Genetic topography of brain morphology

Chi-Hua Chen,a,b,1 Mark Fiecasaab, E. D. Gutiérrez2, Matthew S. Panizzonaab, Lisa T. Eylerd, Eero Vuoksimiaab,e, Wesley K. Thompsona, Christine Fennema-Notestinea,f, Donald J. Hagler, Jr.f, Terry L. Jerniganas,e,f, Michael C. Nealeg, Carol E. Franzh, Michael J. Lyonsi, Bruce Fischlij, Ming T. Tsuangk,b,e, Anders M. Dalek,d,e, and William S. Kremenab,m,2

Departments of aPsychiatry, cCognitive Science, bRadiology, and fNeurosciences, hCenter for Behavioral Genomics, and hInstitute of Genomic Medicine, University of California, San Diego, La Jolla, CA 92093; iCenter of Excellence for Stress and Mental Health and jVeterans Integrated Service Network 22 Mental Illness Research, Education, and Clinical Center, Veterans Affairs San Diego Healthcare System, San Diego, CA 92161; kDepartments of Psychiatry and Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA 23220; Department of Radiology, Harvard Medical School and Massachusetts General Hospital, Boston, MA 02115; fDepartment of Psychology, Boston University, Boston, MA 02215; lDepartment of Public Health, University of Helsinki, FI-00014, Helsinki, Finland; and mMIT Computer Science and Artificial Intelligence Laboratory, Boston, MA 02139

Edited by Charles F. Stevens, The Salk Institute for Biological Studies, La Jolla, CA, and approved September 5, 2013 (received for review April 30, 2013)

Animal data show that cortical development is initially patterned by genetic gradients largely along three orthogonal axes. We previously reported differences in genetic influences on cortical surface area along an anterior-posterior axis using neuroimaging data of adult human twins. Here, we demonstrate differences in genetic influences on cortical thickness along a dorsal-ventral axis in the same cohort. The phenomenon of orthogonal gradations in cortical organization evident in different structural and functional properties may originate from genetic gradients. Another emerging theme of cortical patterning is that patterns of genetic influences recapitulate the spatial topography of the cortex within hemispheres. The genetic patterning of both cortical thickness and surface area corresponds to cortical functional specializations. Intriguingly, in contrast to broad similarities in genetic patterning, two sets of analyses distinguish cortical thickness and surface area genetically. First, genetic contributions to cortical thickness and surface area are largely distinct; there is very little genetic correlation (i.e., shared genetic influences) between them. Second, organizing principles among genetically defined regions differ between thickness and surface area. Examining the structure of the genetic similarity matrix among clusters revealed that, whereas surface area clusters showed close genetic similarity, cortical thickness clusters correspond to have close genetic relatedness with clusters that have similar maturational timing. The discrepancies are in line with evidence that the two traits follow different mechanisms in neurodevelopment. Our findings highlight the complexity of genetic influences on cortical morphology and provide a glimpse into emerging principles of genetic organization of the cortex.

Significance

How diverse functional cortical regions develop is an important neuroscience question. Animal experiments show that regional differentiation is controlled by genes that express in a graded and regionalized pattern; however, such investigation in humans is scarce. Using noninvasive imaging techniques to acquire brain structure data of genetically related subjects (i.e., twins), we estimated the spatial pattern of genetic influences on cortical structure. We developed a genetic parcellation of cortical thickness to delineate the boundaries of cortical divisions that are—within each division—maximally under control of shared genetic influences. We also found differences in genetic influences on cortical surface area and thickness along two orthogonal axes. The concept of gradients is crucial for understanding the organization of the human brain.


Conflict of interest statement: A.M.D. is a founder of and holds equity in CorTechs Laboratories, Inc., and also serves on its Scientific Advisory Board. The terms of this arrangement have been reviewed and approved by the University of California, San Diego, in accordance with its conflict of interest policies. All other authors report no biomedical financial interests or potential conflicts of interest.

Conflict of interest statement: A.M.D. is a founder of and holds equity in CorTechs Laboratories, Inc., and also serves on its Scientific Advisory Board. The terms of this arrangement have been reviewed and approved by the University of California, San Diego, in accordance with its conflict of interest policies. All other authors report no biomedical financial interests or potential conflicts of interest.

This article is a PNAS Direct Submission.

1To whom correspondence should be addressed. E-mail: chc101@ucsd.edu.

2A.M.D. and W.S.K. contributed equally to this work.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1308091110/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1308091110

PNAS | October 15, 2013 | vol. 110 | no. 42 | 17089-17094
According to the radial unit hypothesis (21), these two traits reflect different mechanisms in cortical development. Cortical surface area is primarily driven by the number of radial columns perpendicular to the pial surface, and cortical thickness is largely determined by the horizontal layers in the cortical columns (including neurons and neuropil) (22). Thus, the human cortex is organized into columnar and laminar patterns with important functional implications. Cortical columns are thought to be functional units of the cortex, and the cortical layers are composed of diverse neuronal populations that organize cortical connectivity (22, 23). Consistent with this hypothesis, twin and family studies indicate that, to a large extent, global measures of surface area and thickness are influenced by distinct genetic contributions (20, 24), compatible with the notion of different developmental processes (25).

Little is known about how genetic influences on surface area and thickness are distributed across the human cortex (i.e., genetic patterning), including whether gradients of genetic influence can be seen for both measures and how these might be similar or different. Previously, we mapped the genetic patterning of surface area based on the MRI scans of over 400 adult twins (26, 27). Using the twin design, which compares monozygotic and dizygotic twins, we can estimate the relative influence of genes and the environment on variance of a phenotype. This method can be further extended to determine the magnitude of genetic and environmental covariance between phenotypes. In the present study, heritability represents the proportion of variance in any location on the cortical surface that is due to genes. Genetic correlations represent shared genetic influences on cortical structure between different points on the cortical surface. The estimated genetic effect is the aggregate effect of all genes as opposed to specific effects of a few individual genes. This approach is advantageous for estimating genetic influences on a complex trait like human brain structure, which probably involves large numbers of genes and possible gene–gene interactions (27). By examining shared genetic influences across the cortex, we demonstrated that maps of genetic patterning of surface area followed specific spatial features comparable with those demonstrated in mouse models, such as an A-P division and four basic cortical regions, which correspond to a mostly lobar pattern in the human cortex (10, 27). Additional partitions were nested within lobes that correspond to structurally and functionally meaningful regions, suggesting a central role of genetic control on regional differentiation. Of note, the genetic divisions were not the same as traditional regions that have been defined on the basis of structure or function.

Here, we adopt a similar analytic procedure in an examination of genetic patterning of cortical thickness in the normal adult human brain. Magnetic resonance (MR) images do not have developmental processes (25).

We constrained the number of clusters to be two, the solution identified a D-V division as the most distinct partition in the basic genetic patterning of cortical thickness. This division separated the motor/premotor, parietal, and occipital regions (the dorsal cluster) from prefrontal and temporal regions (the ventral cluster) (Fig. 1).

We also performed seed-point analyses to map genetic correlations of cortical–thickness measures between selected seed regions and all other cortical locations after adjusting for mean thickness averaged across the entire cortex. The D-V division shown in the two-cluster solution was supported by a multisite seed-point analysis resulting in similar D-V patterns regardless of initial seed locations (Fig. S1).

**Results**

**Genetic Influences on Cortical Thickness: Two-Cluster Solution.** When we constrained the number of clusters to be two, the solution identified a D-V division as the most distinct partition in the basic genetic patterning of cortical thickness. This division separated the motor/premotor, parietal, and occipital regions (the dorsal cluster) from prefrontal and temporal regions (the ventral cluster) (Fig. 1).

We also performed seed-point analyses to map genetic correlations of cortical–thickness measures between selected seed regions and all other cortical locations after adjusting for mean thickness averaged across the entire cortex. The D-V division shown in the two-cluster solution was supported by a multisite seed-point analysis resulting in similar D-V patterns regardless of initial seed locations (Fig. S1).

**Genetic Patterning of Cortical Thickness.** Here, we sought to determine the most appropriate number of clusters to explain patterns in the genetic correlation data based on several different clustering algorithms. These algorithms included spectral cluster analysis and the silhouette coefficients calculated from the results of fuzzy clustering. After testing the stability of the clustering using different approaches (Fig. S2), which all identified correled (i.e., under control of shared genetic influences on cortical thickness). The resulting cortical maps of cluster or seed-point analyses are referred to as the genetic patterning of cortical thickness, i.e., the spatial layout of the maps. It is worth clarifying at the outset that the term “patterning” is also used to refer to molecular or cellular differences between functional cortical areas as a result of a developmental process (9, 29). However, here we use genetic patterning to describe the spatial distribution of genetic correlation based on MRI data and modeling tools used in this study. Second, we explore the organization among the genetically based clusters, i.e., the genetic relationships among the clusters. In particular, we use a dendrogram based on the genetic similarity matrix among clusters to inspect the genetic proximity between clusters to search for underlying organizational principles related to genetic control over cortical thickness (e.g., whether clusters that are near to each other on the brain surface are more genetically related than those that are farther away). Third, we compare genetic influences on cortical thickness and surface area to see whether the spatial layout of the maps (i.e., the genetic patterning) looks the same and, independent of this result, whether the genetic relationships among the clusters (i.e., the genetic organization) are similar between the two brain phenotypes. Thickness and surface area could, for example, have similar spatial layouts (patterning) but the genetic correlations between parcels comprising that layout (organization) might be different. Finally, we examine the degree to which genetic influences on thickness and surface area within the identified genetic clusters are shared across the two measures to discover whether, within genetically meaningful boundaries, there are unique genetic influences on these two cortical dimensions.
SI Methods and Figs. S3 and S4), we focused on the cluster map of the 12-cluster solution based on the adjusted data (Fig. 2). We chose 12 clusters because (i) a lower number is more parsimonious, all other things being equal; and (ii) it enables a direct comparison with our previous report on the 12-cluster map of cortical surface-area measures (26). However, it is worth noting that whether we chose 11 or 12 clusters did not alter the interpretation of results in the present study.

**Organization Among Genetic Clusters.** We calculated the genetic similarity matrix to determine the genetic relatedness between clusters and constructed a dendrogram, a tree-structured graph, derived from hierarchical clustering to summarize the genetic relations among them (Fig. 3). The most distinct genetic partitions located at the highest level of the hierarchy corresponded to the D-V division. The ventral division included the temporal and prefrontal clusters. The dorsal division included the occipital, parietal clusters, and motor/premotor clusters. Below that, there were four branches including (i) motor/somatosensory/parietal/parisylvian, (ii) occipital lobe, (iii) temporal/ventral frontal, and (iv) dorsal prefrontal regions. Finally, at the bottom of the dendrogram, there were clusters related to human functional specialization. These divisions can also be observed through the progression of cluster solutions, from 2 to 12 clusters (Fig. S5).

**Comparisons Between Genetic Patterning of Cortical Thickness and Surface Area.** The genetic patterns of cortical thickness and surface area were similar, especially in the lateral aspect of the cortical surface; however, differences between the patterning were noted in the medial wall (Fig. 2). The spatial extent of clusters also differed. For example, cortical thickness clusters near the midline extended to both the lateral and medial parts of the hemispheres, whereas the most corresponding surface-area clusters were restrictively placed on either side of the hemispheres. We computed the Rand index to

---

**Fig. 2.** Genetic clustering maps. Cortical thickness: 1, motor-premotor-supplementary motor area; 2, superior parietal cortex; 3, inferior parietal cortex; 4, perisylvian region; 5, occipital cortex; 6, ventromedial occipital cortex; 7, ventral frontal cortex; 8, temporal pole; 9, medial temporal cortex; 10 middle temporal cortex; 11, dorsolateral prefrontal cortex; 12, medial prefrontal cortex. Surface area: 1, motor-premotor cortex; 2, dorsolateral prefrontal cortex; 3, dorsomedial frontal cortex; 4, orbitofrontal cortex; 5, pars opercularis and subcentral region; 6, superior temporal cortex; 7, posteroslateral temporal cortex; 8, anteromedial temporal cortex; 9, inferior parietal cortex; 10, superior parietal cortex; 11, precuneus; and 12: occipital cortex (26).

**Fig. 3.** Organization among genetic clusters. (Upper) The dendrograms derived from hierarchical clustering based on genetic correlations among clusters. (Lower) The genetic clustering maps show anatomical locations of the clusters, and the heatmaps represent the weighted mean genetic correlations within and between clusters.
calculate the similarity between the clustering for cortical thickness and surface area. Despite the medial differences, it showed a considerable degree of similarity (Rand index = 0.91; adjusted Rand index = 0.44) (30, 31). The Rand index is a number between 0 and 1 that quantifies the degree of similarity between the two cluster solutions by computing the proportion of vertices that were given the same cluster label in both cluster solutions. It is possible for some vertices to have the same cluster label by chance, and this is accounted for in the adjusted Rand index.

**Distinct Genetic Contributions to Cortical Thickness and Surface Area.** The cortical surface was divided into 12 cortical subregions for thickness and area separately using the corresponding genetically based cluster map of each trait. We calculated the weighted average cortical thickness and surface area for each cluster according to the partial memberships of all of the clusters assigned to each vertex. We then estimated genetic correlations between cortical thickness and surface area. The genetic correlation between cortical thickness and surface area for the dorsolateral prefrontal cluster was 0.01 [95% confidence interval (CI): 0.13–0.16], and for the inferior parietal cluster it was 0.15 [95% CI: 0.01–0.28]. These two clusters were selected because they showed the highest degree of similarity in boundaries between the two traits. Applying these comparisons to the same cortical regions showed an interesting contrast: namely, that the boundaries of genetic clusters can be very similar between the two traits but the genetic determinants underlying them appear quite distinct.

**Discussion**

**Genetic Influences on Cortical Thickness Showed a Dorsal-Ventral Division (Two-Cluster Solution).** The most substantial genetic partition of cortical thickness corresponded to a D-V division. Although the boundary of the D-V division did not match any cortical folding patterns, it did correspond to a border characterized by cytoarchitectonic features, which separates the granular and agranular cortex in the frontal lobe (11). The granular cortex is defined by the presence of a granule cell layer IV (11). The prefrontal region is classified as the granular cortex whereas the motor-premotor region is classified as the agranular cortex (11). Thus, differences in cytoarchitecture might reflect the D-V genetic boundary of cortical thickness.

The genetic D-V division of cortical thickness may also be related to cortical connectivity patterns. For example, despite their lack of spatial contiguity, prefrontal and temporal regions’ cortical thickness had similar genetic influences. These regions are known to be structurally and functionally connected. This result is consistent with our previous report based on a limited seed-point analysis, which showed that the pattern of genetic correlations for cortical thickness partially correspond to neuroanatomical connectivity, with high genetic correlations between distal, noncontiguous regions (28). The cross-regional genetic patterning of cortical thickness may relate to underlying fiber tract structures (e.g., thalamocortical or intracortical connections).

We hypothesize that the D-V division of genetic influences on cortical thickness may relate to morphogenetic gradients along the D-V axis. The genetic divisions or gradients are in line with the cytoarchitectonic gradation. The gradation hypothesis describes gradations from the motor cortex toward the more rostrally located frontal regions with increasing granularity, and also suggests that the cytoarchitectonic changes are accompanied by myeloarchitectonic changes. Ample evidence on cortical gradients seen in genetic patterning, morphology, connectivity, and function provides a unique insight on the organization of the human brain (17, 19, 32).

The Genetic Patterning of Cortical Thickness Largely Corresponds to Functional Specializations (12-Cluster Solution). Without any incorporation of anatomical knowledge in the clustering algorithm, the genetically based subdivisions corresponded closely to meaningful structural and functional regions (Fig. 2). Some cluster boundaries mapped onto traditionally parcellated regions, such as cytoarchitectural areas (e.g., Brodmann area (BA) 7) or gyral patterns, whereas others did not. Evidence of genetic expression domains recapitulating phenotypic boundaries has been observed in various body parts and diverse species (33–35). These findings partly address a challenging question about how genes shape initially homogeneous cells into different body parts (36). The close correspondence between genetic patterning and phenotypic boundaries is consistent with the idea that body patterning is blueprinted by genetic patterning (21, 37). We previously found such evidence in the genetic patterning of cortical surface area (26, 27), and here we see similar results for cortical thickness. Gene expression profiles between regions recapitulating the spatial topography of the cortex were also shown in an atlas of the adult human brain transcriptome (38). The genes included in that study coded proteins involved in many different aspects of structure and function other than cortical thickness and surface area, yet the genetic topography corresponded to the spatial topography of the brain.

All Genetic Divisions Were Bilaterally Located in Homologous Regions. The genetic patterning of cortical thickness was predominantly bilaterally symmetrical. All clusters were bilaterally located in the homologous regions between hemispheres. Because the clustering algorithm is blind to location in space and the clustering was conducted on both hemispheres simultaneously, with no constraint for hemispheric symmetry, the results clearly indicate a predominantly bilaterally symmetrical feature. This characteristic is similar to that of the genetic patterning of cortical surface area, which also showed a predominantly symmetrical feature (26, 27).

Comparisons Between Genetic Patterning of Cortical Thickness and Surface Area. The boundaries of genetic clusters of cortical thickness were most similar to those of surface area in the lateral parts of the hemispheres, except for the temporal clusters (Fig. 2). Differences were most notable in the medial part of the hemispheres. At first glance, it may seem surprising to see similar genetic patterning between cortical thickness and surface area, given that they are very different measures along two orthogonal dimensions on the cortical ribbon. The observed similarity in genetic patterning between surface area and thickness might be explained by results showing that both genetically informed parcellations largely conform to functionally specialized domains of the human cortex.

It is also important to note that the genetic patterning of thickness and surface area was not identical. It remains unclear as to why genetic parcellations of these measures are most similar on the lateral surface, and least on the medial surface. Possibly, these differences in genetic patterning may be due to distinct genetic mechanisms controlling these two traits during development. Also possible are methodological limitations regarding low spatial resolution in MR measurements. There may be more genetic divisions than can be detected with our method. Our analysis can demarcate only the broad and prominent landscape of genetic patterning; thus, it is hard to predict how fine details that are missed in our analysis come into play in the observed discrepancies.

Similar Genetic Patterning but Distinct Genetic Organization and Contributions for Thickness and Surface Area. Although the genetic patterns of the surface area and thickness 12-cluster solutions show a high degree of correspondence, the genetic organization...
among the clusters is fundamentally different (Fig. 3). We saw a
D-V division as the most distinguished partition in the genetic
patterning of cortical thickness, in contrast to the A-P division
observed in the genetic patterning of cortical surface area. The
dendrogram structure indicated that the genetic divisions of
surface area are largely organized by four lobes such that the
clusters that belong to the same lobes are genetically more
similar to one another than to those from different lobes.
However, the genetic divisions of thickness did not follow this
organizing principle. Neither the A-P division nor the lobes
were present in the dendrogram structure of cortical thickness.
Clusters from different lobes showed close genetic proximity.
For example, the motor and somatosensory clusters were first
merged together in the dendrogram (see thickness cluster 1 and
2 in Fig. 3). Unlike the surface area clusters, spatially discon-
nected cortical-thickness clusters from different lobes also show
close genetic relatedness such as anterior temporal and ventral
prefrontal regions (see thickness cluster 7, 8, and 9 in Fig. 3).
The organizational principles of thickness clusters appear to
reflect maturation timing or primary vs. association cortex.
Sensorimotor region and frontal pole, temporal, and occipital
poles mature before higher-order association areas (12, 39).
Interestingly, we indeed found that the genetic clusters in
regions that mature early are grouped together (e.g., primary
motor merged with somatosensory clusters, anterior temporal
merged with ventral frontal including frontal pole, and two
occipital clusters merged together).

The patterning of high cross-regional genetic correlations was
a feature of the genetic organization of cortical thickness.
In contrast, we have demonstrated predominantly spatially contig-
uous patterns of genetic correlations for cortical surface area (26,
27). From the sequential cluster analyses from 2- to 12-cluster
solutions, the way that each additional cluster was “carved out
from” or “fit into” the previous cluster solution differed greatly
between surface area and thickness measures, even though the
eventual 12-cluster solution was very similar in patterning (Fig.
S5). This result is additional support for the conclusion of distinct
organizational principles for genetic relations among clusters
between surface area and thickness.

Furthermore, we found very little genetic overlap between
measures of regional cortical thickness and surface area, i.e., very
low genetic correlations between them, consistent with our pre-
vious report (20), as well as a report from an independent sample
(24). In the current analysis, we calculated the genetic correla-
tions using the genetic divisions defined by the cluster analyses.
Even within largely spatially corresponding regions based on the
genetic cluster maps, we failed to find evidence of common gen-
etic influences between surface area and thickness. It is worth
noting that minimal genetic overlap between cortical thickness
and surface area is most evident for global measures. In our
previous study, we did find evidence for nonzero genetic corre-
lations between cortical thickness and surface area for some
regional measures (e.g., lateral orbital frontal cortex) (20). It is
therefore possible that there are common genes related to these
traits locally in certain cortical regions or specific layers, which
we could not detect with the present analysis. For example, ro-
dent studies show that genes, controlling for area patterning, also
regulate thickness in certain cortical layers (40–42). The mol-
ecular and cellular mechanisms in cortical development are
complex and intricate and cannot be fully appreciated by MRI
measures that focus on the examination of brain structure at the
systems level.

Distinct organization among genetic clusters and few shared
genetic influences between the two traits could be explained by
separate underlying developmental trajectories. According to the
radial-unit hypothesis, the overall expansion of surface area is
largely determined by the number of radial cortical columns,
starting from a layer of founder neural stem cells. The exponential
growth of founder cells occurs before the onset of neurogenesis.
After the onset of neurogenesis, the number of neurons in-
creases linearly, derived from founder cells within each col-
umn that determine initial cortical thickness (43). Thus, our
current knowledge suggests that different genetic mechanisms
and time courses during neurodevelopment are involved in the
features of cortical thickness and surface area.

Conclusion. The human cortex is a complex structure with many
cortical functional areas that are distinguishable by differences in
gene expression, morphology, and connectivity (44–46). Despite
the existence of a considerable degree of cortical plasticity, our
study supports the concept of early genetic determination of
cortical regionalization. We have examined the genetic pattern-
ning and organization of cortical morphology along two orthog-
onal dimensions, the tangential (horizontal) expansion of surface
area and the radial (vertical) growth of cortical thickness. We
showed that the genetic patternings of cortical thickness and
surface area are largely similar to one another and consistent
with functional specialization boundaries. However, the organi-
zational principles among genetic clusters differ between the two
traits. The genetic effects that influence cortical thickness and
surface area were largely distinct, in line with previous findings.
Further study of genetic patterning with different age groups is
warranted to determine whether the pattern changes during
development or aging. Our collective understanding of the ge-
netic underpinnings of complex human traits is in its infancy.
The current findings provide a glimpse into the genetic architec-
ture of the human cortex and may facilitate future genomic in-
vestigation into diverse complex traits of the human brain.

Methods
Participants and Imaging Data. The participants were 406 middle-aged men
from the Vietnam Era Twin Study of Aging (VETSA) (47). There were 110
monozygotic and 93 dizygotic twin pairs. Based on demographic and health
characteristics, the sample is representative of US men in their age range
(47). T1-weighted MR images were acquired on Siemens 1.5 Tesla scanners at
University of California, San Diego and Massachusetts General Hospital. The
data were analyzed using the FreeSurfer software package to calculate
cortical thickness measures (48, 49). In brief, the cortical surface was recon-
structed to measure thickness at each surface location, or vertex. The
resulting surface was covered with a polygonal tessellation. The boundary of
the gray and white matter was identified and then deformed outwards to
obtain a representation of the pial surface. Each subject's cortical surface
was aligned to an atlas space in a spherical surface-based coordinate system.
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 (50). The thickness of the gray matter can be
computed at any point in the cortex as the shortest distance between the
graywhite boundary and pial surfaces (2). See full method in SI Methods.

Twin Analysis. We estimated genetic correlations of cortical-thickness meas-
ures between brain areas using Mx, a structural equation modeling software
for genetically informative data (51). In brief, a standard bivariate twin (AE)
model was used to estimate the proportion of phenotypic variance between
vertices accounted for by additive genetic effects (A) and individual-specific
environmental effects (E) for each measure (S2). Before the model fitting, the
cortical thickness data were adjusted for age and site effects, and then nor-
malized to adjust for global effects (i.e., the average cortical thickness was
subtracted from the vertex-wise data at each cortical location).

Cluster Analysis. Clustering methods partition the dataset into clusters based
on the pair-wise genetic correlations of thickness measures between every
two vertices. The fuzzy clustering procedure was performed by the cluster
package implemented in R (www.r-project.org). A commonly used quanti-
tative index called the silhouette coefficient was used to approximately
determine the correct number of clusters. We also used an entirely different
clustering algorithm called spectral clustering to validate the fuzzy cluster-
ning results (53, 54). See full method in SI Methods.
ACKNOWLEDGMENTS. This work was funded by National Institute on Aging Grants AG02881, AG018836, AG018384, AG022982, and AG031224; National Institute on Drug Abuse Grant DA029475; National Institute of Neurological Disorders and Stroke Grant NS055883; and Academy of Finland Grant 257075 (to E.V.). The US Department of Veterans Affairs has provided financial support for the development and maintenance of the Vietnam Era Twin (VET) Registry.

45. Brodmann K (1909) Vergleichende lokalisationlehre der grosshirnrinde (Barth, Leipzig).
Supporting Information

Chen et al. 10.1073/pnas.1308091110

SI Methods

Participants. The Vietnam Era Twin Study of Aging (VETSA) project has been described previously (1). The VETSA sample was drawn from the Vietnam Era Twin (VET) Registry (2), a sample of male–male twin pairs born between 1939 and 1957 who had both served in the United States military between 1965 and 1975. The study sample is not a VA or patient group; the majority of individuals were not exposed to combat. For this analysis, 474 individual VETSA participants were included. Of those, 406 were paired (i.e., 203 twin pairs): 110 monozygotic (MZ) and 93 dizygotic (DZ) pairs. Zygosity for 92% of the sample was determined by analysis of 25 satellite markers that were obtained from blood samples. For the remainder of the sample, zygosity was determined through a combination of questionnaire and blood-group methods (3).

Mean age of the MRI participants was 55.8 (2.0) years (range, 51–59), mean years of education was 13.9 (SD = 2.1). There were 88.3% non-Hispanic white, 5.3% African-American, 3.4% Hispanic, and 3.0% “other” participants. There were no significant demographic differences between MZ and DZ twins (1,4). The VETSA sample is representative of US men in their age range based on sociodemographic and health characteristics determined by US census and Center for Disease Control data (1,4). 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. Sagittal T1-weighted MPRAGE images were acquired on Siemens 1.5 Tesla scanners (241 at University of California, San Diego; 233 at Massachusetts General Hospital). Scan parameters were: Inversion time (TI) = 1,000 ms, echo time (TE) = 3.31 ms, repetition time (TR) = 2,730 ms, flip angle = 7 degrees, slice thickness = 1.33 mm, voxel size 1.3 × 1.3 × 1.3 mm. Data were reviewed for quality, registered, and averaged to improve signal-to-noise. Of the 493 scans available at the time of these analyses, quality control measures excluded 0.6% (3 cases) due to scanner artifact and 3% (16 cases) due to inadequate image-processing results (e.g., poor contrast caused removal of nonbrain to fail). The resultant 474 available cases included 203 twin pairs (406 individuals) that were used in the present study.

Image Processing. The cortical surface was reconstructed to measure surface area and cortical thickness at each surface location (a total of more than 160,000 locations for each hemisphere) using a semiautomated approach provided by the FreeSurfer software (5–7). Variations in image intensity due to radio frequency (RF) coil sensitivity inhomogeneities were corrected, a normalized intensity image was created, and the skull (nonbrain) was removed from this image. A preliminary segmentation was then partitioned using a semiautomated approach. The lesion was then removed from the data set at each cortical location, 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 (6). 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 (8). 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 (9). The thickness of the gray matter can be computed at any point in the cortex as the shortest distance between the gray/white and pial surfaces. Finally, the vertex-wise 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 estimates. We used three levels of smoothing from small to large corresponding to 176, 705, and 2,819 iterations, respectively, of nearest-neighbor smoothing on the standardized atlas tessellation. The 2,819-iteration was found to be of the smallest order that can yield sufficiently high heritability values to enable accurate estimation of genetic correlations, which are critical for the stability of the subsequent cluster analysis.

Twin Analysis. Based on our previous findings of minimal common environmental influences on surface area (10, 11) and cortical thickness (12, 13), we used a twin model that estimated contributions of additive genetic effects (A) and individual-specific environmental effects (E) to the variance in cortical thickness at each vertex. The variance–covariance patterns were examined by fitting models with Mx, a maximum-likelihood-based structural equation modeling program (14). We sought to map shared genetic effects on cortical-thickness measure between each pair of cortical locations. To accomplish this aim, univariate AE models are easily extended to the bivariate case (15). In addition to genetic and environmental sources of variance, genetic and environmental sources of covariance can also be examined in the bivariate model. In the present study, we used bivariate models to compute genetic correlations of cortical-thickness measures between each pair of 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.

Before the model fitting, the cortical thickness data were adjusted for age and site effects and then normalized to adjust for global effects (i.e., the average cortical thickness was subtracted from the vertex-wise data at each cortical location). The thickness measure at each location was then standardized to z-scores across all subjects.

Fuzzy Cluster Analysis. Clustering methods partition the dataset into clusters based on the chosen proximity relations. We calculated pair-wise genetic correlations of thickness measures between
every two vertices on the entire cortex to generate interregional
genetic correlation matrices for the left and right hemispheres
simultaneously. To reduce computation time and make the cluster
analysis feasible, we subsampled the standardized cortical-surface
tessellation from the original 163,842–2,562 vertices per hemi-
sphere. We then transformed the genetic correlation matrix into
the distance matrix by subtracting each genetic correlation in the
genetic correlation matrix from 1. The value of this distance
measure ranges between 0 and 2, indicating that two objects are
closely related or very different, respectively. In fuzzy clustering,
objects can belong to more than one cluster and with different
degrees of membership to the different clusters: between 0 (ab-
solutely doesn’t belong) and 1 (absolutely belongs). Thus, the
memberships of objects at the overlapping boundaries can ex-
press the ambiguity of the cluster assignment.

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

\[ \sum_{i=1}^{k} \sum_{j=i+1}^{n} \frac{1}{2} u_{i}^{1} u_{j}^{1} d(i, j) \]

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 ob-
servations \( i \) and \( j \) (16). The cluster memberships \( u \) are nonneg-
ative and sum to one for a given data point. To investigate the
stability of the clustering in relation to initialization, we ran-
domly initialized the algorithm for 100 runs and picked the clus-
ter solution that maximized the likelihood function.

Silhouette Coefficient. Two cluster properties are usually evalu-
ated: cohesion, which determines how closely related the objects in
a cluster are, and separation, which determines how distinct or
well-separated a cluster is from other clusters. Quantitative in-
dices called silhouette coefficients combining both cohesion and
separation are commonly used to approximately determine the
correct number of clusters (16, 17). The silhouette coefficients
can be computed by

\[ s_{i} = (b_{i} - a_{i}) / \max(a_{i}, b_{i}) \]

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 of the objects in a given cluster and \( b_{i} \) is the mini-
imum 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 clus-
ters in a dataset 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.

Spectral Cluster Analysis. To validate the stability of the fuzzy
clustering results, we used an entirely different clustering algo-
rithm called spectral clustering (18, 19). Instead of directly using
distance matrices like in fuzzy clustering, spectral clustering
methods transform distance matrices into affinity matrices (ref.
18); for reviews, see ref. 20). Each element of an affinity matrix
\( A \), as defined by Ng et al., can be computed by

\[ A_{ij} = \exp\left(-\frac{||s_{i} - s_{j}||^2}{2\sigma^2}\right) \]

The affinity matrix \( A \) measures the affinity between data points
\( s_{i} \) and \( s_{j} \). In our study, the distance between \( s_{i} \) and \( s_{j} \) is equal to 1
minus genetic correlations. \( \sigma \) is a scaling parameter. The affinity
matrix is then normalized \( L = D^{-1/2}AD^{-1/2} \) where \( D \) is a diagno-
mal matrix with \( D_{ii} = \sum_{j} A_{ij} \). The dataset is now represented
in an eigenspace using the top eigenvectors of the \( D \) matrix (in this
study, we used the top 30 eigenvectors) (18). Projecting the data
into the eigenspace of the affinity matrix allows us to capture the
primary distribution of the original data points. We would like to
ensure that the selected eigenvectors are sufficiently relevant to
inform the variation of data distribution. We have validated the
relevance of the top 30 eigenvectors using an eigenvector selec-
tion algorithm to confirm that the selected eigenvectors were
informative to separate data into groups (19). Including only
top and relevant eigenvectors makes clustering results less sus-
cceptible to noise in the data (19).

Next, we used Gaussian mixture models to cluster the data into
different groups using the new data representation in the
eigenspace. Gaussian mixture models are widely used in statistics
for clustering data as a model-based clustering algorithm (17).
The number of clusters was approximated by evaluating the
number of Gaussian components using Bayesian Information
Criterion (BIC) (19). In the context of high-dimensional data,
fitting Gaussian mixture models is challenging due to the large
number parameters relative to the small sample size. Thus,
regularization was used to improve estimation of the covariance
matrices in Gaussian mixture models. Moreover, because the
penalty term in the BIC will dominate the likelihood function
because of the large number of parameters, we instead used a 10-
fold cross-validation procedure. We used nine-tenths of the data
as a training set to fit the Gaussian mixture models and then
evaluated the fit of the model via the penalized likelihood on the
one-tenth of the data held out as a testing set. This procedure
was repeated until all data points have served as part of the
testing set. The results indicated ~11–15 clusters existing in the
dataset depending on the methods of regularization.

Determining the Number of Clusters. Here, we sought to determine
the most appropriate number of clusters to explain patterns in the
genetic correlation data based on the silhouette coefficient, which
evaluates how appropriate a cluster solution is based on maxi-
mizing within-cluster cohesion and between-cluster separation
(17). The silhouette plot, based on the clustering in the un-
adjusted thickness data before controlling for global effects,
demonstrated a plateau starting around the 11–12 cluster solu-
tion (Fig. S3). The adjustment for global cortical thickness ef-
effects allows us to examine region-specific genetic effects. For
data adjusted for average cortical thickness, the coefficient began
to reach a plateau around the 12–14 cluster solutions (Fig. S4).
However, it continued to increase beyond even the 20-cluster
solution. We suspect that the continuous increase in the sil-
houette coefficient after the 20-cluster solution may be due to
lower heritability estimates for each cortical location after ad-
justing for global effects. Some locations may have low herita-
bility estimates, and therefore less reliable genetic correlation
estimates.

To validate the number of clusters, we also used spectral
clustering, which uses a fundamentally different clustering al-
gorithm. Spectral clustering relies on the eigen-structure of the
similarity matrix rather than on the original similarity matrix to
partition points into clusters (18). An important advantage of the
spectral algorithm is its robustness to atypical cluster features
and noise in the data (18). Spectral clustering was performed for
a simultaneous cluster-number estimation and data clustering
based on Gaussian mixture models. Results yielded 11–15 clus-
ters, confirming that the true number of clusters is unlikely to be
more than 20. The genetic cluster maps derived from these ap-
proaches revealed very similar features (Fig. S2), providing fur-
ther evidence of the converging results.
Genetic Similarity Matrix and Dendrogram. 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. The dendrogram was then produced by hierarchical cluster analysis.


Fig. S1. A grid placement of genetic correlation maps from 24 different seeds. We also performed seed-point analyses to map genetic correlations of cortical-thickness measures between selected seed regions and all other cortical locations after adjusting for mean thickness averaged across the entire cortex. Genetic correlation maps are a simple way to visualize the genetic patterning because the color codes directly reflect the strength of genetic correlations between the seed regions and all other cortical points. For the selection of seed points, we used a grid of regularly spaced seeds distributed across the entire lateral aspect of one cortical hemisphere. The location of each small genetic correlation map on the gray brain map represents the location of the seed for that correlation map. Color scale indicates the strength of genetic correlations between the surface area at the seed region and at all other locations on the cortical surface; these correlations range from positive to negative.
Fig. S2. Thickness clustering derived from different clustering approaches. (Top) Fuzzy clustering in the adjusted data. (Middle) Fuzzy clustering in the unadjusted data. (Bottom) Spectral clustering in the adjusted data.

Fig. S3. Silhouette Plot of the genetic clustering based on unadjusted data. The plot shows the silhouette coefficients against the number of clusters. When the number of clusters reached 11–12, the silhouette coefficients started to reach a plateau.
Fig. S4. Silhouette Plot of the genetic clustering based on adjusted data. The plot shows the silhouette coefficients against the number of clusters. When the number of clusters reached 12–14, the silhouette coefficients started to reach a plateau.
Fig. S5. The order of emergence of the 12 clusters. The two columns on the Left show the cluster maps sequentially from 2 to 12 clusters for surface area. The two columns on the Right show the cluster maps sequentially from 2 to 12 clusters for cortical thickness.