Massively expedited genome-wide heritability analysis (MEGHA)

Tian Ge*, Thomas E. Nichols, Phil H. Lee, Avram J. Holmes, Joshua L. Roffman, Randy L. Buckner, Mert R. Sabuncu, and Jordan W. Smoller

*Department of Radiology, Harvard Medical School, Boston, MA 02115; †Department of Neurology, Massachusetts General Hospital, Boston, MA 02114; ‡Department of Psychiatry, Massachusetts General Hospital, Boston, MA 02114; §Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115; ¶Division of Biomedical Informatics, University of Washington, Seattle, WA 98195; ¶¶Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom; ||Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115

Edited by C. Thomas Caskey, Baylor College of Medicine, Houston, TX, and approved January 15, 2015 (received for review August 12, 2014)

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.

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.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

1To whom correspondence may be addressed. Email: tge1@mgh.harvard.edu or jsmoller@hms.harvard.edu.

2M.R.S. and J.W.S. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1415603112/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1415603112
PNAS | February 24, 2015 | vol. 112 | no. 8 | 2479–2484

GENETICS

STATISTICS

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

As a demonstration of application, MEGA H 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 measures, 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).

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 mean cortical thickness, overall mean cortical thickness, left/right hemisphere 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 highlyheritable, 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. MEGA H estimates of heritability magnitude are tightly correlated with GCTA results (Fig. 1).

We next applied both MEGA H 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 MEGA H 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 MEGA H and GCTA, although MEGA H 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 MEGA H 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 MEGA H 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>
<thead>
<tr>
<th>Table 1. Analysis of global morphometric measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Intracranial volume (ICV)</td>
</tr>
<tr>
<td>Total brain volume</td>
</tr>
<tr>
<td>Left hemispheric cortical GM volume</td>
</tr>
<tr>
<td>Right hemispheric cortical GM volume</td>
</tr>
<tr>
<td>Total cortical GM volume</td>
</tr>
<tr>
<td>Total subcortical GM volume</td>
</tr>
<tr>
<td>Total GM volume</td>
</tr>
<tr>
<td>Left hemispheric WM volume</td>
</tr>
<tr>
<td>Right hemispheric WM volume</td>
</tr>
<tr>
<td>Total WM volume</td>
</tr>
<tr>
<td>Left hemispheric mean cortical thickness</td>
</tr>
<tr>
<td>Right hemispheric mean cortical thickness</td>
</tr>
<tr>
<td>Overall mean cortical thickness</td>
</tr>
<tr>
<td>Left hemispheric total surface area</td>
</tr>
<tr>
<td>Right hemispheric total surface area</td>
</tr>
<tr>
<td>Total surface area</td>
</tr>
</tbody>
</table>

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 $h^2$ with the corresponding SE and P value obtained by GCTA, the familywise error corrected (FWEc) P value using Bonferroni correction, the heritability estimate $h^2$ and P value computed by MEGA H, 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.

2480 | www.pnas.org/cgi/doi/10.1073/pnas.1415603112
Ge et al.
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 heritability analyses of cortical thickness measures, 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, 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 distribution.

**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 non-parametric 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, 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 distribution.
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 (PhWAS) (40, 41). The high-dimensional phenotypic resources contained in electronic health records and population-based biobanks have spurred interest in PhWAS.

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 for 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 (GWAS) 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

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

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 \( i \) (i.e., age, sex, and top principal components to adjust for population stratification), \( \beta \) is a \( p \times 1 \) vector of fixed effects, \( G_i \) is the set of genotypes of \( L \) SNPs 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) \( K \) that can be estimated from SNP data, and \( \epsilon_i \) is a Gaussian distributed random error with zero-mean and homogeneous variance \( \sigma^2_\epsilon \). It has been shown that the semiparametric kernel machines model (1) can be converted into a linear mixed effects model (11)

\[ y = \mathbf{X}\beta + \mathbf{g} + \epsilon, \quad \text{var}(y) = \mathbf{V} = \sigma^2_\epsilon \mathbf{I}, \]

where \( y = [y_1, \ldots, y_N]^\top, \mathbf{x} = [x_1, \ldots, x_N]^\top, \mathbf{g} \) is an \( N \times 1 \) vector of the aggregate genetic effects of the individuals with \( g \sim \mathcal{N}(0, \sigma^2_g \mathbf{I}), \epsilon = [\epsilon_1, \ldots, \epsilon_N]^\top, \sigma^2_\epsilon \) is the variance explained by all of the \( L \) SNPs combined, and \( I \) 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

---

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

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

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

---

**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.
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 $\sigma^2_a$ and $\sigma^2_e$ in the model (2) and gives an estimate of heritability by $h^2 = \frac{\sigma^2_a}{\sigma^2_a + \sigma^2_e}$, where $\sigma^2_a$ 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 $\sigma^2_a$. A score test has been proposed in the kernel machines literature (11, 12)

$$S(\sigma^2_a) = \frac{1}{2\sigma_0^2} \left( y^T P_k K P_k y - 1 \right) K \sigma_0^2$$  \[3\]

where $\sigma_0$ is the maximum likelihood estimate (MLE) of the residuals under the null model $y = X \beta_0 + e$, $\sigma^2_a$ is the variance of $\beta_0$, and $P_k = I - X(X'X)^{-1}X'$ is the projection matrix. $S(\sigma^2_a)$ is a quadratic function of $y$ and follows a mixture of $\chi^2$ distributions under the null. We use the Satterthwaite method to approximate the distribution of $S(\sigma^2_a)$ by a scaled $\chi^2$ distribution $\chi^2_\nu$, where $\nu$ is the scale parameter and $\nu$ is the degrees of freedom that captures the effective number of independent SNPs combined by the kernel function. The two parameters are calculated by determining the first two moments, mean and variance, of $S(\sigma^2_a)$ with those of $\chi^2_\nu$.

$$\begin{align*}
\delta &= E[S(\sigma^2_a)] = \frac{1}{2} \text{tr}(P_k K) = \text{E}[|X|^2] = \sigma_0^2, \\
\rho &= \text{var}(S(\sigma^2_a)) = \frac{1}{2} \text{tr}(P_k K P_k K P_k) = \text{var}[\chi^2_\nu] = 2\nu \sigma_0^2.
\end{align*}$$  \[4\]

Solving the two equations gives $\nu = \frac{\rho}{\delta^2}$ and $\nu = \frac{\rho}{\delta^2}$. In practice, the unknown model parameter $\sigma^2_a$ is replaced by its REML estimate, $\sigma^2_a$, under the null model. To account for this substitution, we replace $\rho$ by $\tilde{\rho} = \text{tr}(Y Y' - \text{null}) = \text{tr}(P_k K P_k K P_k)$, where $Y = \text{tr}(P_k K P_k K P_k)$, $I_2 = \text{tr}(P_k K P_k)$, and $I_1 = \text{tr}(P_k P_k)$. With the adjusted parameters $\delta = \frac{\rho}{\delta^2}$ and $\nu = \frac{\rho}{\delta^2}$, the $P$ value of an observed score statistic $S(\sigma^2_a)$ is then computed using the scaled $\chi^2$ distribution $\chi^2_\nu$. In high-dimensional heritability analyses, we note that only a simple linear regression model under the null, $y^* = X \beta_0 + e$, needs to be fit for each phenotype $v$. If all phenotypes share the same covariate matrix $X$, the projection matrix $P_k$ 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 $H_0: \sigma^2_a = 0$, which is distributed as

$$\frac{\hat{\sigma}^2_a}{\text{var}[\hat{\sigma}^2_a]} \sim \chi^2_1$$  \[5\]

where $\hat{\sigma}^2_a$ is the maximum likelihood estimate (MLE) of the residuals under the null model (2). Therefore, if we discard the last $p$ columns of $U$ and denote the resulting $N \times (N-p)$ matrix as $U$, we have $P_k = U^T U$ and $U^T U = I_{N-p}$. Now applying $U$ to both sides of the model (2), and noticing the fact that $X' U X' = X' P_k X = 0$, the transformed model is on an $N-p$ dimensional space

$$y = U' y = U' g + U' e = g' + \epsilon.$$  \[7\]

Because $g \sim N(0, \sigma^2_a U K U')$, the new GRM is $K = U K U'$. We note that the transformed score test statistic is invariant

$$S(\sigma^2_a) = \frac{1}{2\sigma_0^2} \left( y' K y - 1 \right) K \sigma_0^2.$$  \[8\]

Shuffling the rows and columns of the transformed GRM $K$ is now equivalent to shuffling the transformed phenotype $y$, and importantly, we have $\text{cov}(y') = \sigma^2_a U = \sigma^2_a I$, and thus $y \sim N(0, \sigma^2_a)$, which indicates that $y'$, a permuted sample of $y$, has the same distribution as $y$. 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 greater flexibility in making inferences. For example, to obtain an FWE 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 $S^*(r)$ for each ROI $r$, and record the maximal statistic over the $R$ ROIs, $M_{r} = \max_{1,...,R} S^*(r)$, for each permutation $k = 1,...,N_{\text{perm}}$. Then for an observed score test statistic $S(r)$ for the $r$th ROI, the FWE $P$ value can be computed as

$$P_{\text{FWE}} = \frac{\#(M_{r} \geq c)}{N_{\text{perm}}}$$  \[9\]

This permutation procedure implicitly accounts for the correlation structure among the ROIs and therefore produces more accurate FWE $P$ 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 $k = 1,...,N_{\text{perm}}$, we compute the permuted score test statistic $S^*(v)$ for each voxel/vertex $v$, threshold the statistical map, and record the maximal cluster size $M_{c}$. Then for an observed cluster $C$ with size $c$, the FWE $P$ value is

$$P_{\text{FWE}} = \frac{1}{N_{\text{perm}}} \#(M_{c} \geq c)$$  \[10\]

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-handed, 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: geno-
type call rate ≥95%, minor allele frequency (MAF) ≥1%, and Hardy-Weinberg
equilibrium. Quality control was carried out as described previously [8]. SNPs rema-
in 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 (V6), 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 stratification, the top five principal
components (PCs) of the GRM were also included in the model as nuisance
covariates. To account for population substratification, the top five principal
analyses of global, regional, and vertex-wise
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 stratification, the top five principal
components (PCs) of the GRM were also included in the model as nuisance
covariates. To account for population substratification, the top five principal
analyses of global, regional, and vertex-wise

Aviability. A MATLAB implementation of MEGHA is available for download at
www.massgeneral.org/psychiatry/research/pngu_software.aspx.

Acknowledgments. This research was carried out in whole or in part at the
Athinoula A. Martinos Center for Biomedical Imaging at Massachusetts
General Hospital (MGH), using resources provided by the Center for
Functional Neuroimaging Technologies, P41EB015896, a P41 Biotechnology
Resource Grant supported by the National Institute of Biomedical Imaging
and Bioengineering (NIBIB), National Institutes of Health (NIH). Data were
provided by the Brain Genomics SuperProject (GSP) of Harvard
University and MGH, with support from the Center for Brain Science
Neuroinformatics Research Group, Athinoula A. Martinos Center for Bio-
medical Imaging, Center for Human Genetic Research, and Stanley Center
for Psychiatric Research. Twenty individual investigators at Harvard and
MGH generously contributed data to the overall project. This research was
also supported in part by NIH Grants R01 EB015611-01 and US4 MH091657-03
(to T.E.N.); K99MH101367 (to P.H.L.); K01MH99232 (to A.J.H.); R01
NS085334, R01 NS070963, and NIH NIBIB 1K25EB013649-01 (to M.R.S.);
K24MH094614 and R01 MH101486 (to J.W.S.); Welcome Trust Grants
100309/21/Z/Z and 098369/21/Z (to T.E.N.); and Brightfund Grant AHAF-
A201233 (to M.R.S.). J.W.S. is a Tepper Family MGH Research Scholar.

1. Falconer D (1960) Introduction to Quantitative Genetics (Oliver and Boyd, Edinburgh,
Scotland).
(Springer, New York).
4. Vischer PM, et al. (2006) Assumption-free estimation of heritability from genome-
8. Lee SH, et al; Schizophrenia Psychiatric Genome-Wide Association Study
Consortium (PSC:SCZ); International Schizophrenia Consortium (ISC); Molecular Genetics of
(Springer, New York).
4. Vischer PM, et al. (2006) Assumption-free estimation of heritability from genome-
8. Lee SH, et al; Schizophrenia Psychiatric Genome-Wide Association Study
Consortium (PSC:SCZ); International Schizophrenia Consortium (ISC); Molecular Genetics of
(Springer, New York).
4. Vischer PM, et al. (2006) Assumption-free estimation of heritability from genome-