Supplementary Material for

Hierarchical Genetic Organization of Human Cortical Surface Area

Chi-Hua Chen, E. D. Gutierrez, Wes Thompson, Matthew S. Panizzon, Terry L. Jernigan, Lisa T. Eyler, Christine Fennema-Notestine, Amy J. Jak, Michael C. Neale, Carol E. Franz, Michael J. Lyons, Michael D. Grant, Bruce Fischl, Larry J. Seidman, Ming T. Tsuang, William S. Kremen* Anders M. Dale

*To whom correspondence should be addressed. E-mail: wkremen@ucsd.edu

Published 30 March 2012, *Science* **335**, 1634 (2012)
DOI: 10.1126/science.1215330

This PDF file includes:

Materials and Methods

Participants

Participants were part of the Vietnam Era Twin Study of Aging (VETSA). An overview of the VETSA project can be found in (*1*). The VETSA is a longitudinal behavior genetic study of cognitive and brain aging. A total of 1237 male-male twins participated in VETSA wave 1, and a subset underwent MRI. Twins in the VETSA project were randomly selected within a given age range (with one exception) from over 8,000 twins who had participated in a previous Vietnam Era Twin Registry study (*2*). The exception to the random selection was that both members of a twin pair had to agree to participate. To participate in VETSA wave 1, participants had to be between the ages of 51 and 59 at the time of recruitment. The goal was to have an age-homogenous midlife sample to provide a baseline for this longitudinal aging study. Having a narrow age range meant that there would be enhanced ability to examine individual differences in trajectories over time compared to more typical cohort-sequential designs which have smaller subsets of people in a range of age cohorts.

The Vietnam Era Twin Registry comprises a national sample male-male twin pairs born between 1939 and 1957 who both served in the United States military sometime between 1965 and 1975. The Registry is not a patient sample, nor a sample of people selected on the basis of using the U.S. Department of Veterans Affairs hospital system. The majority of Registry members were not in Vietnam or exposed to combat (*3,4*). Registry members are currently middle-aged men living throughout the United States. Those in the VETSA project were demographically similar to the entire Registry. The VETSA sample is representative of U.S. men in their age range based on sociodemographic and health characteristics determined by U.S. census and Center for Disease Control data (*1,5*).

In most of the twins, zygosity was determined by 25 microsatellite markers. Zygosity for the small portion of the sample, for whom genotyping data were not available, was determined by questionnaire and blood group. This method had 95% accuracy relative to our DNA analysis. In cases where genotyping and questionnaire-based zygosity differed, we used the DNA-based classification.

The MRI study began in year 3 of the primary VETSA study. Additional details regarding the MRI sample and methods can be found elsewhere (*6,7*). Only 6% of those invited to participate in the MRI study declined to do so. Others were excluded for the following reasons: possible metal in the body (7%); claustrophobia (3%); inability to travel to the test site (5%); exclusion of co-twin (9%); other reasons (3%). Scanner problems resulted in a loss of data for 8% of the participants. The present study is based on analyzable scans from 474 twins. The genetic analyses involved 406 twins: 110 monozygotic (MZ) and 93 dizygotic (DZ) pairs. The mean age was 55.8 (±2.6) years (range: 51-59), and the mean level of education was 13.9 years (±2.1). The subset of participants who had MRIs were demographically similar to the rest of the VETSA sample (*6,7*). All participants gave informed consent to participate in the research and the study was approved by the Institutional Review Boards of the University of California, San Diego, Boston University, and the Massachusetts General Hospital.

Image Acquisition

Imaging acquisition and processing methods have been described in previous VETSA publications (e.g., *6,7*) as well as other publications cited below that describe FreeSurfer methods. Images were acquired on Siemens 1.5 Tesla scanners (241 participants at the University of California, San Diego; 233 participants at Massachusetts General Hospital). Sagittal T1-weighted MPRAGE sequences were employed with a TI=1000ms, TE=3.31ms, TR=2730ms, flip angle=7 degrees, slice thickness=1.33mm, voxel size 1.3x1.0x1.3mm. In almost all cases, twins within a pair were assessed at the same site.

Image Processing

The cortical surface was reconstructed to measure surface areas at each surface location (a total of more than 160,000 locations for each hemisphere) using a semi-automated approach provided by the FreeSurfer software (*8-10*). Variation in image intensity due to RF coil sensitivity inhomogeneities were corrected, a normalized intensity image was created, and the skull (non-brain) was removed from this image. A preliminary segmentation was then partitioned using a connected components algorithm, with connectivity not allowed across the established cutting planes. Interior holes in the components representing white matter were filled, resulting in a single filled volume for each cortical hemisphere. The resulting surface was covered with a polygonal tessellation and smoothed to reduce metric distortions. A refinement procedure was then applied to obtain a representation of the gray/white boundary, and the resulting surface was subsequently deformed outwards to obtain an explicit representation of the pial surface. Once generated, the cortical surface model was manually reviewed and edited for anatomical accuracy. Minimal manual editing was performed in accordance with standard, objective editing rules. Each subject's cortical surface was mapped to spherical atlas space, using a diffeomorphic registration procedure based on folding patterns (*9*). The surface alignment method used is not anchored to specific anatomical landmarks (e.g., fundus of the central sulcus). Rather, it uses the entire pattern of surface curvature at every vertex across the cortex to register individual subjects to atlas space (*11*). Although registration into atlas space is driven by cortical folding patterns, there is evidence that the folds are good predictors of the locations of functionally and histologically distinct regions (*11,12*). Then, for each subject, the standardized atlas surface tessellation was transformed into subject space based on the inverse of the subject to atlas mapping. Vertex-wise estimates of areal expansion or compression from atlas space to subject space, for each subject, were then obtained using standard FreeSurfer functions (*13*). These expansion maps were then smoothed using iterative nearest neighbor averaging.

The optimal size of smoothing was determined empirically by reanalyzing the data with various levels of smoothing to investigate the effect of smoothing on the heritability estimate and the clustering. We used three levels of smoothing from small to large corresponding to 176, 705 and 2819 iterations, respectively, of nearest-neighbor smoothing on the standardized atlas tessellation (**fig. S3**). The 2819-iteration was found to be required in order to obtain sufficiently high heritability values to enable accurate estimation of genetic correlations, which are critical for the stability of the subsequent cluster analysis. **Fig. S4** illustrates the smoothing kernel, thresholded at half-max, for the 2819-iteration smoothing, displayed on a 3D surface rendering of the mean pial cortical surface in atlas space.

Statistical analysis

In the classical twin study of sets of MZ and DZ twins, four latent factors can account for the variance of any phenotype: additive genetic effects (A); non-additive genetic effects, including dominance (D); common or shared environmental effects (C); and non-shared or individual-specific environmental effects (E) (*14*). Because MZ twins are presumed to be genetically identical, they correlate perfectly ($r = 1.0$) with respect to both additive and non-additive genetic effects. DZ twins share, on average, 50% of their genes, resulting in correlations of 0.50 for additive genetic effects and 0.25 for non-additive genetic effects. The C term is defined as environmental factors that make twins similar; hence, common environmental factors correlate 1.0 across twin pairs, regardless of zygosity. The E term represents environmental factors that lead to differences between twins. Because these are individual-specific factors, they are assumed to be uncorrelated across twins. Error is assumed to be random across individuals, so measurement error forms part of the estimate of E in these analyses. These latent factors comprise what are referred to as the univariate ACE or ADE models; due to model under-identification, an ACDE model cannot be tested in the classical twin design (*14*).

The ACE and ADE models are easily extended to the multivariate case (*14*). In addition to genetic and environmental sources of variance, sources of covariance can also be examined in the bivariate model. In the present study, we used bivariate models to compute genetic correlations of surface area measures between two points or vertices on the cortex. A phenotypic correlation measures shared variance; a genetic correlation measures shared genetic variance. More specifically, a phenotypic correlation is defined as the total covariance (genetic plus environmental) of two variables divided by the square root of the product of the total variance of variable 1 and the total variance of variable 2. After decomposing the sources of variance in the bivariate model, we computed genetic correlations. These are defined as the *genetic* covariance divided by the square root of the product of the *genetic* variance of variable 1 and the *genetic* variance of variable 2.

The analyses were performed using Mx, a maximum likelihood-based structural equation modeling software package for genetically informative data (*15*). The raw surface measure data were standardized to z-scores across all subjects. Prior to the model fitting, the surface area measure at each location was normalized to adjust for global effects (each measure was divided by the total cortical surface area of each individual). Age and site were covariates in all models. All models were fitted against the variance-covariance matrices of surface area measures from the MZ and DZ twins, with each location treated as an independent and continuous variable. In conjunction with Mx, we utilized the MATLAB (The MathWorks, Natick, Massachusetts) software environment to sequentially perform the analyses. Preliminary analyses using the ACE model revealed that most of the estimates of common environmental variance were at or near zero. With 406 participants, we were also underpowered to differentiate between A and D effects (*14*). Overall, AE models tended to provide the best fits to the data. Consequently, the genetic effects estimated in these AE models refer to broad-sense heritability, reflecting the proportion of phenotypic variance accounted for by the combined effect of all genetic influences (A+D). The basic bivariate AE model is shown in **fig. S5**.

Fuzzy cluster analysis

The complexity of the genetic correlation matrix computation is quadratic in the number of surface points considered. Thus, in order to reduce computation time and make the cluster analysis feasible, we subsampled the standardized cortical surface tessellation from the original 163,842 to 2,562 vertices per hemisphere (i.e., from 8^{th}- to 5^{th}-order icosahedral sub-tessellation, increasing the mean vertex spacing from <1mm to ~6mm).

Clustering is an unsupervised learning task that attempts to partition a given set of objects into subgroups, or clusters, based on similarity. It is primarily a pattern recognition technique for determining whether a natural grouping exists in data. One of many popular algorithms is fuzzy clustering, which finds groups in data, especially when the groups are imperfectly separated and thus a crisp assignment of objects to clusters is sub-optimal. In fuzzy clustering, objects can belong to more than one cluster and with different degrees of membership to the different clusters: between 0 (absolutely doesn't belong) and 1 (absolutely belongs). Thus, the memberships of objects at the overlapping boundaries can express the ambiguity of the cluster assignment.

Clustering methods work jointly with distance functions that measure the dissimilarity of given objects, which is equivalent to measuring their similarity. Thus, one can partition the dataset into clusters based on the chosen proximity relations. We calculated pair-wise genetic correlations of surface area measures between every two vertices on the entire cortex to generate inter-regional genetic correlation matrices for the left and right hemispheres together. Genetic correlations are standardized indices of genetic covariance between measures, in contrast with phenotypic correlations, which reflect both genetic and environmental covariance. We then transformed the genetic correlation matrix into the distance matrix by subtracting genetic correlations from 1. The value of this distance measure ranges between 0 and 2, indicating that two objects are closely related or very different, respectively.

The clustering procedure was performed by the *cluster* package implemented in R (http://www.r-project.org/). Fuzzy clustering aims to minimize the objective function

$$\sum_{v=1}^{k} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} u_{iv}^{r} \, u_{jv}^{r} \, d(i,j)}{2 \sum_{j=1}^{n} u_{jv}^{r}}$$

where n is the number of observations, k is the number of clusters (from 2 to 12, e.g.), r is the membership exponent, u is the cluster membership, and $d(i, j)$ is the dissimilarity between observations i and j (*16*). The cluster memberships u are non-negative and sum to one for a given data point. To investigate the stability of the clustering in relation to initialization, we randomly initialized the algorithm for 100 runs. **Table S1** shows the means, standard deviations and minimums of the objective function among the 100 random initializations. When we have 2-7 clusters in the analysis, identical minima of the objective function was found. The analysis for a larger number of clusters showed some variation, but this variation was very small, which suggested stability in the minima of the objective function that we have found and, consequently, in the results of each cluster solution. Thus, the robustness of the emergence order of clusters should also be benefited from the stability of each cluster solution.

This method is a prototype-less, partitional clustering analysis. The prototype-less procedure does not require prior knowledge of cluster centers. The prototype of a cluster is often a centroid, e.g., the mean of all the points in the cluster. K-means clustering is a commonly used prototype-based clustering technique. One advantage of prototype-less clustering is that the partitioning can be performed directly on the distance matrix rather than on the original data, which is beneficial for us to reduce computation time to recalculate genetic correlations at every iteration. Another advantage is that it can handle non-globular clusters or clusters of different sizes (*17*).

Cluster evaluation

It is important to conduct some kind of cluster validation, because any clustering analysis will find clusters in a data set, even if that data set has no natural cluster structure (*17*). Two cluster properties are usually evaluated: cohesion, which determine how closely related the objects in a cluster are, and separation, which determine how distinct or well-separated a cluster is from other clusters. Quantitative indices called silhouette coefficients combining both cohesion and separation is commonly used to approximately determine the correct number of clusters.

$$s_i = (b_i - a_i)/\max(a_i, b_i)$$

where s_i is the silhouette coefficient for the i^{th} object; a_i is the average distance between the i^{th} object and all other objects in the same cluster; calculate the average distance between the i^{th} object to all the objects in a given cluster and b_i is the minimum value with respect to all clusters. An overall measure of the goodness of a cluster can be obtained by computing the average silhouette coefficient of all objects. The natural number of clusters in a data set can be determined by looking at the number of clusters at which there is a peak in the plot of the silhouette coefficients when it is plotted against the number of clusters.

To validate the stability of the fuzzy clustering results, we employed an entirely different clustering algorithm, spectral clustering (*18*). The results consistently indicated approximate 12 clusters existing in the data set. The 12 cluster solution revealed almost identical patterning to that of fuzzy clustering (**fig. S6**), except that the cluster boundaries were sharp rather than fuzzy. This is a consequence of spectral clustering using a hard clustering algorithm, in which each data point can only belong to one cluster and no partial memberships are allowed. The converging results from two different types of the clustering algorithms provide further evidence for stability of our results.

After clustering the data into 12 clusters, we used these 12 cluster map to parcel the heritability map into 12 genetically-based regions. We then calculated the weighted average heritability for each cluster according to the partial memberships of all the clusters assigned to each vertex (**fig. S7**). We expect that using the genetically-based parcellation system is more optimal than using anatomical atlases to calculate average heritability within regions of interest, although directly testing this supposition is beyond the scope of this study.

Phenotypic clustering

We also explored the phenotypic patterning of the surface area in order to observe both genetic and environmental influences on the patterning, given that phenotypic variation can be

decomposed into genetic and environmental variation. We conducted clustering to phenotypic correlations generated in the same way as genetic correlations did, but the twins were treated as unrelated individuals. Phenotypic correlations simply mean correlations that measure the degree of how two variables (e.g., area measures of two cortical points) vary together. The result of the phenotypic clustering showed that the algorithm cannot find the approximate number of clusters based on silhouette coefficients (**fig. S8**). Furthermore, the fit of the objective function was poorer for phenotypic than genetic clustering. Nevertheless, we specified 12 clusters for the phenotypic clustering based on the result of the genetic clustering. The overall pattern between genetic and phenotypic patterning appeared to be similar (**fig. S9**). The observed similarity provided strong evidence to confirm that the surface area is a heritable trait. However, the result also showed a noticeable difference in the phenotypic patterning with a very asymmetric pattern around the perisylvian regions. Two clusters around this region are exclusively located only in one hemisphere (indicated by * in **fig. S9**). The result may suggest a role of environmental influences on hemispheric asymmetric of the surface area.

Plotting genetic similarity matrix

To visualize the genetic proximity between clusters, we plotted the pair-wise genetic correlation matrix. The rows and columns of the matrix were sorted by the cluster labels. Thus, all data points belonging to the same clusters are grouped together, and the matrix has roughly a block diagonal structure. Each cell in the genetic similarity matrix is the weighted mean of genetic correlations within and between clusters. The patterns displayed in the similarity matrix can reveal the relationship between clusters.

Dendrogram

The dendrogram was produced by single linkage hierarchical clustering using the *hclust* function implemented in R (http://www.r-project.org/). The hierarchical cluster analysis used the weighted mean genetic correlations between and within the 12 clusters to build a hierarchy of clusters. Hierarchical algorithms find successive clusters using previously established clusters, which lack a global objective function, so clustering is decided locally at each step in attempt to approximate a global optimum.

Table S1: 100 random initializations

	Objective Function Value 100 random initializations		
	Mean	Standard Deviation	Minimum
2 cluster solution	1890.311	0.00	1890.311
3 cluster solution	1610.217	0.000	1610.217
4 cluster solution	1428.892	0.000	1428.892
5 cluster solution	1296.005	0.000	1296.005
6 cluster solution	1189.021	0.000	1189.021
7 cluster solution	1097.765	0.000	1097.765
8 cluster solution	1025.115	6.153	1018.990
9 cluster solution	957.099	7.317	952.268
10 cluster solution	898.383	5.544	895.312
11 cluster solution	847.440	7.432	843.013
12 cluster solution	792.790	5.204	790.308

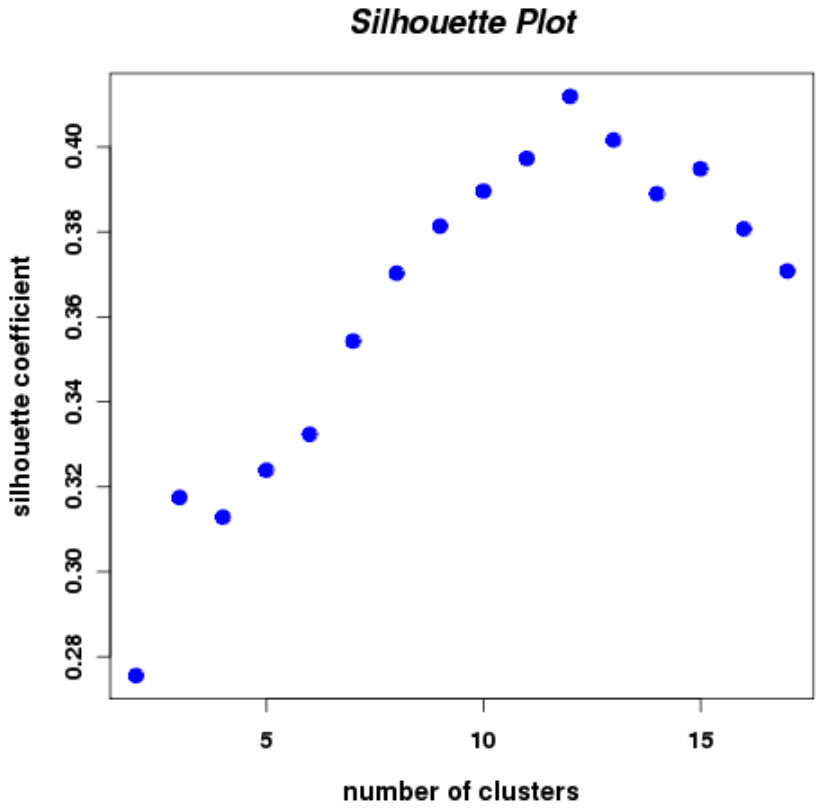

Fig. S1: Silhouette Plot of the genetic clustering

The plot shows the silhouette coefficients against the number of clusters. When the number of clusters reached 12, the silhouette coefficients started to peak.

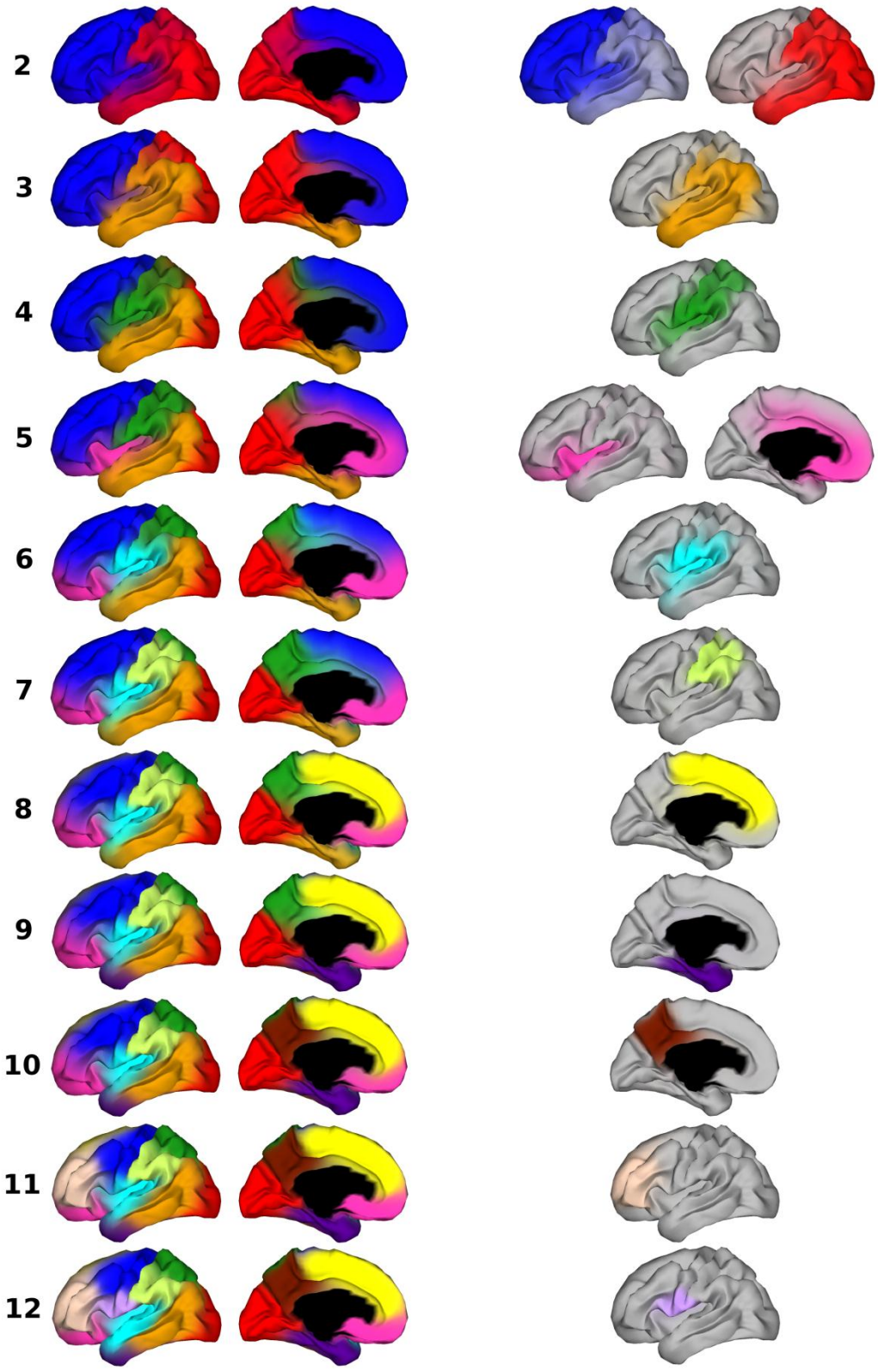
2
3
4
5
6
7
8
9
10
11
12

Fig. S2: The successive fuzzy cluster analysis from 2-12 clusters
The two columns on the left show the cluster maps sequentially from 2 to 12 clusters. The numbers 2-12 at the left of the figure refer to the number of clusters in the sequence of clustering algorithms, not the cluster numbers that are in Fig. 1. The right column shows the new cluster at each level, i.e., the one with the least spatial overlap with the preceding clusters. The extent of spatial overlap was estimated by correlational analyses. Fig. S3 depicts the pattern for each level of the cluster analysis. The 2-cluster solution demonstrated a division between anterior (frontal) and posterior regions (non-frontal). Some examples of nested subdivisions consistent with hierarchical structure are as follows: The 4-cluster solution, which corresponds to the lobar divisions reported in (*19*), is consistent with the non-frontal cluster being subdivided into occipital, temporal and postercentral clusters. Further subdivisions of the basic lobar structure continue in the sucssive cluster solutions. For example, the motor-premotor, dorsolateral prefrontal, orbitofrontal clusters in the 12-cluster solution are all subdivisions of the frontal cluster of the 2-cluster solution. Also, in the 12-cluster solution the superior, posterolateral, and anteromedial temporal clusters are subdivisions of temporal cluster in the 4-cluster solution.

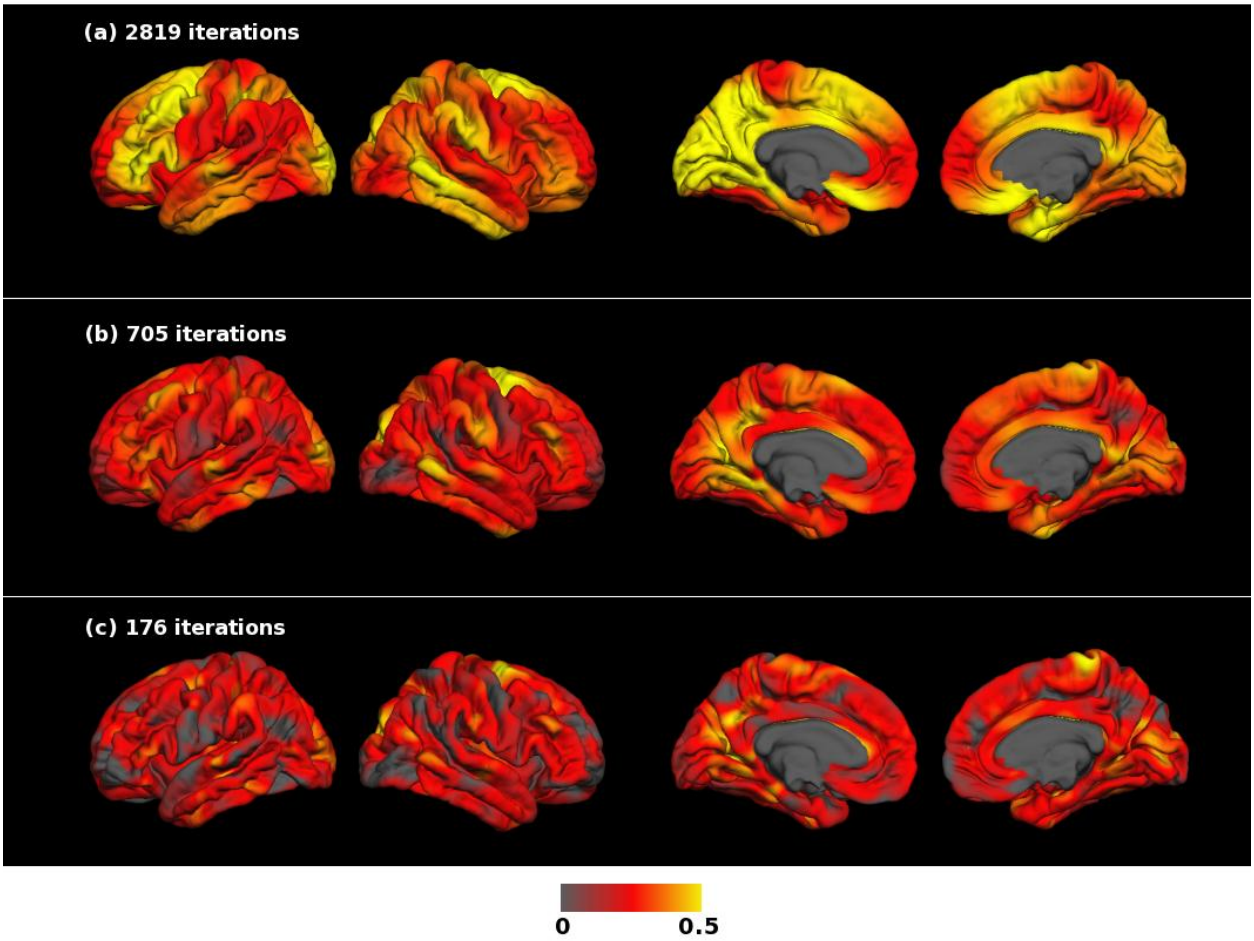

Fig. S3: The effect of smoothing on heritability estimates of relative areal expansion

Statistical power to detect heritability increased with the level of smoothing. Color scale indicates the magnitude of heritability. From (a)-(c), the maps show the heritability estimated on the data with large, mediate and small sizes of smoothing respectively. The areal expansion measures were divided by the total surface area, so the heritability of relative areal expansion was adjusted for global effects (*20*). (a) The heritability using the 2819-iteration: mean ± SD = 0.39 ± 0.103 (left hemisphere); mean ± SD = 0.35 ± 0.087 (right hemisphere), (b) The heritability using the 705-iteration: mean ± SD = 0.24 ± 0.099 (left hemisphere); mean ± SD = 0.24 ± 0.105 (right hemisphere), (c) The heritability using the 176-iteration: mean ± SD = 0.17 ± 0.102 (left hemisphere); mean ± SD = 0.17 ± 0.105 (right hemisphere).

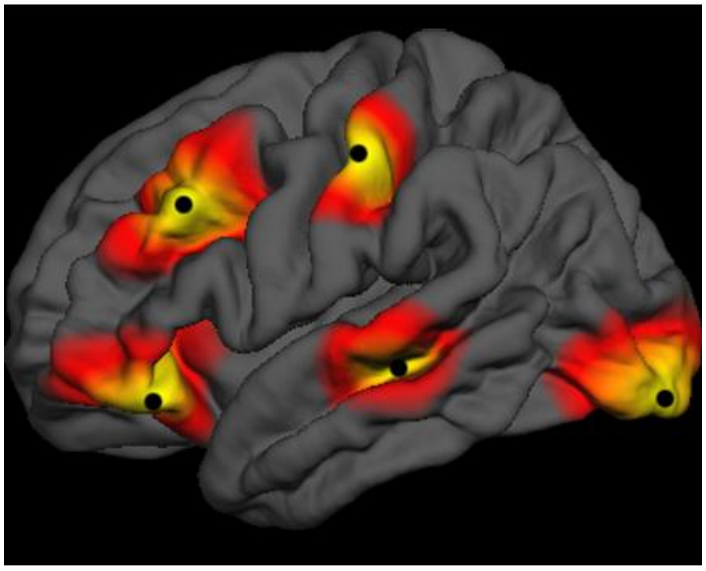

Fig. S4: The map shows examples of smoothing kernels for the 2819-iteration smoothing

The black dots indicate the five locations of individual vertices whos values were initialized to unity prior to smoothing (all other locations were initialized to zero). The color indicates the surface values after smoothing, thresholded at half max, thus illustrating the full width half max (FWHM) dimensions of the effective smoothing kernels. An important advantage of smoothing on the cortical surface, rather than in 3D using an isotropic volumetric kernel, is the improved preservation of cortical maps and sulcal/gyral features (*21*).

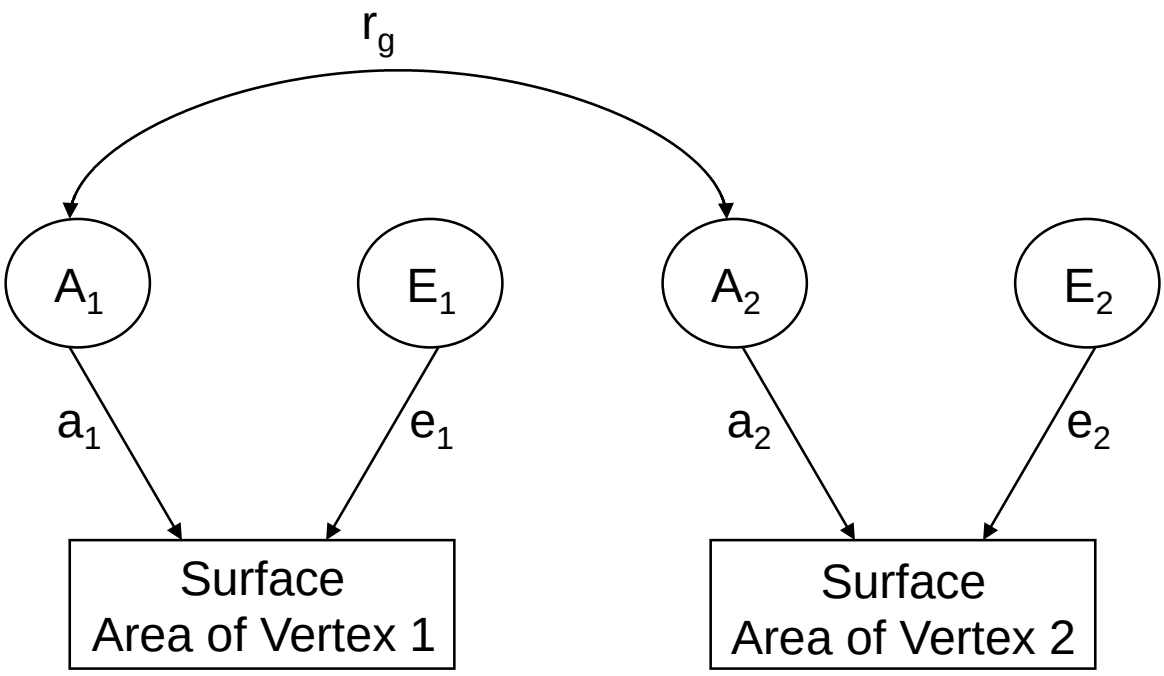

Fig. S5: The basic bivariate AE model

Circles indicate latent variables. Rectangles indicate measured phenotypes. A = additive genetic + non-additive influences; E = non-shared environmental influences; a and e = parameter estimates for contribution of additive + non-additive genetic and non-shared environmental influences, respecitively, to the phenotype; r_g = genetic correlation between the two phenotypes.

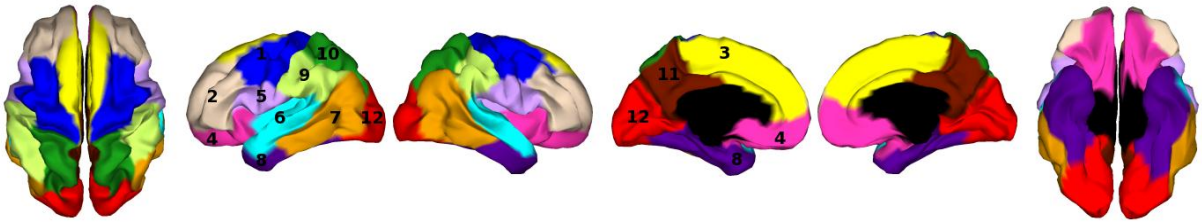

Fig. S6: Genetic clustering map for 12-cluster solution from spectral clustering

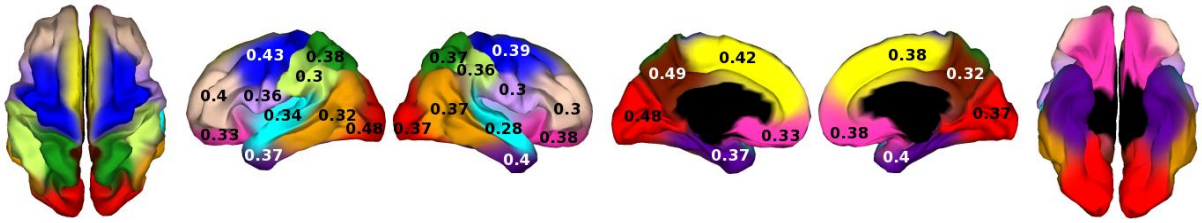

Fig. S7: The weighted average heritability of each cluster

Fig S8: Silhouette Plot of the phenotypic clustering

The plot shows the silhouette coefficients against the number of clusters. In contrast to the genetic clustering, there is absence of a peak in the plot. Also, the silhouette coefficients in the phenotypic clustering were about half of those in the genetic clustering. The approach failed to find the natural number of clusters in the data. It may be due to lack of a well-defined cluster structure in the phenotypic patterning or phenotypic clusters may have atypical structures such as intertwined or overlapping with one another (*17*). In any scenarios, the analysis indicates that there are differences in cluster structures between genetic and phenotypic patterning of the surface area.

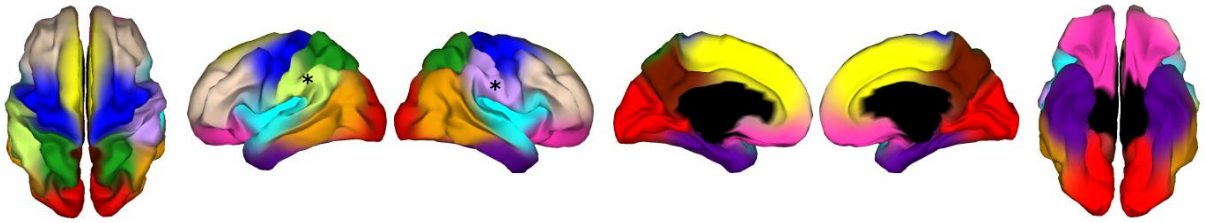

Fig. S9: Phenotypic clustering map for 12-cluster solution from fuzzy clustering

References

1. W. S. Kremen *et al.*, *Twin Research and Human Genetics* **9**, 1009 (2006).
2. M. T. Tsuang, J. L. Bar, R. M. Harley, M. J. Lyons, *Harvard Review of Psychiatry* **9**, 267 (2001).
3. W. G. Henderson *et al.*, *Public Health Reports* **105**, 368 (1990).
4. S. A. Eisen, W. R. True, J. Goldberg, W. Henderson, C. D. Robinette, *Acta Geneticae Medicae et Gemellologiae* **36**, 61 (1987).
5. National Health and Nutrition Examination Survey (NHANES III). (National Center for Health Statistics, 1999-2004), vol. 2007.
6. W. S. Kremen *et al.*, *Neuroimage* **49**, 1213 (2010).
7. M. S. Panizzon *et al.*, *Cerebral Cortex* **19**, 2728 (2009).
8. A. M. Dale, B. Fischl, M. I. Sereno, *Neuroimage* **9**, 179 (1999).
9. B. Fischl, M. I. Sereno, A. M. Dale, *Neuroimage* **9**, 195 (1999).
10. A. M. Dale, M. I. Sereno, *Journal of Cognitive Neuroscience* **5**, 162 (1993).
11. B. Fischl, M. I. Sereno, R. B. H. Tootell, A. M. Dale, *Human Brain Mapping* **8**, 272 (1999).
12. B. Fischl *et al.*, *Cereb Cortex* **18**, 1973 (Aug, 2008).
13. L. M. Rimol *et al.*, *Proceedings of the National Academy of Sciences U.S.A.* **108**, 384 (2010).
14. M. C. Neale, L. R. Cardon, *Methodology for genetic studies of twins and families*. (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1992).
15. M. C. Neale, S. M. Boker, G. Xie, H. H. Maes, *Mx: Statistical Modeling*. (Department of Psychiatry, Medical College of Virginia, Richmond, VA, ed. 6th, 2004).
16. L. Kaufman, P. Rousseeuw, *Finding Groups in Data An Introduction to Cluster Analysis*. (John Wiley & Sons, Inc., 1990).
17. P.-N. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining*. (Pearson Education, Inc., ed. 1st, 2006).
18. A. Ng, M. Jordan, Y. Weiss, *On spectral clustering: analysis and an algorithm*. T. Dietterich, S. Becker, Z. Ghahramani, Eds., Advances in Neural Information Processing Systems 14 (MIT Press, 2002).
19. C. H. Chen *et al.*, *Neuron* **72**, 537 (Nov 17, 2011).
20. L. T. Eyler *et al.*, *Cereb Cortex* **21**, 2313 (Oct, 2011).
21. M. I. Sereno *et al.*, *Science* **268**, 889 (May 12, 1995).