Massively expedited genome-wide heritability analysis (MEGHA)

Citation

Permanent link

Terms of Use

Share Your Story

Massively expedited genome-wide heritability analysis (MEGHA)

Tian Ge[a,b,c,1], Thomas E. Nichols[d], Phil H. Lee[b,c], Avram J. Holmes[e], Joshua L. Roffman[f], Randy L. Buckner[a,f,g], Mert R. Sabuncu[a,h,2], and Jordan W. Smoller[b,c,1,2]

[a]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital/Harvard Medical School, Charlestown, MA 02129; [b]Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114; [c]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02138; [d]Department of Statistics & Warwick Manufacturing Group, The University of Warwick, Coventry CV4 7AL, United Kingdom; [e]Department of Psychology, Yale University, New Haven, CT 06520; [f]Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, MA 02114; [g]Department of Psychology and Center for Brain Science, Harvard University, Cambridge, MA 02138; and [h]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139

The discovery and prioritization of heritable phenotypes is a computational challenge in a variety of settings, including neuroimaging genetics and analyses of the vast phenotypic repositories in electronic health record systems and population-based biobanks. Classical estimates of heritability require twin or pedigree data, which can be costly and difficult to acquire. Genome-wide complex trait analysis is an alternative tool to compute heritability estimates from unrelated individuals, using genome-wide data that are increasingly ubiquitous, but is computationally demanding and becomes difficult to apply in evaluating very large numbers of phenotypes. Here we present a fast and accurate statistical method for high-dimensional heritability analysis using genome-wide SNP data from unrelated individuals, termed massively expedited genome-wide heritability analysis (MEGHA) and accompanying nonparametric sampling techniques that enable flexible inferences for arbitrary statistics of interest. MEGHA produces estimates and significance measures of heritability with several orders of magnitude less computational time than existing methods, making heritability-based prioritization of millions of phenotypes based on data from unrelated individuals tractable for the first time to our knowledge. As a demonstration of application, we conducted heritability analyses on global and local morphometric measurements derived from brain structural MRI scans, using genome-wide SNP data from 1,320 unrelated young healthy adults of non-Hispanic European ancestry. We also computed surface maps of heritability for cortical thickness measures and empirically localized cortical regions where thickness measures were significantly heritable. Our analyses demonstrate the unique capability of MEGHA for large-scale heritability-based screening and high-dimensional heritability profile construction.

heritability | genome-wide complex trait analysis | imaging genetics | endophenotype | phenomics

In quantitative genetics, the variance of a phenotype is commonly attributed to genetic components, environmental factors, and their interactions (1). The proportion of phenotypic variance captured by total additive (allelic) genetic effects is conceptualized as narrow-sense heritability. With the rapid expansion of comprehensive phenotypic data, practical tools to estimate heritability are invaluable as they can be used to prioritize high-dimensional phenotypes for genetic studies.

Classical estimates of narrow-sense heritability require twin or pedigree data (2–4), which can be costly and difficult to acquire. As genome-wide data became widely available, genome-wide complex trait analysis (GCTA) (5, 6) was developed, which assesses aggregate effects of common SNPs spanning the genome on phenotypes and thus provides an SNP-based heritability estimate, a lower bound for narrow-sense heritability. This method has been successfully applied to the heritability analyses of several complex traits and mental disorders (5, 7, 8) and has been used to investigate the puzzle of "missing heritability" (5, 9, 10).

However, GCTA is a computationally expensive procedure. The use of a time-consuming iterative optimization procedure in the fitting of variance component models makes it prohibitive to use for evaluating very large numbers of phenotypes or with nonparametric sampling techniques, such as permutation tests. More practical and computationally efficient methods are needed to facilitate the identification of phenotypes that are most appropriate for genetic studies especially in instances where the complexity of the phenotype provides thousands or even millions of options.

Here we present a fast and accurate statistical method for heritability analysis using genome-wide SNP data from unrelated individuals, which we call massively expedited genome-wide heritability analysis (MEGHA). MEGHA largely falls in the kernel machines framework (11), which subsumes the GCTA model as a special case and uses a variance component score test (12), known as sequence kernel association test (SKAT) (13–15), for efficient statistical inferences. MEGHA provides both magnitude estimates and significance measures of heritability with orders of magnitude less computational effort relative to GCTA, making it possible to analyze millions of phenotypes and develop sampling techniques that produce accurate inferences for

Significance

Practical tools for high-dimensional heritability-based screening are invaluable for prioritizing phenotypes for genetic studies with the dramatic expansion of available phenotypic data. Classical estimates of heritability require twin or pedigree data, which can be costly and difficult to acquire. Alternative methods based on whole-genome data from unrelated individuals exist but are computationally expensive. Here we present a novel, fast, and accurate statistical method for massively expedited genome-wide heritability analysis, making heritability-based prioritization of millions of phenotypes based on data from unrelated individuals tractable for the first time to our knowledge. We apply our method to large-scale heritability analyses of brain imaging measurements and demonstrate its potential for facilitating phenome-wide analyses and characterizing the genetic architecture of complex traits.

arbitrary statistics of interest and accommodate complex correlation structures within phenotypic data.

As a demonstration of application, MEGHA was applied to brain structural MRI and genome-wide SNP data from 1,320 unrelated subjects, as part of the Harvard/Massachusetts General Hospital (MGH) Brain Genomics Superstruct Project (GSP) (16). Brain imaging data are a prototype case where a vast array of potentially relevant phenotypes are routinely collected and phenotypic complexity has grown exponentially as new tools to analyze high-resolution structure and point-to-point connectivity have emerged. A wide range of volume-, surface-, and connection-based measures are of potential interest in analyzing the relationship between genetic and brain data in the context of clinical conditions (17–21). Although, in principle, any measure computable from different brain imaging modalities can be used as phenotypes in genetic studies, ideal candidate imaging traits should be heritable intermediate (or endo-) phenotypes (22–24), to uncover the genetic underpinnings of various neuropsychiatric disorders or biological processes of interest (25). However, due to the computational complexity and inordinate options of brain imaging measurements, few tools exist to enable efficient heritability-based screening of these phenotypes (26, 27), and the exploration of their genetic basis has been limited to a small subset of the search space. All high-dimensional (whole-brain, voxel-/vertex-wise) heritability maps computed to date have relied on twin or pedigree data (28–32). MEGHA may thus offer a powerful method for large-scale heritability screening and high-resolution heritability profile construction in imaging genetics.

Results

Table 1 shows the SNP-based heritability estimates of a number of global morphometric measurements, including intracranial volume (ICV; i.e., head size), total brain volume, left/right hemispheric cortical gray matter (GM) volume, total cortical GM volume, total subcortical GM volume, total GM volume, left/right hemispheric white matter (WM) volume, total WM volume, left/right hemispheric mean cortical thickness, overall mean cortical thickness, left/right hemispheric total surface area, and total surface area. The test–retest reliability of these measurements measured by correlation coefficient was computed

using 42 individuals that had repeated brain MRI scans on separate days. All measurements show high test–retest reliability. ICV, total brain volume, and mean cortical thickness measures are highly heritable, with familywise error corrected (FWEc) significant P values computed by the proposed permutation procedure (*Materials and Methods*). Cortical GM volumes are also heritable, with uncorrected significant P values. Subcortical GM volume, WM volumes, and surface area measures show moderate heritability. The proposed permutation procedure implicitly accounts for the correlation structure among measurements and provides more accurate FWEc P values (based on one million permutations) than Bonferroni-corrected GCTA P values. MEGHA estimates of heritability magnitude are tightly correlated with GCTA results (Fig. 1).

We next applied both MEGHA and GCTA to the heritability analyses of average cortical thickness measures in 68 regions of interest (ROIs; 34 ROIs per hemisphere) defined by the Desikan–Killiany atlas (33), producing SNP-based heritability estimates and significance measures (Table S1). The MEGHA heritability estimates, P values, and permutation P values (based on one million permutations) show excellent concordance with the GCTA results (Fig. 2 and Fig. S1). Four brain regions—the bilateral superior parietal cortex, the left precuneus cortex, and the left rostral anterior cingulate cortex—are significantly heritable after multiple testing corrections over all of the ROIs. The right precuneus cortex and the right supramarginal gyrus are marginally significant with FWEc ($P < 0.10$).

As shown in Table 2, an analysis that involves 50 or 100 phenotypes would be easily handled by both MEGHA and GCTA, although MEGHA is hundreds of times faster (cases 1 and 2). For example, it took ~400 s for GCTA to compute the P values for all of the 68 ROIs, whereas MEGHA required less than 1 s with a MATLAB implementation on a MacBook Pro with 8 GB of memory and a 2.4-GHz Intel Core i7 processor. The dramatic improvement of MEGHA in computational efficiency makes it possible for high-dimensional heritability screening and mapping (case 3), for inferences on arbitrary statistics of interest based on thousands of permutations (case 4), and even for a combination of both (case 5). Using GCTA in any of these analyses would require months, years, or even decades

Table 1. Analysis of global morphometric measurements

| Measurement | Test–retest reliability | GCTA | | | | MEGHA | | | |
		$\hat{h}^2$	SE	P value	FWEc P value (Bonf)	$\hat{h}^2$	P value	Perm P value	FWEc P value (Perm)
Intracranial volume (ICV)	0.995	0.849	0.275	6.95E-4	*0.011*	0.804	3.91E-4	4.16E-4	*4.21E-3*
Total brain volume	0.997	0.981	0.273	1.07E-4	*1.71E-3*	0.929	5.24E-5	7.00E-5	*6.32E-4*
Left hemispheric cortical GM volume	0.992	0.521	0.279	0.033	0.529	0.432	0.036	0.036	0.265
Right hemispheric cortical GM volume	0.991	0.492	0.279	0.041	0.652	0.411	0.043	0.043	0.307
Total cortical GM volume	0.994	0.515	0.279	0.034	0.550	0.429	0.037	0.037	0.270
Total subcortical GM volume	0.968	0.357	0.279	0.104	1.000	0.298	0.107	0.107	0.587
Total GM volume	0.995	0.475	0.279	0.050	0.796	0.382	0.055	0.055	0.374
Left hemispheric WM volume	0.996	0.416	0.279	0.071	1.000	0.344	0.075	0.076	0.467
Right hemispheric WM volume	0.996	0.302	0.279	0.140	1.000	0.257	0.141	0.142	0.691
Total WM volume	0.996	0.369	0.279	0.095	1.000	0.310	0.098	0.098	0.556
Left hemispheric mean cortical thickness	0.899	0.688	0.277	5.60E-3	0.090	0.625	4.54E-3	4.83E-3	*0.043*
Right hemispheric mean cortical thickness	0.885	0.732	0.277	3.48E-3	0.056	0.662	2.84E-3	2.99E-3	*0.027*
Overall mean cortical thickness	0.935	0.734	0.277	3.41E-3	0.055	0.665	2.73E-3	2.93E-3	*0.027*
Left hemispheric total surface area	0.999	0.298	0.279	0.137	1.000	0.270	0.130	0.129	0.658
Right hemispheric total surface area	0.997	0.288	0.279	0.139	1.000	0.274	0.127	0.126	0.650
Total surface area	0.998	0.305	0.279	0.128	1.000	0.283	0.118	0.118	0.625

Test–retest reliability of the measurements measured by correlation coefficient is computed using 42 individuals with repeated brain structural MRI scans on separate days. The heritability estimate $\hat{h}^2$ with the corresponding SE and P value obtained by GCTA, the familywise error corrected (FWEc) GCTA P value using Bonferroni correction, the heritability estimate $\hat{h}^2$ and P value computed by MEGHA, and the uncorrected and corrected P values obtained by the proposed permutation procedure based on one million permutations are provided. FWEc significant P values (<0.05) are shown in italic.
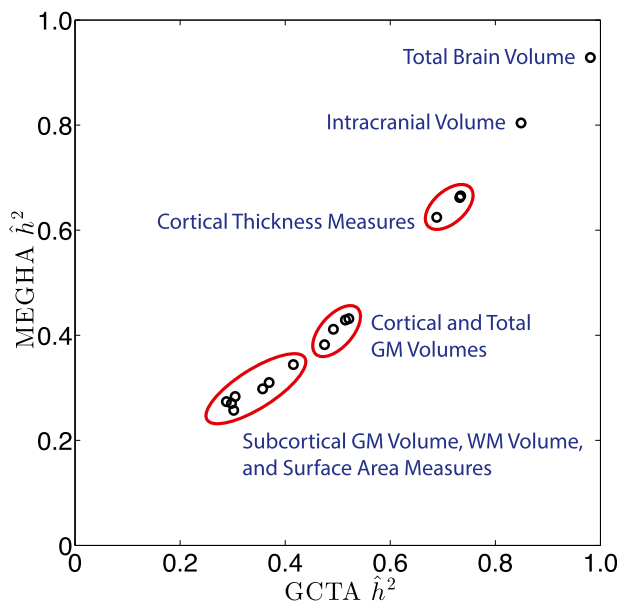
Fig. 1. MEGHA SNP-based heritability estimates for global morphometric measurements are plotted against GCTA results.

of computational time, and could be prohibitively slow even if parallel computation is used.

As a demonstration of the usefulness and flexibility of MEGHA in high-dimensional heritability analyses, we conducted vertex-wise MEGHA of cortical thickness measures to produce high-resolution surface maps for SNP-based heritability estimates (Fig. S2) and significance (Fig. 3). (Also see Fig. S2 for spatial heritability maps of sulcal depth, curvature, and cortical surface area measures.) We then performed surface-based clustering on the significance map, using $P = 0.01$ as a cluster-forming threshold, to localize heritable regions of cortical thickness. These empirically identified clusters are typically not aligned with sulcal/gyral patterns or predefined anatomical/functional ROIs. Cluster inferences (18, 34, 35) using the proposed permutation procedure identified five clusters (white outlined and annotated in Fig. 3) with FWEc significance over the entire cortical surface based on 1,000 permutations. Cluster 1, the largest cluster identified comprising 6,518 vertices with a FWEc $P < 0.001$, spans the left superior parietal cortex, cuneus, precuneus, and the left posterior cingulate cortex. Cluster 2 (FWEc, $P = 0.003$) and cluster 3 (FWEc, $P = 0.015$) largely overlap with the left precentral/postcentral cortex and the left superior temporal cortex, respectively. Clusters 4 and 5 are located on the right hemisphere. Specifically, cluster 4 (FWEc, $P = 0.004$) spans the right supramarginal cortex and the lateral part of the precentral/postcentral cortex. Last, cluster 5, which comprises 4,523 vertices with a FWEc $P < 0.001$, extends from the right superior parietal cortex to the right cuneus and precuneus.

Discussion

In this paper, we present MEGHA, a fast and accurate statistical method for heritability analysis using genome-wide SNP data from unrelated individuals. Our method has excellent concordance with GCTA, but is thousands of times faster. This computational efficiency allows for examination of complex phenotypes that have millions of combinations, and the development of nonparametric sampling techniques such as permutation tests and Jackknife resampling that can produce accurate and flexible inferences for arbitrary statistics of interest. As a case study of its application, MEGHA was used to prioritize brain structural MRI phenotypes based on heritability. We conducted global, regional, and vertex-wise heritability analyses of cortical thickness measures,

which empirically identified significantly heritable regions in superior parietal, precuneus, precentral/postcentral, superior temporal, and visual cortex. Prior studies have also published heritability maps of brain image derived phenotypes, where cortical thickness measures are under substantial genetic influences (28–31). These studies have relied on twin or pedigree samples, each spanning a different age range. In particular, twin-modeling results reported in ref. 30 are the most pertinent to our analyses, as the study sample consisted of young adults with an age span similar to our sample. The heritability maps of thickness presented in ref. 30 are similar to the results observed here based on genotypic data. Several differences do emerge such as the present results emphasizing significantly heritable clusters in bilateral association regions of the parietal cortex extending into precuneus. One explanation for these differences might be that MEGHA only captures SNP-based heritability due to common variation, whereas twin or pedigree based analyses can capture components of heritability due to rare variation.

Methodological Assessment. Although MEGHA provides estimates of heritability magnitude, the values need to be interpreted with caution. The reason is that when searching over a large number of phenotypes, the heritability estimates ranked at the top are highly likely to be inflated by noise. Reporting heritability magnitude extracted from significantly heritable brain regions is also invalid, representing a general problem of double dipping (36, 37). For this reason, although we demonstrate the usefulness of MEGHA to screen heritability of large numbers of phenotypes, we recommend deriving unbiased estimates of the magnitude of heritability in independent, replicate datasets.

We also note that heritability estimates and significant measures can be affected by the reliability of extracted measurements. Caution is thus needed when comparing heritability estimates across different measurements that are computed using different techniques. Although it appears from Fig. S2 that cortical thickness measures are more heritable than other morphometric features, this may be partly due to the fact that cortical surface area measures have much lower vertex-wise test–retest reliability than cortical thickness and sulcal depth measures in this particular data set.

A completely empirical permutation procedure to assess heritability significance would have to break the association between phenotypes and genotypes while retaining the observed phenotypic correlation structure. To the best of our knowledge, no strategy exists to achieve this requirement. The permutation procedure designed in this paper relies on the assumption that the linear mixed effects model is a good description of the data, making the score test valid and accurate. For example, the permutation inference is only valid when the model residuals under the null hypothesis (no additive genetic effects) can be approximately treated as independent and identically Gaussian
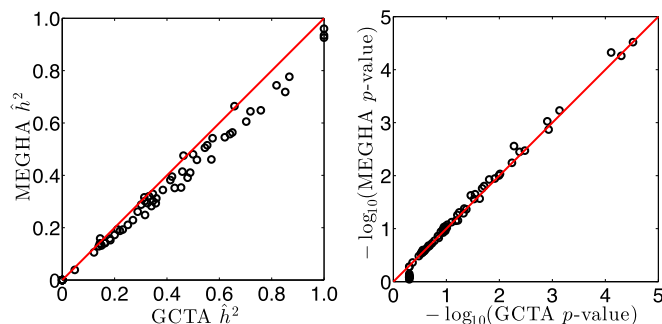

Fig. 2. Evaluation of MEGHA using average cortical thickness measures in 68 regions of interest (ROIs). MEGHA SNP-based heritability estimates (*Left*) and *P* values (*Right*) are plotted against GCTA results for each ROI. There is an excellent concordance between MEGHA and GCTA.

Table 2. Comparison of the computational time of MEGHA and GCTA when applied to different types of analyses

Case	Type of analysis	Effective no. of phenotypes	GCTA	MEGHA
1	Analysis of global morphometric measurements	16	120 s	0.65 s
2	ROI analysis	68	400 s	0.75 s
3	Vertex-wise heritability mapping	299,881	39.01 d	90 s
4	ROI analysis with permutation inference (one million permutations)	68,000,000	24.24 y*	3.5 h
5	Vertex-wise analysis with cluster inference (one thousand permutations)	299,881,000	106.88 y*	6.8 h

*Estimated computational time, using the average computational time, ~11.24 s, for each phenotype in case 3.

distributed. Therefore, the assumptions underlying classical GCTA analyses (e.g., common environmental effects are ignorable across individuals) remain important to our permutation procedure.

Potential Applications and Extensions. Although we demonstrate the capability of MEGHA using brain imaging data, it can be potentially applied to many types of high-dimensional phenotypes. In recent years, the study of complex diseases is shifting from the investigation of an isolated outcome variable to a complete and systematic characterization of the "phenome" (38, 39)—the full set of phenotypes of an individual—to unveil disease etiology and accommodate heterogeneity across individuals. The availability of high-dimensional phenotypic resources contained in electronic health records and population-based biobanks has spurred interest in phenome-wide association studies (PheWAS) (40, 41). The ability to identify or prioritize heritable traits in such large-scale repositories could facilitate important biomedical discoveries.

In addition to the capability of handling extremely high-dimensional phenotypic data, the flexibility of the kernel machines framework, which MEGHA is built on, offers multiple choices on kernel functions and SNP grouping strategies. This opens opportunities to examine different sources of genetic contributions (e.g., additive vs. epistatic effects) and dissect the genetic architecture of complex traits (42). Specifically, as shown in *Materials and Methods*, using a linear kernel function to combine all SNPs spanning the genome essentially produces an equivalent model to GCTA and assesses total additive genetic effects from common variants on phenotypic variables. Other kernels, such as a polynomial kernel or the identity-by-state (IBS) kernel (13, 43, 44), may capture complex genetic interactions (epistasis) and facilitate the analysis of broader-sense heritability. Alternatively, grouping SNPs

based on genes, pathways, findings of previous genome-wide association studies (GWASs) or other biologically informative information, and using different weighting strategies when combining SNPs, enable the investigation of genetic contributions from specific collections of SNPs.

Materials and Methods

MEGHA. MEGHA makes use of the semiparametric kernel machines model

$$y_i = \mathbf{x}_i^\top \beta + h(\mathbf{G}_i) + \epsilon_i, \quad i = 1, 2, \ldots N, \quad [1]$$

where N is the total number of subjects, y_i is a quantitative phenotype for subject i, $\mathbf{x}_i$ is a $p \times 1$ vector of nuisance variables for subject i (e.g., age, sex, and top principal components to adjust for population stratification), β is a $p \times 1$ vector of fixed effects, $\mathbf{G}_i = [G_{i,1}, \cdots, G_{i,L}]^\top$ denotes the genotypes of L SNP markers for subject i, $h(\cdot)$ is a nonparametric function located in a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, defined by an empirical, nonnegative-definite genetic relationship matrix (GRM) $\mathbf{K}$ that can be estimated from SNP data, and ϵ_i is a Gaussian distributed random error with zero-mean and homogeneous variance σ_ϵ^2. It has been shown that the semiparametric kernel machines model (1) can be converted into a linear mixed effects model (11)

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{g} + \epsilon, \quad \text{var}[\mathbf{y}] = \mathbf{V} = \sigma_g^2 \mathbf{K} + \sigma_\epsilon^2 \mathbf{I}, \quad [2]$$

where $\mathbf{y} = [y_1, \cdots, y_N]^\top$, $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]^\top$, $\mathbf{g}$ is an $N \times 1$ vector of the aggregate genetic effects of the individuals with $\mathbf{g} \sim \text{N}(0, \sigma_g^2 \mathbf{K})$, $\epsilon = [\epsilon_1, \cdots, \epsilon_N]^\top$, σ_g^2 is the variance explained by all of the L SNPs combined, and $\mathbf{I}$ is an identity matrix. Using a linear kernel function to combine all of the SNPs spanning the genome assesses the total additive genetic effects from common variants on phenotypes and essentially creates a linear mixed effects model equivalent to the one used in GCTA, which is useful for narrow-sense heritability analyses. The flexibility of the modeling framework allows for the use of other kernel functions (e.g., the quadratic kernel and the IBS kernel) and various SNP grouping strategies (e.g., based on genes, pathways, and previous GWAS findings), making it
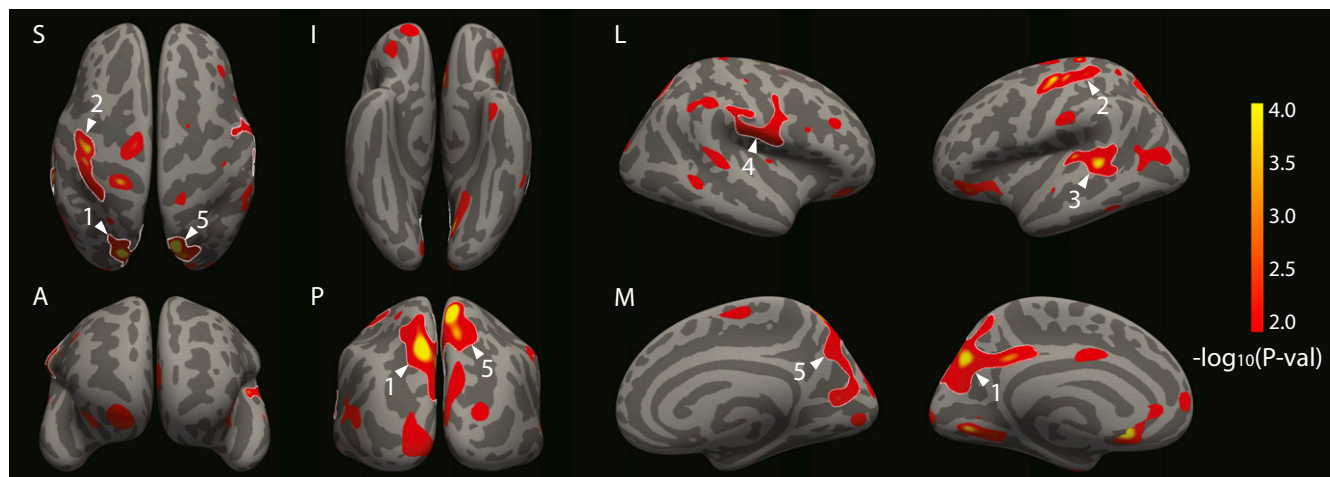

Fig. 3. Superior (S), inferior (I), lateral (L), anterior (A), posterior (P), and medial (M) views of the vertex-wise surface maps for SNP-based heritability significance of cortical thickness measures constructed by MEGHA. All clusters identified with a cluster-forming threshold $P = 0.01$ are shown. Five clusters that are familywise error corrected (FWEc) significant in size (FWEc, $P < 0.05$) based on 1,000 permutations are white outlined and annotated.

Ge et al.

possible to model different sources of genetic contributions (e.g., additive vs. epistatic effects) from a specific collection of SNPs to phenotypes.

GCTA uses the iterative restricted maximum likelihood (ReML) algorithm to estimate the variance components σ_g^2 and σ_ϵ^2 in the model (2) and gives an estimate of heritability by $\hat{h}^2 = \hat{\sigma}_g^2/\hat{\sigma}_p^2$, where $\hat{\sigma}_p^2$ is the estimated phenotypic variance. In contrast, MEGHA relies on a noniterative score test. It can be seen, from the linear mixed effects model (2), that testing for significant genetic effects is equivalent to testing the variance component $\mathcal{H}_0 : \sigma_g^2 = 0$. A score test has been proposed in the kernel machines literature (11, 12)

$$\mathcal{S}(\sigma_0^2) = \frac{1}{2\sigma_0^2} y^\top P_0 K P_0 y = \frac{1}{2\sigma_0^2} \hat{\epsilon}_0^\top K \hat{\epsilon}_0, \qquad [3]$$

where $\hat{\epsilon}_0$ is the maximum likelihood estimate (MLE) of the residuals under the null model $y = X\beta_0 + \epsilon_0$, σ_0^2 is the variance of ϵ_0, and $P_0 = I - X(X^\top X)^{-1} X^\top$ is the projection matrix. $\mathcal{S}(\sigma_0^2)$ is a quadratic function of y and follows a mixture of χ^2s under the null. We use the Satterthwaite method to approximate the distribution of $\mathcal{S}(\sigma_0^2)$ by a scaled χ^2 distribution $\kappa\chi_\nu^2$, where κ is the scale parameter and ν is the degrees of freedom that captures the effective number of independent SNPs combined by the kernel function. The two parameters are calculated by matching the first two moments, mean and variance, of $\mathcal{S}(\sigma_0^2)$ with those of $\kappa\chi_\nu^2$

$$\begin{cases} \delta = \mathbf{E}\left[\mathcal{S}(\sigma_0^2)\right] = \frac{1}{2}\mathbf{tr}\{P_0 K\} = \mathbf{E}\left[\kappa\chi_\nu^2\right] = \kappa\nu, \\ \rho = \mathbf{var}\left[\mathcal{S}(\sigma_0^2)\right] = \frac{1}{2}\mathbf{tr}\{P_0 K P_0 K\} = \mathbf{var}\left[\kappa\chi_\nu^2\right] = 2\kappa^2\nu. \end{cases} \qquad [4]$$

Solving the two equations gives $\kappa = \rho/2\delta$ and $\nu = 2\delta^2/\rho$. In practice, the unknown model parameter σ_0^2 is replaced by its ReML estimate, $\hat{\sigma}_0^2$, under the null model. To account for this substitution, we replace ρ by $\hat{\rho} \approx \hat{\mathcal{I}}_{gg} = \mathcal{I}_{gg} - \mathcal{I}_{g\epsilon}\mathcal{I}_{\epsilon\epsilon}^{-1}\mathcal{I}_{g\epsilon}^\top$, where $\mathcal{I}_{gg} = \mathbf{tr}\{P_0 K P_0 K\}/2$, $\mathcal{I}_{g\epsilon} = \mathbf{tr}\{P_0 K\}/2$, and $\mathcal{I}_{\epsilon\epsilon} = \mathbf{tr}\{P_0 P_0\}/2$ (11). With the adjusted parameters $\hat{\kappa} = \hat{\rho}/2\delta$ and $\hat{\nu} = 2\delta^2/\hat{\rho}$, the P value of an observed score statistic $\mathcal{S}(\hat{\sigma}_0^2)$ is then computed using the scaled χ^2 distribution $\hat{\kappa}\chi_{\hat{\nu}}^2$. In high-dimensional heritability analyses, we note that only a simple linear regression model under the null, $y^{(v)} = X\beta_0^{(v)} + \epsilon_0^{(v)}$, needs to be fit for each phenotype v. If all phenotypes share the same covariate matrix X, the projection matrix P_0 can be precomputed, and thus the computation of the test statistics for all phenotypes can be very efficient.

To obtain a point estimate of the SNP-based heritability, we consider the Wald test statistic for the null hypothesis $\mathcal{H}_0 : \sigma_g^2 = 0$, which is distributed as

$$\frac{\hat{\sigma}_g^4}{\mathbf{var}\left[\hat{\sigma}_g^2\right]} \sim \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2, \qquad [5]$$

a half-half mixture of the χ^2 distribution χ_0^2 with all probability mass at zero and the χ^2 distribution with 1 degrees of freedom χ_1^2, as the null hypothesis lies on the boundary of the parameter space (45). Because Wald tests are asymptotically equivalent to score tests, we assume that the Wald test P value is identical to the score test P value, which can be converted into a Wald test statistic T using the mixture of χ^2s in Eq. 5. Then, assuming that the phenotypic variance σ_p^2 is estimated with very high precision and noticing the approximation $\mathbf{std}[\hat{h}^2] \approx 316/N$ derived recently (46), we have

$$\hat{h}^2 = \hat{\sigma}_g^2/\hat{\sigma}_p^2 \approx \min\left\{\frac{316}{N}\sqrt{T}, 1\right\}. \qquad [6]$$

Permutation Procedure. The efficient computation of the test statistic allows for the use of standard permutation procedures. However, with the presence of covariates X, shuffling the rows and columns of the GRM K in Eq. 3 produces inaccurate inferences. Inspired by the ideas of the Huh–Jhun permutation (47), we propose a permutation procedure that involves a transformation that projects the data from N dimensional space onto an $N-p$ dimensional subspace and removes the effect of nuisance variables. Specifically, because P_0 is a symmetric and idempotent (i.e., $P_0^2 = P_0$) matrix of rank $N-p$, it can be decomposed as $P_0 = UDU^\top$, where U is an $N \times N$ matrix satisfying $U^\top U = I_{N \times N}$, D is a diagonal matrix with $N-p$ ones and p zeros on the diagonal. Without loss of generality, we assume that the first $N-p$ diagonal elements are one. Therefore, if we discard the last p columns of U and denote the resulting $N \times (N-p)$ matrix as $\tilde{U}$, we have $P_0 = \tilde{U}\tilde{U}^\top$ and $\tilde{U}^\top \tilde{U} = I_{(N-p) \times (N-p)}$. Now applying $\tilde{U}^\top$ to both sides of the model (2), and noticing the fact that $X^\top \tilde{U}\tilde{U}^\top X = X^\top P_0 X = 0$, the transformed model is on an $N-p$ dimensional space

$$\bar{y} := \tilde{U}^\top y = \tilde{U}^\top g + \tilde{U}^\top \epsilon := \tilde{g} + \tilde{\epsilon}. \qquad [7]$$

Because $\tilde{g} \sim N(0, \sigma_g^2 \tilde{U}^\top K \tilde{U})$, the new GRM is $\tilde{K} = \tilde{U}^\top K \tilde{U}$. We note that the transformed score test statistic is invariant

$$\tilde{\mathcal{S}}(\hat{\sigma}_0^2) = \frac{1}{2\hat{\sigma}_0^2}\bar{y}^\top \tilde{K}\bar{y} = \frac{1}{2\hat{\sigma}_0^2} y^\top \tilde{U}\tilde{U}^\top K \tilde{U}\tilde{U}^\top y = \frac{1}{2\hat{\sigma}_0^2} y^\top P_0 K P_0 y = \mathcal{S}(\hat{\sigma}_0^2). \qquad [8]$$

Shuffling the rows and columns of the transformed GRM $\tilde{K}$ is now equivalent to shuffling the transformed phenotype $\bar{y}$, and importantly, we have $\mathbf{cov}[\bar{y}] = \sigma_0^2 \tilde{U}^\top \tilde{U} = \sigma_0^2 I$, and thus $\bar{y} \sim N(0, \sigma_0^2 I)$, which indicates that $\bar{y}^*$, a permuted sample of $\bar{y}$, has the same distribution as $\bar{y}$. Therefore, the proposed permutation procedure ensures that all permutation samples of the score test statistic follow the same null distribution.

Permutation allows us to consider arbitrary statistics of interest and offers great flexibility in making inferences. For example, to obtain an FWEc P value of heritability for each region in an ROI-based analysis, using a predetermined anatomical or functional atlas with a total of R regions, we compute the permuted score test statistic $\mathcal{S}_k^*(r)$ for each ROI r, and record the maximal statistic over the R ROIs, $M_k = \max_{r=1,2,\cdots,R} \mathcal{S}_k^*(r)$, for each permutation $k = 1, \cdots, N_{\text{perm}}$. Then for an observed score test statistic $\mathcal{S}(r)$ for the rth ROI, the FWEc P value can be computed as (48)

$$p_{\text{FWE}}(r) = \frac{\#\{M_k \geq \mathcal{S}(r)\}}{N_{\text{perm}}}, \quad r = 1, \cdots, R. \qquad [9]$$

This permutation procedure implicitly accounts for the correlation structure among the ROIs and therefore produces more accurate FWEc P values than Bonferroni-type corrections, which treats each measure as independent. As a second example, we consider cluster inferences on voxel-/vertex-wise significance maps of heritability, as commonly used in the neuroimaging literature (18, 34, 35). Clusters are defined by contiguous voxels/vertices with test statistics above a predefined threshold (or equivalently, P values below a threshold). For each permutation $k = 1, \cdots, N_{\text{perm}}$, we compute the permuted score test statistic $\mathcal{S}_k^*(v)$ for each voxel/vertex v, threshold the statistical map, and record the maximal cluster size M_k. Then for an observed cluster $\mathcal{C}$ with size c, the FWEc P value is (48)

$$p_{\text{FWE}}(\mathcal{C}) = \frac{\#\{M_k \geq c\}}{N_{\text{perm}}}. \qquad [10]$$

Inferences on other statistics, e.g., a weighted voxel-/vertex-wise average of test statistics to summarize heritability into a single number and provide an overall significance, can be easily made following similar procedures.

The Brain GSP. The Harvard/MGH Brain GSP is a neuroimaging and genetics study of brain and behavioral phenotypes. More than 3,500 native English-speaking adults with normal or corrected-to-normal vision were recruited from Harvard University, MGH, and the surrounding Boston communities. To avoid spurious effects resulting from population stratification, we restricted our analyses to 1,320 young adults (18–35 y old) of non-Hispanic European ancestry with no history of psychiatric illnesses or major health problems (age, 21.54 ± 3.19 y old; female, 53.18%; right-handedness, 91.74%). All participants provided written informed consent in accordance with guidelines set by the Partners Health Care Institutional Review Board or the Harvard University Committee on the Use of Human Subjects in Research. For further details about the recruitment process and participants, and imaging data acquisition, we refer the reader to ref. 16.

Imaging Data Processing. We used FreeSurfer (freesurfer.net) (49), version 4.5.0, a freely available, widely used, and extensively validated brain MRI analysis software package, to process the structural brain MRI scans and compute global and regional morphological measurements. For vertex-wise heritability analyses of cortical thickness, sulcal depth, curvature, and cortical surface area, we resampled subject-specific measures onto FreeSurfer's *fsaverage* representation, which consists of more than 300,000 vertices across the two hemispheres with an intervertex distance of ~1 mm, and smoothed the coregistered surface maps using a Gaussian kernel with full width at half maximum (FWHM) 20 mm. We also defined a neighborhood structure on the surface mesh for surface-based clustering.

Genetic Analysis. We used PLINK (pngu.mgh.harvard.edu/purcell/plink/) (50), version 1.07, to preprocess the GSP genome-wide SNP data. Major procedures included sex discrepancy check, removing population outliers, spuriously related subjects, and subjects with low genotype call rate (<97%). Individual markers that contained an ambiguous strand assignment or that did not

satisfy the following quality criteria were excluded from the analyses: genotype call rate ≥95%, minor allele frequency (MAF) ≥1%, and Hardy-Weinberg equilibrium of $P \geq 1 \times 10^{-6}$; 580,479 SNPs remained for analysis after quality control. We performed a complete linkage clustering of individuals and a multidimensional scaling (MDS) analysis (Fig. S3), based on autosomal genome-wide SNP data in PLINK, to ensure that no clear population stratification and outliers exist in the sample. We used the GCTA toolbox (6), version 1.24.4 (www.complextraitgenomics.com/software/gcta/download.html), to estimate the GRM used in the heritability analyses from all genotyped autosomal SNPs.

Heritability Analyses of Brain Imaging Measurements. For all MEGHA and GCTA analyses of global, regional, and vertex-wise brain imaging measurements, we included age, sex, handedness, scanner group, console group, and coil type as covariates. To account for population substratification, the top five principal components (PCs) of the GRM were also included in the model as nuisance variables. We adjusted for ICV in all of the analyses of cortical/subcortical gray/white matter volume measures, and sulcal depth, curvature, and cortical surface area measures, but not in the cortical thickness analyses because cortical thickness is not correlated with ICV.

1. Falconer D (1960) *Introduction to Quantitative Genetics* (Oliver and Boyd, Edinburgh, Scotland).
2. Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62(5):1198–1211.
3. Neale M, Cardon L (1992) *Methodology for Genetic Studies of Twins and Families* (Springer, New York).
4. Visscher PM, et al. (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet* 2(3):e41.
5. Yang J, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42(7):565–569.
6. 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.
7. Yang J, et al. (2011) Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 43(6):519–525.
8. Lee SH, et al.; Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ); International Schizophrenia Consortium (ISC); Molecular Genetics of Schizophrenia Collaboration (MGS) (2012) Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 44(3):247–250.
9. Manolio TA, et al. (2009) Finding the missing heritability of complex diseases. *Nature* 461(7265):747–753.
10. Lee SH, Wray NR, Goddard ME, Visscher PM (2011) Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 88(3):294–305.
11. Liu D, Lin X, Ghosh D (2007) Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics* 63(4):1079–1088.
12. Lin X (1997) Variance component testing in generalised linear models with random effects. *Biometrika* 84(2):309–326.
13. Kwee LC, Liu D, Lin X, Ghosh D, Epstein MP (2008) A powerful and flexible multilocus association test for quantitative traits. *Am J Hum Genet* 82(2):386–397.
14. Wu MC, et al. (2010) Powerful SNP-set analysis for case-control genome-wide association studies. *Am J Hum Genet* 86(6):929–942.
15. Wu MC, et al. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 89(1):82–93.
16. Holmes AJ, et al. (2012) Individual differences in amygdala-medial prefrontal anatomy link negative affect, impaired social functioning, and polygenic depression risk. *J Neurosci* 32(50):18087–18100.
17. Stein JL, et al.; Alzheimer's Disease Neuroimaging Initiative (2010) Voxelwise genome-wide association study (vGWAS). *Neuroimage* 53(3):1160–1174.
18. Ge T, Feng J, Hibar DP, Thompson PM, Nichols TE (2012) Increasing power for voxel-wise genome-wide association studies: The random field theory, least square kernel machines and fast permutation procedures. *Neuroimage* 63(2):858–873.
19. Shen L, et al.; Alzheimer's Disease Neuroimaging Initiative (2010) Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage* 53(3):1051–1063.
20. Glahn DC, et al. (2010) Genetic control over the resting brain. *Proc Natl Acad Sci USA* 107(3):1223–1228.
21. Thompson PM, Ge T, Glahn DC, Jahanshad N, Nichols TE (2013) Genetics of the connectome. *Neuroimage* 80:475–488.
22. Gottesman I, Shields J (1972) *Schizophrenia Genetics: A Twin Study Vantage Point* (Academic Press, New York).
23. Gottesman II, Gould TD (2003) The endophenotype concept in psychiatry: Etymology and strategic intentions. *Am J Psychiatry* 160(4):636–645.
24. Blangero J (2004) Localization and identification of human quantitative trait loci: King harvest has surely come. *Curr Opin Genet Dev* 14(3):233–240.
25. Meyer-Lindenberg A, Weinberger DR (2006) Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nat Rev Neurosci* 7(10):818–827.
26. Winkler AM, et al. (2010) Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage* 53(3):1135–1146.
27. Glahn DC, et al. (2012) High dimensional endophenotype ranking in the search for major depression risk genes. *Biol Psychiatry* 71(1):6–14.
28. Eyler LT, et al. (2012) A comparison of heritability maps of cortical surface area and thickness and the influence of adjustment for whole brain measures: A magnetic resonance imaging twin study. *Twin Res Hum Genet* 15(3):304–314.
29. Lenroot RK, et al. (2009) Differences in genetic and environmental influences on the human cerebral cortex associated with development during childhood and adolescence. *Hum Brain Mapp* 30(1):163–174.
30. Joshi AA, et al. (2011) The contribution of genes to cortical thickness and volume. *Neuroreport* 22(3):101–105.
31. Yoon U, Fahim C, Perusse D, Evans AC (2010) Lateralized genetic and environmental influences on human brain morphology of 8-year-old twins. *Neuroimage* 53(3):1117–1125.
32. Thompson PM, et al. (2001) Genetic influences on brain structure. *Nat Neurosci* 4(12):1253–1258.
33. 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.
34. Friston KJ, Worsley KJ, Frackowiak RS, Mazziotta JC, Evans AC (1994) Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp* 1(3):210–220.
35. Friston KJ, Holmes A, Poline JB, Price CJ, Frith CD (1996) Detecting activations in PET and fMRI: Levels of inference and power. *Neuroimage* 4(3 Pt 1):223–235.
36. Vul E, Harris C, Winkielman P, Pashler H (2009) Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect Psychol Sci* 4(3):274–290.
37. Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: The dangers of double dipping. *Nat Neurosci* 12(5):535–540.
38. Houle D, Govindaraju DR, Omholt S (2010) Phenomics: The next challenge. *Nat Rev Genet* 11(12):855–866.
39. Imboden M, Probst-Hensch NM (2013) Biobanking across the phenome: At the center of chronic disease research. *BMC Public Health* 13:1094.
40. Kohane IS (2011) Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet* 12(6):417–428.
41. Jensen PB, Jensen LJ, Brunak S (2012) Mining electronic health records: Towards better research applications and clinical care. *Nat Rev Genet* 13(6):395–405.
42. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012) The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci USA* 109(4):1193–1198.
43. Schaid DJ (2010) Genomic similarity and kernel methods I: Advancements by building on mathematical and statistical foundations. *Hum Hered* 70(2):109–131.
44. Schaid DJ (2010) Genomic similarity and kernel methods II: Methods for genomic information. *Hum Hered* 70(2):132–140.
45. Molenberghs G, Verbeke G (2007) Likelihood ratio, score, and Wald tests in a constrained parameter space. *Am Stat* 61(1):22–27.
46. Visscher PM, et al. (2014) Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet* 10(4):e1004269.
47. Huh M, Jhun M (2001) Random permutation testing in multiple linear regression. *Commun Statist - Theory Meth* 30(10):2023–2032.
48. Westfall P, Young SS (1993) *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment* (John Wiley & Sons, New York).
49. Fischl B (2012) FreeSurfer. *Neuroimage* 62(2):774–781.
50. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.