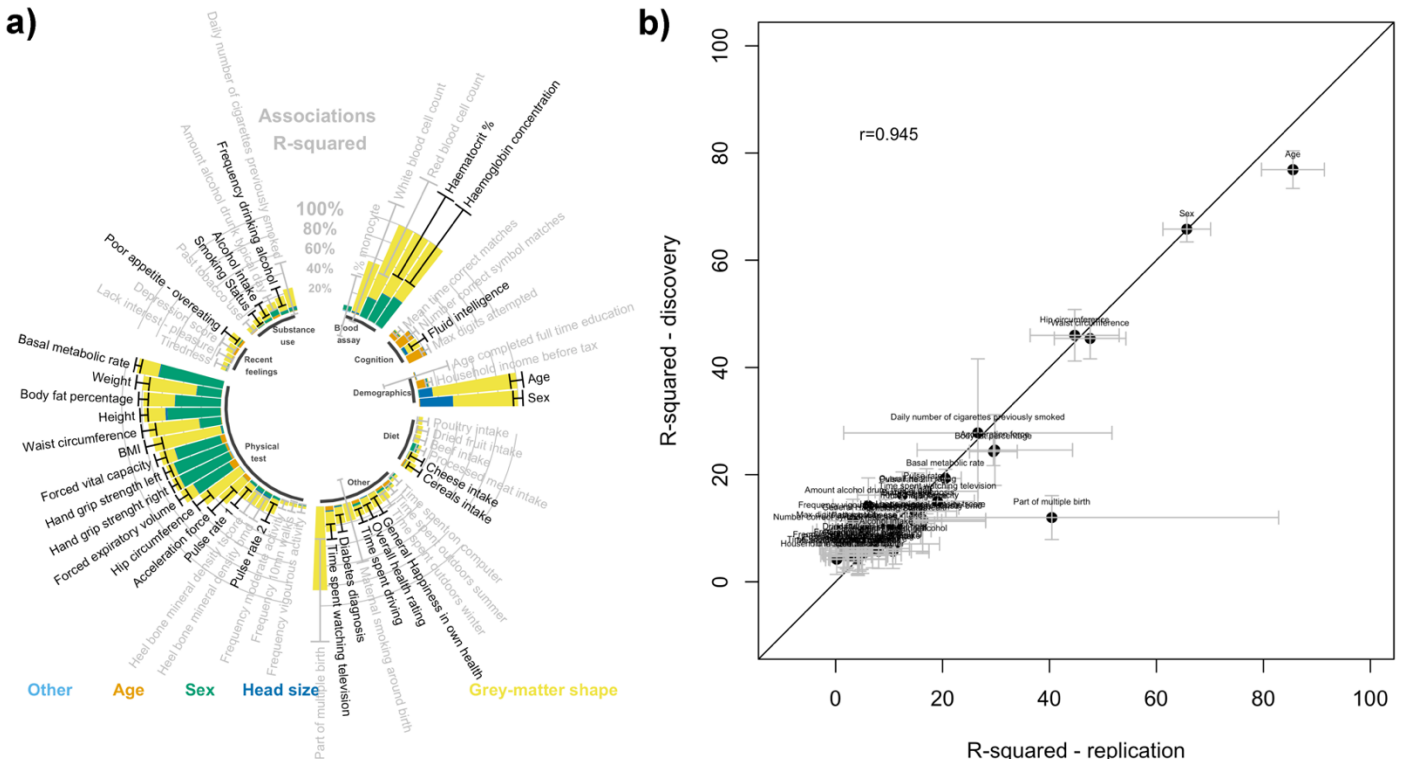


Fig. S3: Morphometricity in the UKB replication sample under baseline covariates.

Panel a) presents the summary of the replication analysis in the UKB. Only the 58 significant phenotypes from the UKB discovery sample were included in the analysis and are shown here. Black bars correspond to 95% confidence intervals. Phenotypes showing significant morphometricity in the replication sample ($p<0.05/58$) are labelled in black. To note, the replication sample was only half the size of the discovery one, with a clear incidence on power (and 95%CIs). Blood assay variables and being part of multiple birth suffered from a small number of observations (N~300), which explains the even larger confidence intervals. Panel b) compares the replication morphometricity estimates (X axis) to the discovery ones (Y-axis). We excluded blood assay variables, for which the small N led to unstable parameter estimation. We found a great concordance of results between the 2 independent UKB samples as indicated by a correlation of 0.95 between the 2 sets of results (panel b).

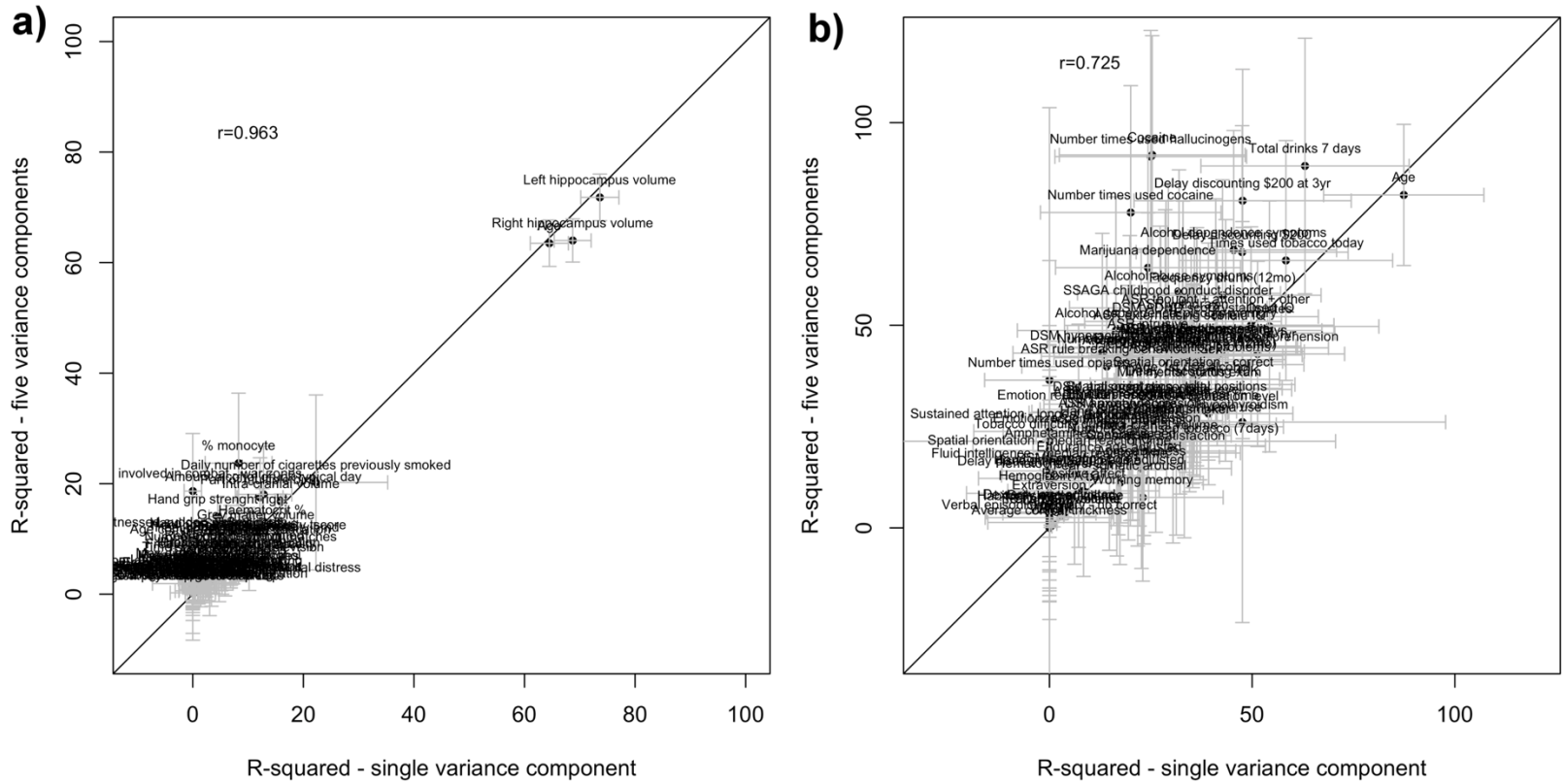


Fig. S4: Scatterplots of association R^2 from LMMs, comparing results when fitting a single BRM versus 4 BRMs corresponding to all different brain modalities (cortical thickness, cortical surface area, subcortical thickness and subcortical curvature).

Panel (a) shows the results for the UKB sample, panel (b) shows the results for the HCP sample. Note that fitting multiple variance components comes at an increased computational cost and a slightly increased standard error of the estimate of the overall variance explained(47). In addition, when modelling 5 variance components, the AI-REML algorithm failed to converge for 58 UKB phenotypes and 71 HCP phenotypes (in particular for phenotypes with low morphometricity) because of the increased uncertainty in the estimate of variance explained by each component due to smaller number of brain measurements in an individual component in comparison with the total. The correlation between the association R^2 from the two approaches appears on the plot.

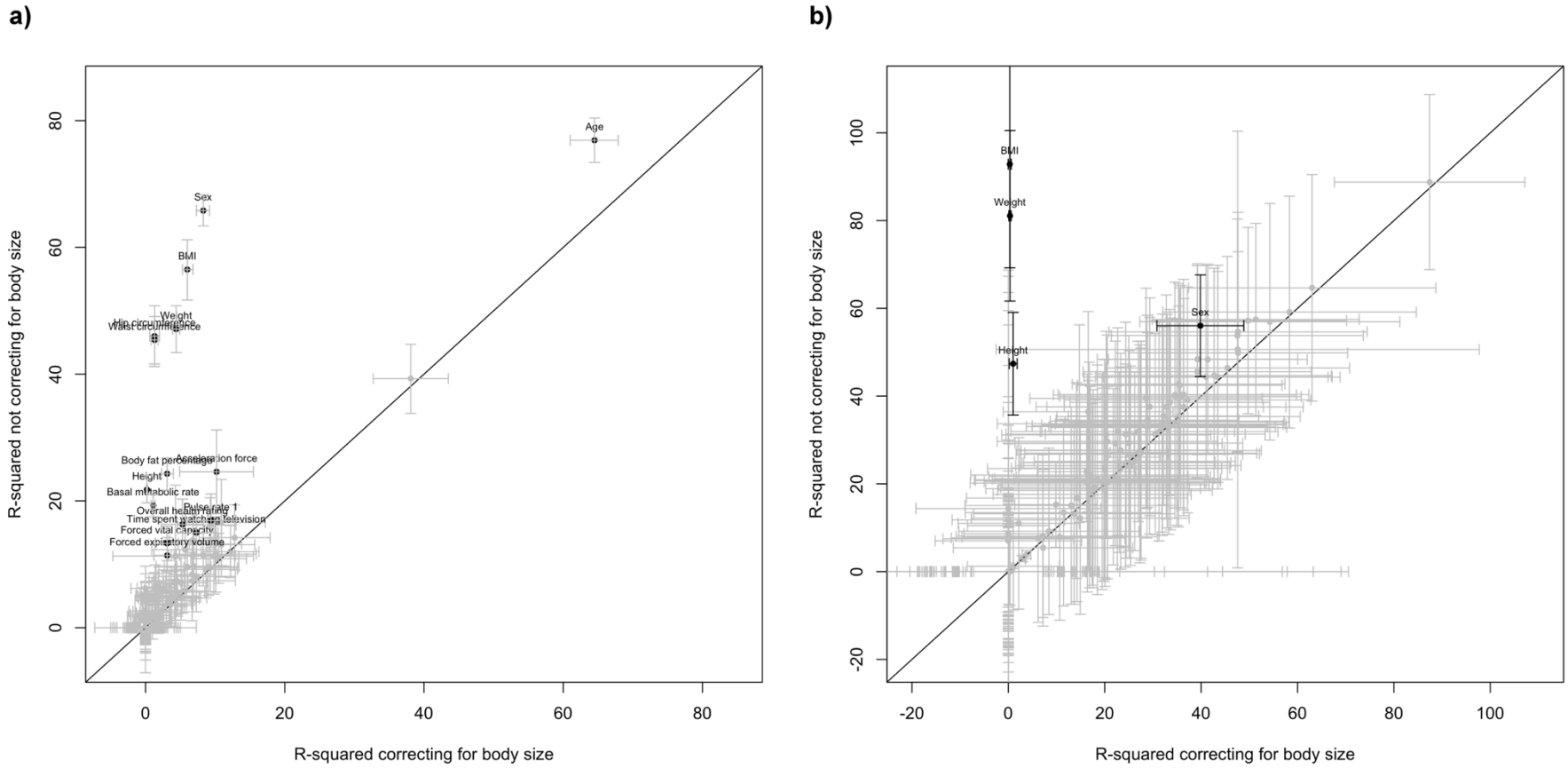


Fig. S5: Effect of correcting for body size on morphometricity estimates. Scatterplots of association R^2 for all the phenotypes before and after correcting for body-size variables.

Results are shown for the UKB discovery sample (panel (a)) and the HCP (panel (b)). Bars represent the 95% confidence intervals.

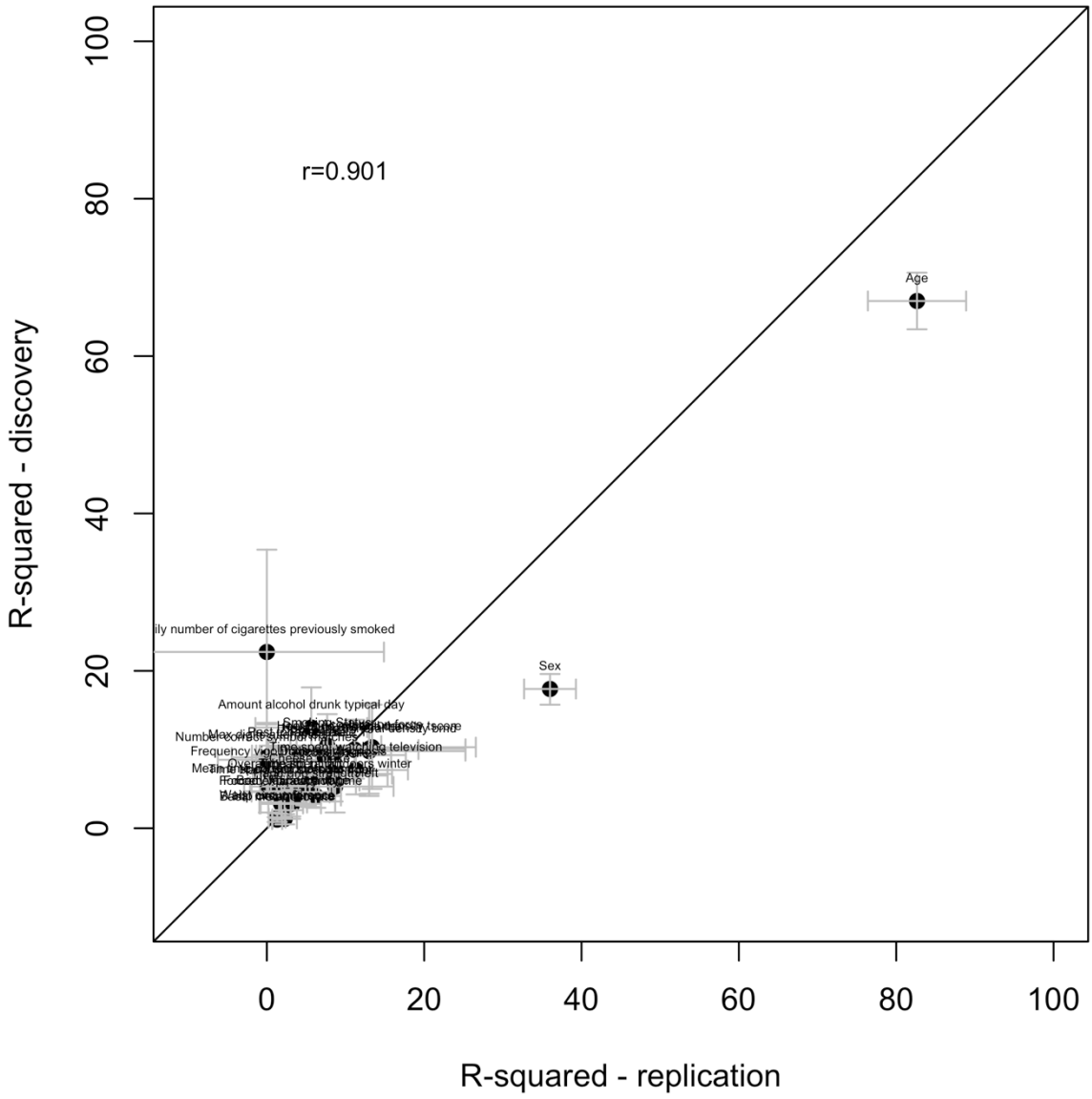


Fig. S6: Morphometricity estimates correcting for height, weight and BMI in addition to baseline covariates, in the UKB replication sample (X-axis) and the UKB discovery sample (Y-axis).

Only phenotypes displaying significant morphometricity in the UKB discovery set are included. Bars correspond to the 95% confidence intervals.

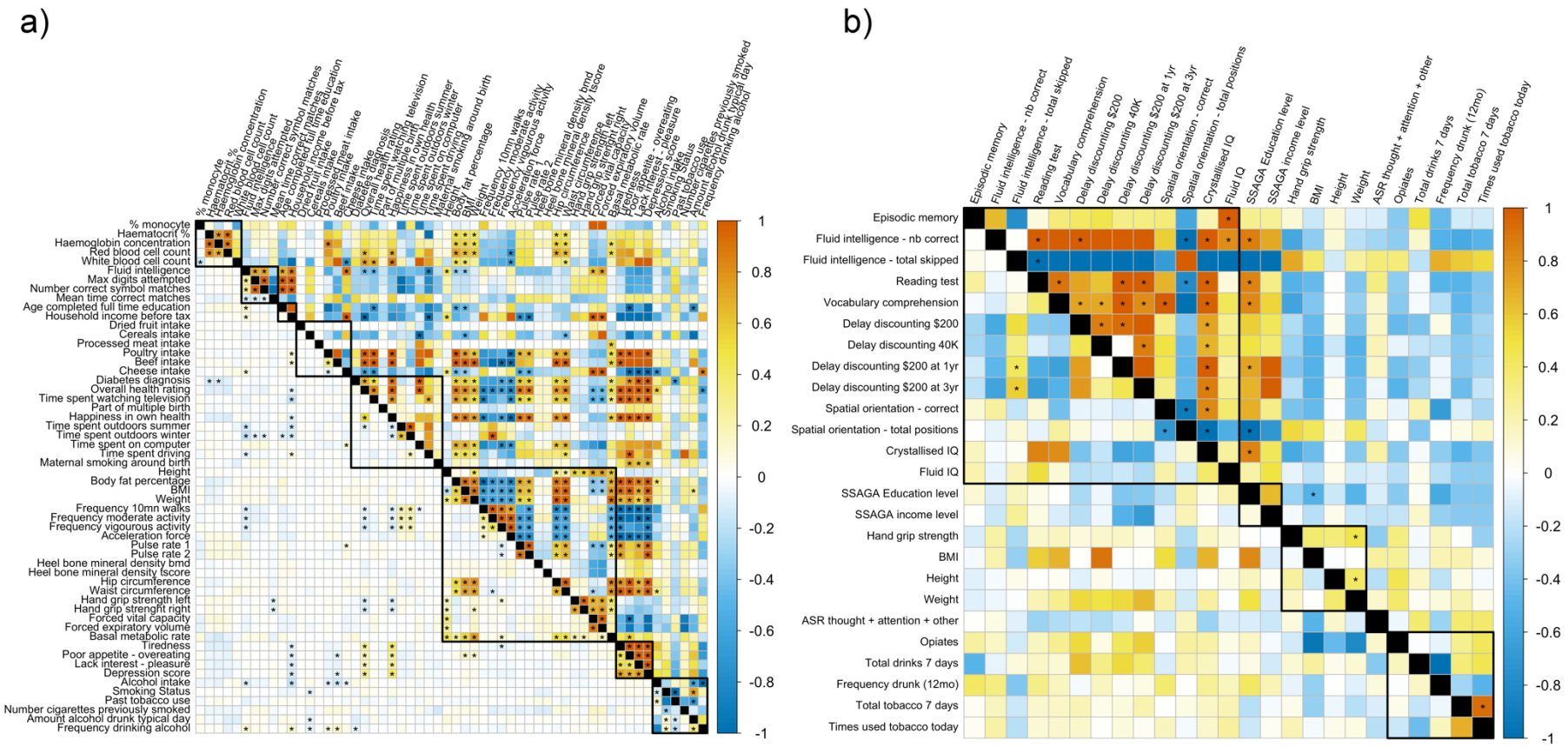


Fig. S7: Grey-matter and residual correlations under the baseline model (i.e. not correcting for body size).

Estimates are shown for the UKB (panel a) and HCP (b). Grey-matter correlation is shown above the diagonal, and residual correlation below the diagonal. Stars indicate significant correlations after multiple testing correction (Bonferroni).

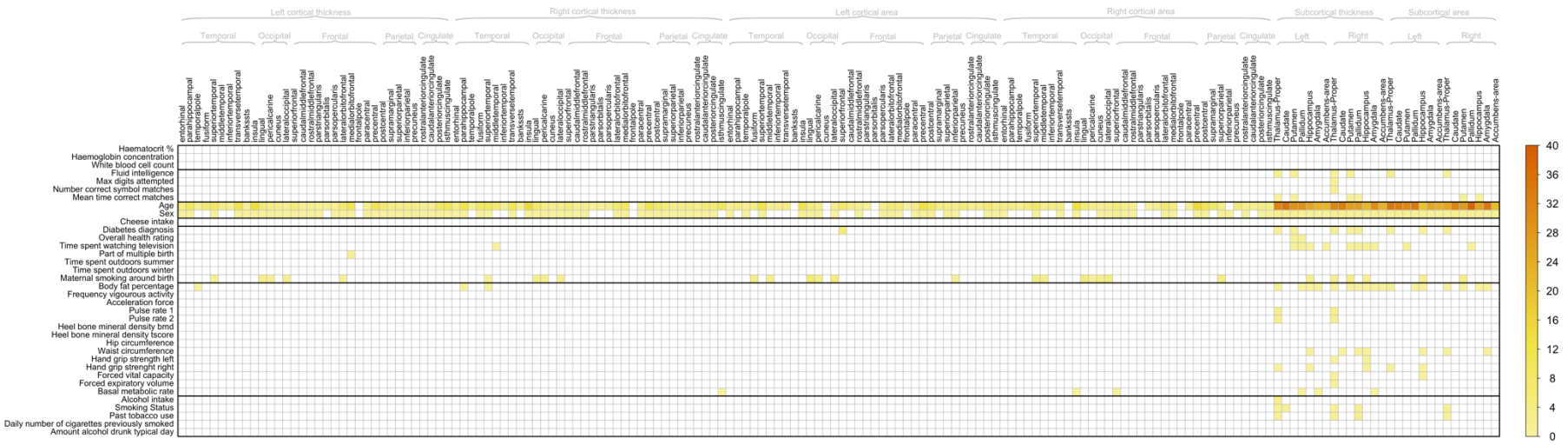


Fig. S8: Region Of Interest (ROI) based LMMs in the UKB.

Plot displays the significant association R2 between each UKB phenotype associated with grey-matter shape in Figure 1a and the grey-matter vertices from each of the Desikan (48) atlas ROI. Results include baseline covariates as well as height, weight and BMI.

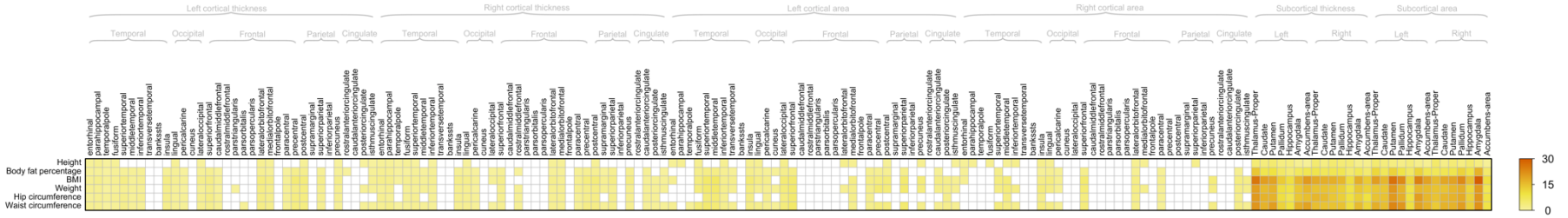


Fig. S9: Region Of Interest (ROI) based LMMs in the UKB for body size variables

Baseline covariates were used.

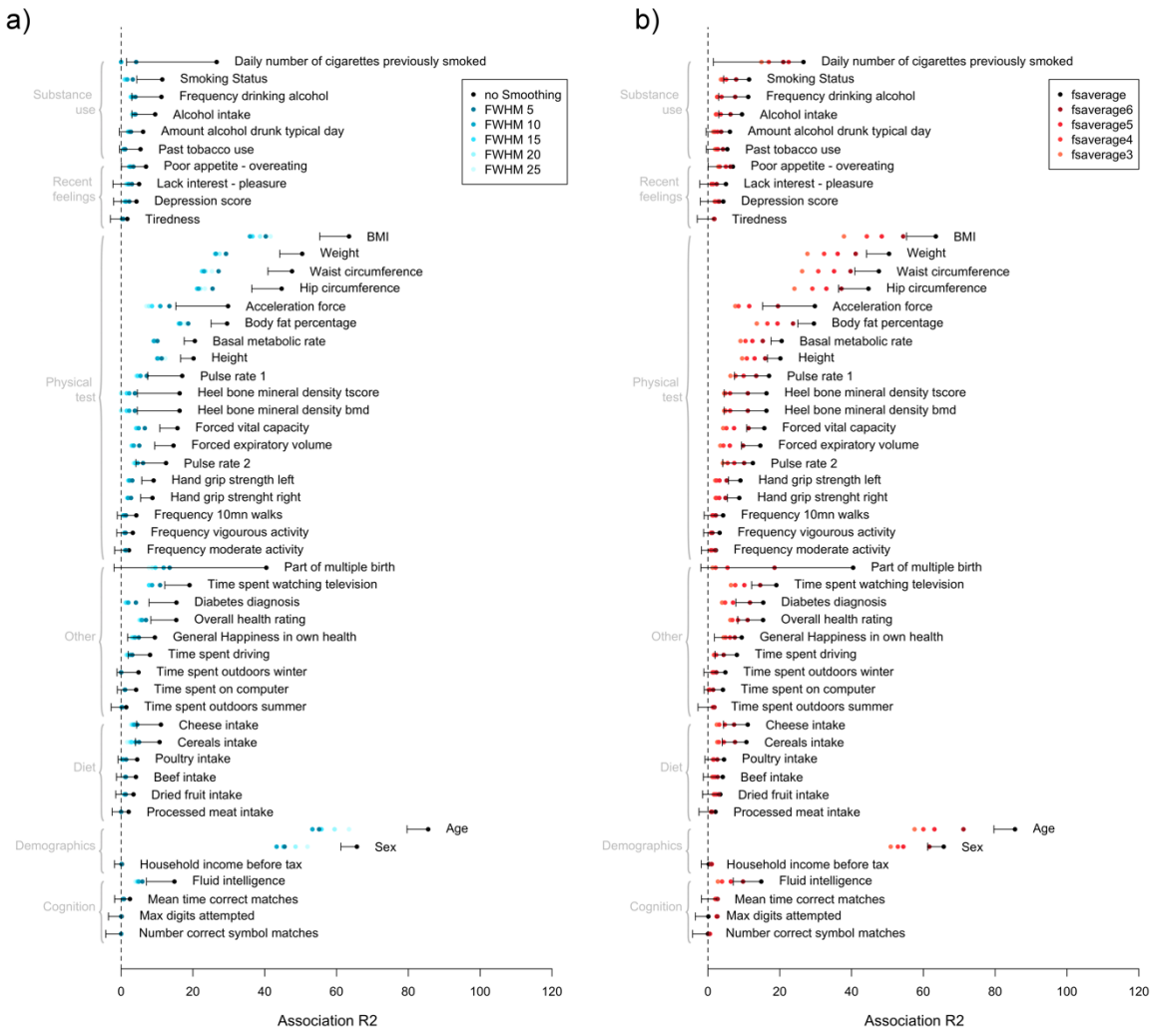


Figure S10: Effect of cortical mesh smoothing and mesh choice on the brain-morphometricity estimates (UKB replication samples).

Blood assay Variables (with n<500 observations) were excluded from the analysis due to unstable estimates and large SE.

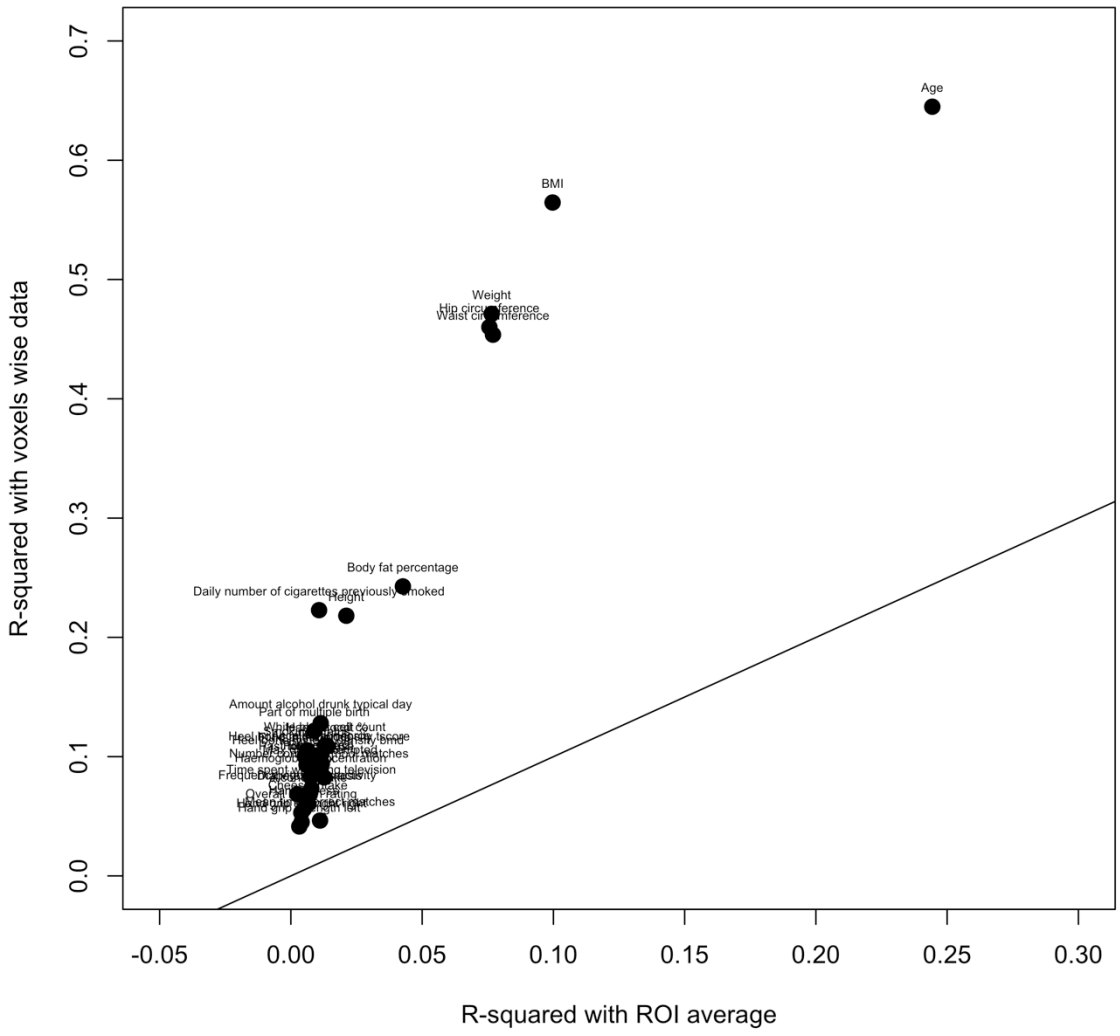


Fig. S11: Scatter plot comparing, for each UKB phenotype, our association R2 from vertex-wise processing and that obtained from standard ENIGMA ROI based processing.

All the points are above the diagonal indicate a greater amount of information retained using the vertex-wise processing compared to the ROI based dimension reduction approach. For example: $R^2_{age\_Vertex}$=0.64 vs. $R^2_{age\_ROI}$=0.24; $R^2_{CigarettesPreviouslySmoked\_Vertex}$=0.22, $R^2_{CigarettesPreviouslySmoked\_ROI}$=0.011, $R^2_{AmountDrunkTypicalDay\_Vertex}$=0.13, $R^2_{AmountDrunkTypicalDay\_ROI}$=0.011, and also $R^2_{BMI\_Vertex}$=0.56 vs. $R^2_{BMI\_ROI}$=0.10. Results are compared using baseline covariates.

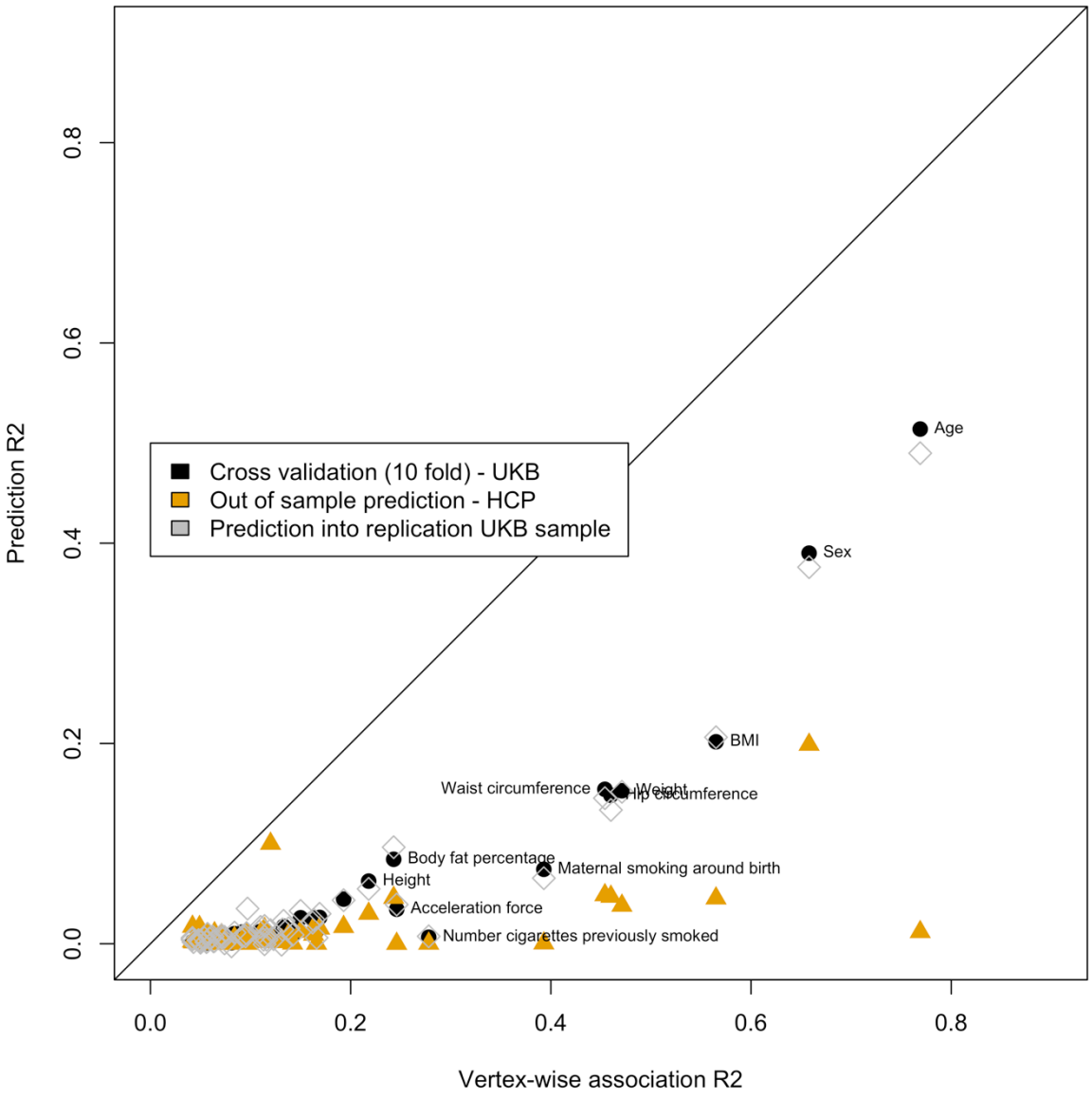


Figure S12: In sample and out of sample prediction accuracy as a function of the total association R^2 (baseline covariates)

Labels highlight some of the significant prediction.

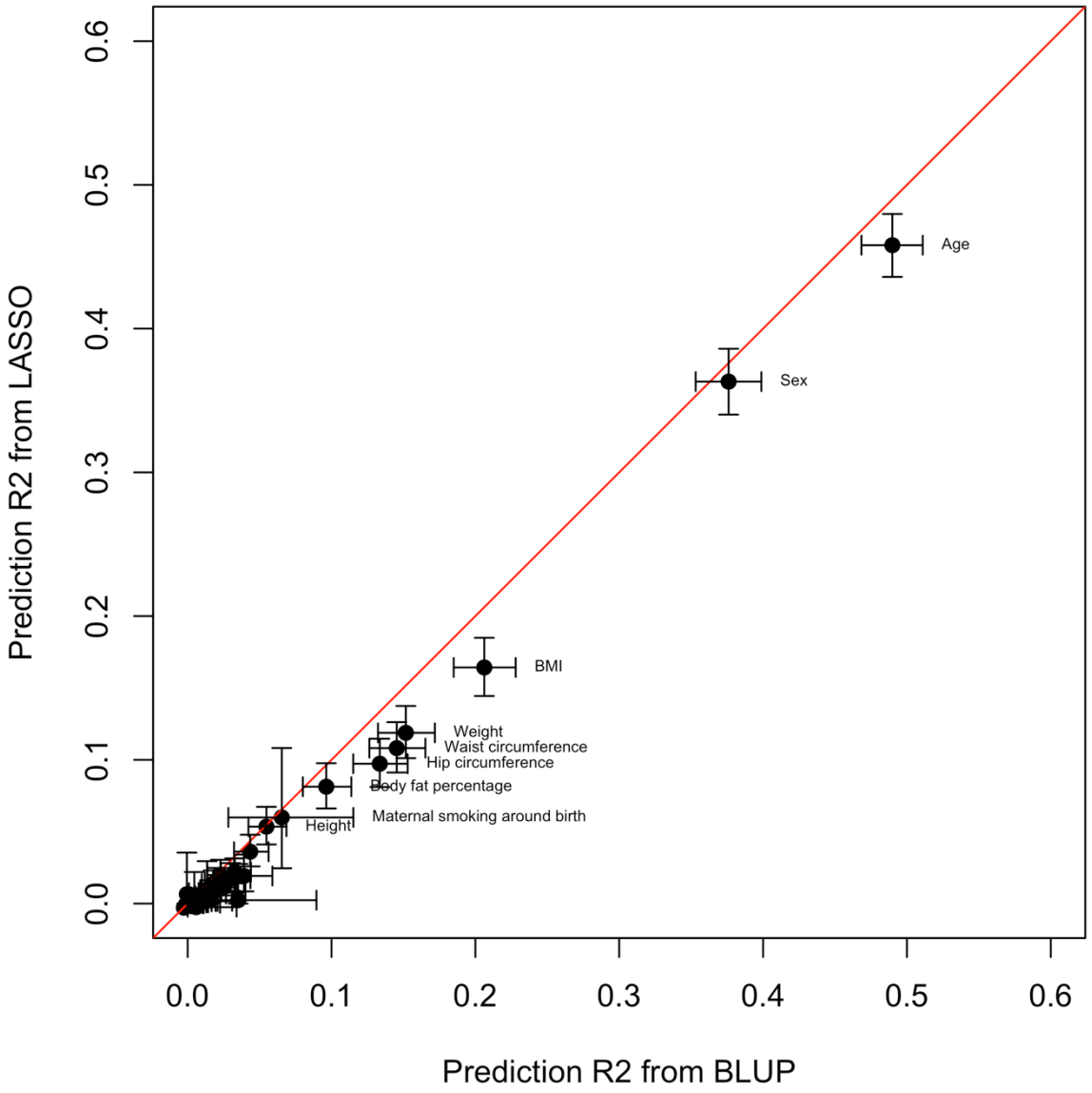


Figure S13: Prediction accuracy of BLUP and LASSO brain scores.

Models were trained and evaluated by controlling for baseline covariates on the UKB discovery sample. We evaluated prediction R^2 in the UKB replication sample. Confidence intervals (95%) are shown around the point estimates.

Table S1. Replication of grey-matter correlations identified in the UKB discovery sample

Variable 1	Variable 2	Discovery	Replication		
		rGM	rGM	SE	pvalue
Fluid intelligence	Max digits attempted	0.71	1.0	7.3	0.29469
Fluid intelligence	Number correct symbol matches	0.72	1.0	25.1	0.31246
Max digits attempted	Number correct symbol matches	1.0	1.0	25.1	0.31246
Fluid intelligence	Cheese intake	1.0	0.82	0.2	0.0015689
Time spent outdoors summer	Time spent outdoors winter	1.0	1.0	0.2	0.1928
Time spent watching television	Body fat percentage	0.73	0.81	0.2	4.42E-05
Frequency vigorous activity	Acceleration force	1.0	0.62	0.4	0.13604
Overall health rating	Pulse rate 1	1.0	1.0	0.5	0.01224
Overall health rating	Pulse rate 2	1.0	1.0	0.5	0.012028
Pulse rate 1	Pulse rate 2	0.99	1.0	0.01	0.00049312
Overall health rating	Waist circumference	1.0	0.39	0.4	0.19506
Body fat percentage	Waist circumference	0.52	0.45	0.2	0.02677
Pulse rate 1	Waist circumference	0.67	0.31	0.3	0.1675
Pulse rate 2	Waist circumference	0.73	0.56	0.3	0.047662
Hand grip strength left	Hand grip strength right	0.92	0.84	0.06	1.76E-07
Body fat percentage	Forced vital capacity	-0.66	-0.45	0.2	0.026155
Hand grip strength right	Forced vital capacity	0.69	0.70	0.2	0.0015424
Fluid intelligence	Forced expiratory volume	1.0	0.48	0.3	0.062686
Hand grip strength right	Forced expiratory volume	1.0	0.79	0.2	0.00044463
Body fat percentage	Basal metabolic rate	-0.69	-0.75	0.1	3.89E-06
Pulse rate 1	Basal metabolic rate	-0.55	-0.46	0.2	0.038444
Waist circumference	Basal metabolic rate	-0.75	-0.55	0.2	0.0055741
Smoking Status	Past tobacco use	-0.98	-0.93	0.03	0.0011549
Alcohol intake	Amount alcohol drunk typical day	-0.89	-1.0	0.1	0.00042018
Smoking Status	Amount alcohol drunk typical day	0.71	1.0	0.2	0.0044935
Past tobacco use	Amount alcohol drunk typical day	-0.64	-1.0	0.8	0.091905

Table S2: Summary of the prediction accuracy (R^2) of the BLUP grey-matter scores.

We constructed BLUP scores for the 39 UKB variables showing significant morphometricity and evaluated their predictive power in the UKB (10 fold-cross validation) and HCP sample. When the phenotype corresponding to the grey-matter score was not available in the HCP, we chose the closest available (e.g. waist circumference grey-matter score evaluated against BMI). We evaluate the prediction accuracy by fitting GLM controlling for height, weight and BMI as well as for the baseline covariates (acquisition, age, sex and head size); except for (#) denoting associations not controlling for height, weight and BMI. Rows in bold indicate significant association after correcting for multiple testing ($p<0.05/39=1.3e-3$) both in and out of sample.

	In sample prediction (UKB)				Prediction into UKB replication				Out of sample prediction (HCP)				
	r	**pvalue**	**R^2**	**AUC (SE)**	**r**	**pvalue**	**R^2**	**AUC (SE)**	**HCP variable predicted**	**r**	**pvalue**	**R^2**	**AUC (SE)**
Haemoglobin concentration	0.05	1.6e-05	0.0025		0.035	5.5e-01	0.0013		Hemoglobin A1C	0.015	6.8e-01	<0.001	
Haematocrit %	0.077	1.4e-10	0.006		0.037	5.7e-01	0.0013		Hematocrit level 1	0.061	2.1e-02	0.0037	
White blood cell count	0.042	2.2e-03	0.0018		0.016	3.4e-01	<0.001		Hematocrit level 1	-0.00051	9.8e-01	<0.001	
Mean time correct matches	0.046	2.0e-06	0.0021		0.053	2.9e-04	0.0028		Crystallised IQ	-0.046	9.7e-02	0.0021	
Number correct symbol matches	0.09	1.1e-13	0.0081		0.069	2.5e-04	0.0048		Crystallised IQ	0.064	2.2e-02	0.0041	
Max digits attempted	0.091	2.9e-14	0.0083		0.083	1.1e-05	0.0069		Crystallised IQ	0.069	1.3e-02	0.0048	
Fluid intelligence	0.077	4.1e-14	0.0059		0.11	7.2e-12	0.011		Fluid IQ	0.027	3.5e-01	<0.001	
Age	0.64	0.0e+00	0.41		0.68	0.0e+00	0.46		Age	0.15	3.1e-08	0.024	
Heel bone mineral density bmd	0.065	2.1e-06	0.0042		0.058	2.8e-03	0.0034		Age	-0.015	5.8e-01	<0.001	
Heel bone mineral density tscore	0.065	1.8e-06	0.0043		0.056	3.9e-03	0.0031		Age	-0.015	5.9e-01	<0.001	
Sex	0.26	0.0e+00	0.067	0.58 (0.0059)	0.33	9.8e-305	0.11	0.8 (0.0064)	Sex	-0.25	8.0e-42	0.061	0.68 (0.016)
Cheese intake	0.061	2.0e-09	0.0037		0.076	1.9e-07	0.0058		SSAGA Education level	0.029	3.2e-01	<0.001	
Part of multiple birth	0.078	4.1e-14	0.0061	0.66 (0.022)	0.13	1.5e-03	0.016	0.72 (0.065)	Being a twin	0.31	1.1e-28	0.098	0.69 (0.016)
Body fat percentage	0.29	0.0e+00	0.085		0.31	7.7e-190	0.095		BMI	0.21	5.6e-13	0.045	
Waist circumference	0.39	0.0e+00	0.16		0.38	2.0e-205	0.14		BMI	0.21	3.5e-13	0.046	

Frequency vigourous activity	0.044	9.4e-06	0.002		0.033	1.6e-02	0.0011		BMI	-0.0015	6.7e-01	<0.001	
BMI	0.45	0.0e+00	0.2		0.45	7.4e-235	0.20		BMI	0.21	2.4e-12	0.042	
Diabetes diagnosis	0.062	4.7e-10	0.0038	0.6 (0.014)	0.085	2.5e-09	0.0073	0.63 (0.019)	BMI	0.0055	1.0e-01	<0.001	
Overall health rating	0.052	8.2e-08	0.0027		0.062	7.8e-06	0.0039		BMI	-0.002	5.5e-01	<0.001	
Time spent watching television	0.083	3.0e-17	0.0068		0.12	1.8e-17	0.014		BMI	-0.0025	4.5e-01	<0.001	
Basal metabolic rate	0.029	1.0e-24	0.00082		0.031	3.7e-14	<0.001		BMI	0.0046	1.7e-01	<0.001	
Hip circumference	0.38	0.0e+00	0.15		0.36	7.3e-143	0.13		BMI	0.21	5.2e-13	0.045	
Time spent outdoors summer	0.047	2.1e-06	0.0022		0.033	1.3e-02	0.0011		BMI	-0.006	7.4e-02	<0.001	
Time spent outdoors winter	0.028	3.8e-03	0.00077		0.036	7.9e-03	0.0013		BMI	-0.0073	3.0e-02	<0.001	
Pulse rate 1	0.1	2.5e-23	0.01		0.13	4.0e-13	0.017		BMI	0.00012	9.7e-01	<0.001	
Pulse rate 2	0.091	2.5e-19	0.0083		0.1	9.0e-09	0.0099		BMI	-0.00057	8.7e-01	<0.001	
Height	0.25	6.5e-318	0.062		0.23	2.6e-132	0.054		Height	0.17	1.8e-17	0.03	
Acceleration force	0.07	7.5e-08	0.0049		0.098	4.1e-06	0.0096		Hand grip strength	0.014	4.5e-01	<0.001	
Hand grip strenght right	0.039	1.8e-09	0.0015		0.052	1.4e-08	0.0027		Hand grip strength	-0.0051	7.9e-01	<0.001	
Forced vital capacity	0.03	7.4e-06	0.00092		0.045	1.1e-05	0.002		Hand grip strength	0.0081	6.7e-01	<0.001	
Forced expiratory volume	0.027	2.3e-04	0.00072		0.06	7.8e-08	0.0036		Hand grip strength	0.019	3.3e-01	<0.001	
Hand grip strength left	0.039	1.6e-09	0.0015		0.064	2.9e-12	0.0041		Hand grip strength	-0.019	3.1e-01	<0.001	
Weight	0.39	0.0e+00	0.15		0.39	5.8e-231	0.15		Weight	0.19	1.2e-12	0.036	
Alcohol intake	0.074	1.0e-13	0.0055		0.096	2.9e-11	0.0092		Frequence alcohol use (12mo)	0.06	4.3e-02	0.0036	
Amount alcohol drunk typical day	0.063	5.9e-08	0.0039		0.075	6.1e-06	0.0056		Frequence alcohol use (12mo)	-0.0055	8.5e-01	<0.001	

Past tobacco use	0.043	1.7e-05	0.0019		0.076	2.2e-07	0.0057		FTND score	0.024	6.8e-01	<0.001	
Smoking Status	0.06	2.6e-09	0.0036		0.1	2.5e-12	0.01		FTND score	-0.013	8.3e-01	<0.001	
Number cigarettes previously smoked	0.066	1.0e-03	0.0043		0.053	4.9e-02	0.0028		FTND score	0.011	8.5e-01	<0.001	
Maternal smoking around birth	0.26	9.8e-132	0.069	0.66 (0.0067)	0.25	1.7e-08	0.063	0.65 (0.027)	FTND score	0.19	8.9e-04	0.037	

Table S3: Variables with significantly lower brain-morphometricity estimates compared to the fsaverage – no smoothing processing

We defined significance as non-overlapping confidence intervals of the morphometricity estimates in the UKB discovery sample. To account for multiple testing (9 tests for each of the 58 traits for which we observed significant morphometricity), we used 1-0.05/(9*58)=99.990% confidence intervals. We acknowledge this is a stringent definition of significance, which likely limited the statistical power of the test. Proper statistical testing was made difficult by the fact that the models are not nested, or that we did not know the covariance between the different morphometricity estimates.

	fwhm=5	fwhm=10	fwhm=15	fwhm=20	fwhm=25	Fsaverage6	Fsaverage5	Fsaverage4	Fsaverage3
Age_MRI	x	x	x				x	x	x
sex _f31_0_0	x	x	x	x	x		x	x	x
body_mass_index_bmi_f21001_2_0	x	x							
waist_circumference_f48_2_0	x	x							
weight_f21002_2_0	x	x							
hip_circumference_f49_2_0		x							
maternal_smoking_around_birth_f1787_2_0		x	x	x					x

Supplementary Dataset descriptions

Dataset S1: Descriptive table of the UKB variables used in the analysis (for the discovery and replication UKB sets). Comparison between final sample participants excluded from the analysis due to failed processing and QC. UKB discovery and replication samples used in the analyses are further compared.

Dataset S2: Descriptive table of the HCP variables used in the analysis and comparison with participants excluded from the analysis due to QC.

Dataset S3: Detailed results of variance components analysis in the UKB (discovery and replication samples). Includes results presented in **Figure 1a**. Results include fixed effect associations (associations with covariates) as well as morphometricity. Baseline covariates are used in the first part of the table but results after accounting for height, weight and BMI are also presented.

Dataset S4: Detailed results of variance components analysis in the HCP. Results of **Figure 1b.**

Dataset S5: Grey-matter and residual correlations in the UKB. Table of the values presented in **Figure 2a** (UKB). The table contains correlations and p-values of the phenotypic and brain correlations. Format: r; (SE; p-value).

Dataset S6: Grey-matter and residual correlations in the HCP. Table of the values presented in **Figure 2b** (HCP). The table contains correlations and p-values of the phenotypic and brain correlations. Format: r; (SE; p-value).

Dataset S7: Association R^2 between phenotypes and each ROI in the UKB discovery sample. Data is shown under the form of %variance (SE), p-value. This data corresponds to the **Figure S8**. Covariates include baseline + body size (height, weight and BMI).

Dataset S8: Association R^2 between body size phenotypes and each ROI in the UKB discovery sample. Data is shown under the form of %variance (SE), p-value. This data corresponds to the **Figure S9**. Baseline covariates only were used.

Dataset S9: Association R^2 between phenotypes and each ROI in the UKB replication sample. Data is shown under the form of %variance (SE), p-value.

Dataset S10: Association R^2 between phenotypes and each ROI in the HCP sample. Data is shown under the form of %variance (SE), p-value.

Dataset S11: Brain-morphometricity results varying coarseness of cortical meshes in the UKB discovery sample. They are plotted in **Figure 4**.

Dataset S12: Brain-morphometricity results varying smoothing and coarseness of cortical meshes in the UKB replication sample. All UKB phenotypes are included. They are presented in **Figure S10**.

Dataset S13: Accuracy from BLUP scores achieved using baseline covariates only. Results include 10-fold cross validation (UKB discovery), as well as prediction into UKB replication and HCP. They also include LASSO prediction into the UKB replication sample.

References

1. Miller KL, *et al.* (2016) Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* 19(11):1523-1536.
2. Mugler JP, 3rd & Brookeman JR (1990) Three-dimensional magnetization-prepared rapid gradient-echo imaging (3D MP RAGE). *Magn Reson Med* 15(1):152-157.
3. Mugler JP, 3rd, *et al.* (2000) Optimized single-slab three-dimensional spin-echo MR imaging of the brain. *Radiology* 216(3):891-899.
4. Van Essen DC, *et al.* (2013) The WU-Minn Human Connectome Project: an overview. *NeuroImage* 80:62-79.
5. Van Essen DC, *et al.* (2012) The Human Connectome Project: a data acquisition perspective. *NeuroImage* 62(4):2222-2231.
6. Glasser MF, *et al.* (2013) The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* 80:105-124.
7. Couvy-Duchesne B, *et al.* (2016) Head Motion and Inattention/Hyperactivity Share Common Genetic Influences: Implications for fMRI Studies of ADHD. *Plos One* 11(1).
8. Kong XZ, *et al.* (2014) Individual differences in impulsivity predict head motion during magnetic resonance imaging. *PLoS One* 9(8):e104989.
9. Siegel JS, *et al.* (2017) Data Quality Influences Observed Links Between Functional Connectivity and Behavior. *Cerebral Cortex* 27(9):4492-4502.
10. Van Essen DC, Glasser MF, Dierker DL, Harwell J, & Coalson T (2012) Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cereb Cortex* 22(10):2241-2262.
11. Marcus DS, *et al.* (2013) Human Connectome Project informatics: quality control, database services, and data visualization. *NeuroImage* 80:202-219.
12. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, & Smith SM (2012) FSL. *NeuroImage* 62(2):782-790.
13. Jenkinson M, Bannister P, Brady M, & Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* 17(2):825-841.
14. Fischl B (2012) FreeSurfer. *NeuroImage* 62(2):774-781.
15. Yang J, Lee SH, Goddard ME, & Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88(1):76-82.
16. Self SG & Liang KY (1987) Asymptotic Properties of Maximum-Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions. *J Am Stat Assoc* 82(398):605-610.
17. Stram DO & Lee JW (1994) Variance components testing in the longitudinal mixed effects model. *Biometrics* 50(4):1171-1177.
18. Pinheiro J & Bates D (2000) *Mixed-Effects Models in S and S-PLUS* (Springer New York).
19. Crainiceanu CM & Ruppert D (2004) Likelihood ratio tests in linear mixed models with one variance component. *J Roy Stat Soc B* 66:165-185.
20. Bates D, Machler M, Bolker B, & Walker S (2015) Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67:1-48.
21. Zhang F, *et al.* (2019) OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol* 20(1):107.

22. Bijma P & Bastiaansen JW (2014) Standard error of the genetic correlation: how much data do we need to estimate a purebred-crossbred genetic correlation? *Genetics Selection Evolution* 46(1):79.
23. Visscher PM (1998) On the sampling variance of intraclass correlations and genetic correlations. *Genetics* 149(3):1605-1614.
24. Lee SH, Yang J, Goddard ME, Visscher PM, & Wray NR (2012) Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 28(19):2540-2542.
25. Visscher PM, *et al.* (2014) Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data in Unrelated Samples. *Plos Genetics* 10(4).
26. Elston RC, Buxbaum S, Jacobs KB, & Olson JM (2000) Haseman and Elston revisited. *Genet Epidemiol* 19(1):1-17.
27. Vinkhuyzen AAE, Wray NR, Yang J, Goddard ME, & Visscher PM (2013) Estimation and Partitioning of Heritability in Human Populations using Whole Genome Analysis Methods. *Annual review of genetics* 47:75-95.
28. Yang J, Zeng J, Goddard ME, Wray NR, & Visscher PM (2017) Concepts, estimation and interpretation of SNP-based heritability. *Nat Genet* 49(9):1304-1310.
29. Sabuncu MR, *et al.* (2016) Morphometricity as a measure of the neuroanatomical signature of a trait. *P Natl Acad Sci USA* 113(39):E5749-E5756.
30. Ritchie SJ, *et al.* (2018) Sex Differences in the Adult Human Brain: Evidence from 5216 UK Biobank Participants. *Cereb Cortex* 28(8):2959-2975.
31. Cole JH, *et al.* (2013) Body mass index, but not FTO genotype or major depressive disorder, influences brain structure. *Neuroscience* 252:109-117.
32. Gupta A, *et al.* (2015) Patterns of brain structural connectivity differentiate normal weight from overweight subjects. *Neuroimage-Clin* 7:506-517.
33. Kurth F, *et al.* (2013) Relationships between gray matter, body mass index, and waist circumference in healthy adults. *Human Brain Mapping* 34(7):1737-1746.
34. Masouleh SK, *et al.* (2016) Higher body mass index in older adults is associated with lower gray matter volume: implications for memory performance. *Neurobiology of Aging* 40:1-10.
35. Medic N, *et al.* (2016) Increased body mass index is associated with specific regional alterations in brain structure. *Int J Obesity* 40(7):1177-1182.
36. Opel N, *et al.* (2017) Prefrontal gray matter volume mediates genetic risks for obesity. *Mol Psychiatr* 22(5):703-710.
37. Hanlon CA, *et al.* (2016) Lower subcortical gray matter volume in both younger smokers and established smokers relative to non-smokers. *Addict Biol* 21(1):185-195.
38. Gallinat J, *et al.* (2006) Smoking and structural brain deficits: a volumetric MR investigation. *Eur J Neurosci* 24(6):1744-1750.
39. Gillespie NA, *et al.* (2018) Testing associations between cannabis use and subcortical volumes in two large population-based samples. *Addiction* 0(ja).
40. Durazzo TC, Meyerhoff DJ, Yoder KK, & Murray DE (2017) Cigarette smoking is associated with amplified age-related volume loss in subcortical brain regions. *Drug Alcohol Depen* 177:228-236.
41. Prom-Wormley E, *et al.* (2015) Genetic and Environmental Contributions to the Relationships Between Brain Structure and Average Lifetime Cigarette Use. *Behavior Genetics* 45(2):157-170.

42. Pitel AL, Segobin SH, Ritz L, Eustache F, & Beaunieux H (2015) Thalamic abnormalities are a cardinal feature of alcohol-related brain dysfunction. *Neurosci Biobehav Rev* 54:38-45.
43. Cardenas VA, Studholme C, Gazdzinski S, Durazzo TC, & Meyerhoff DJ (2007) Deformation-based morphometry of brain changes in alcohol dependence and abstinence. *NeuroImage* 34(3):879-887.
44. van Holst RJ, de Ruiter MB, van den Brink W, Veltman DJ, & Goudriaan AE (2012) A voxel-based morphometry study comparing problem gamblers, alcohol abusers, and healthy controls. *Drug Alcohol Depen* 124(1-2):142-148.
45. Mijnhout GS, *et al.* (2006) Diabetic encephalopathy: A concept in need of a definition. *Diabetologia* 49(6):1447-1448.
46. Moheet A, Mangia S, & Seaquist ER (2015) Impact of diabetes on cognitive function and brain structure. *Ann N Y Acad Sci* 1353:60-71.
47. Yang J, *et al.* (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet* 47(10):1114-1120.
48. Desikan RS, *et al.* (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31(3):968-980.