

ORIGINAL ARTICLE

Genetic Effects on Fine-Grained Human Cortical Regionalization

Yue Cui^{1,2,†}, Bing Liu^{1,2,†}, Yuan Zhou^{4,5,†}, Lingzhong Fan^{1,2}, Jin Li^{1,2}, Yun Zhang⁶, Huawang Wu^{6,7}, Bing Hou^{1,2}, Chao Wang⁶, Fanfan Zheng^{1,2}, Chengxiang Qiu^{1,2}, Li-Lin Rao^{4,5}, Yuping Ning⁷, Shu Li^{4,5} and Tianzi Jiang^{1,2,3,6,8}

¹Brainnetome Center, ²National Laboratory of Pattern Recognition, ³CAS Center for Excellence in Brain Science, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, ⁴Key Laboratory of Behavioral Science, ⁵Magnetic Resonance Imaging Research Center, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China, ⁶Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, China, ⁷Guangzhou Brain Hospital, The Affiliated Brain Hospital of Guangzhou Medical University, Guangzhou 510370, China and

⁸Queensland Brain Institute, University of Queensland, Brisbane QLD 4072, Australia

Address correspondence to Tianzi Jiang, Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.
Email: jiangtz@nlpr.ia.ac.cn; Shu Li, Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China.
Email: lishu@psych.ac.cn

[†]Y.C., B.L., and Y.Z. contributed equally to this work.

Abstract

Various brain structural and functional features such as cytoarchitecture, topographic mapping, gyral/sulcal anatomy, and anatomical and functional connectivity have been used in human brain parcellation. However, the fine-grained intrinsic genetic architecture of the cortex remains unknown. In the present study, we parcellated specific regions of the cortex into subregions based on genetic correlations (i.e., shared genetic influences) between the surface area of each pair of cortical locations within the seed region. The genetic correlations were estimated by comparing the correlations of the surface area between monozygotic and dizygotic twins using bivariate twin models. Our genetic subdivisions of diverse brain regions were reproducible across 2 independent datasets and corresponded closely to fine-grained functional specializations. Furthermore, subregional genetic correlation profiles were generally consistent with functional connectivity patterns. Our findings indicate that the magnitude of the genetic covariance in brain anatomy could be used to delineate the boundaries of functional subregions of the brain and may be of value in the next generation human brain atlas.

Key words: cortical regionalization, genetic correlation, genetics, surface area, twins

Introduction

Mapping fine-grained, anatomically distinct, and functionally specialized cortical subregions is fundamental for understanding brain function. Anatomical microstructure is currently primarily used to define cortical boundaries, and cyto-, myelo-, and receptor-architectonic maps have become the “gold standard” for

cortical parcellation (Brodmann 1909; Amunts et al. 2013). In addition, many other techniques for parcellating the human brain, such as topographic mapping (Wandell and Winawer 2011), gyral/sulcal anatomy (Van Essen et al. 2012), and anatomical (Behrens et al. 2003) and functional (Kim et al. 2010) connectivity with *in vivo* magnetic resonance imaging (MRI), have been explored. These approaches used various brain imaging measures

The cortical surface reconstruction used the publicly available FreeSurfer software package, version 5.3.0 (<http://surfer.nmr.mgh.harvard.edu/>). The details of the processing techniques have been described elsewhere (Dale et al. 1999; Fischl et al. 1999; Fischl and Dale 2000). Briefly, the following stages were included: Talairach transformation, intensity inhomogeneity correction, removal of nonbrain tissues, intensity normalization, tissue segmentation, automated correction of topology defects, and surface deformation to form gray matter/white matter (white) and gray matter/cerebrospinal fluid (pial) boundary surface triangulations (Dale et al. 1999; Fischl and Dale 2000). The generated cortical surfaces were then carefully reviewed and manually edited for technical accuracy. Vertex-wise estimates of the surface area were calculated by assigning one-third of the area of each triangle to each of its vertices. We used 2819-iteration nearest-neighbor averaging to smooth the vertex-wise maps to ensure an accurate estimation of genetic correlations, as previously investigated by Chen and colleagues (Chen et al. 2012).

Definition of Seed Regions

Our genetically based parcellation scheme (Fig. 1) was investigated using representative cortical regions with evolutionary and functional diversity: the superior medial frontal cortex

(SMFC), frontal pole (FP), inferior frontal gyrus (IFG), and primary motor cortex (M1). The seed SMFC was extracted from the AAL template and then projected to a cortical surface model. We used the same seed FP as our previous parcellation that used a connectivity-based technique; see Liu et al. (2013) for the details. In brief, the FP was manually edited on the basis of the FP extracted from the Harvard–Oxford cortical structure atlas. The volumetric FP was projected to a cortical surface to obtain a surface-based region. The IFG in the present study represented traditionally language-related regions, comprising the lateral triangular and opercular areas (pars opercularis and pars triangularis; approximately corresponding to Brodmann's areas (BA) 44 and 45 [Nishitani et al. 2005]). This seed region was extracted from the FreeSurfer Desikan surface-based atlas (Desikan et al. 2006). The M1 seed region was obtained from the PALS Brodmann area atlas (Van Essen 2005), a cortical surface parcellation atlas included in the FreeSurfer software package.

Twin Analysis

In twin analyses, the variance of a phenotype can be accounted for by additive genetic influences (A), common environmental influences (C), or unique environmental influences, including measurement errors (E). A univariate model is able to estimate

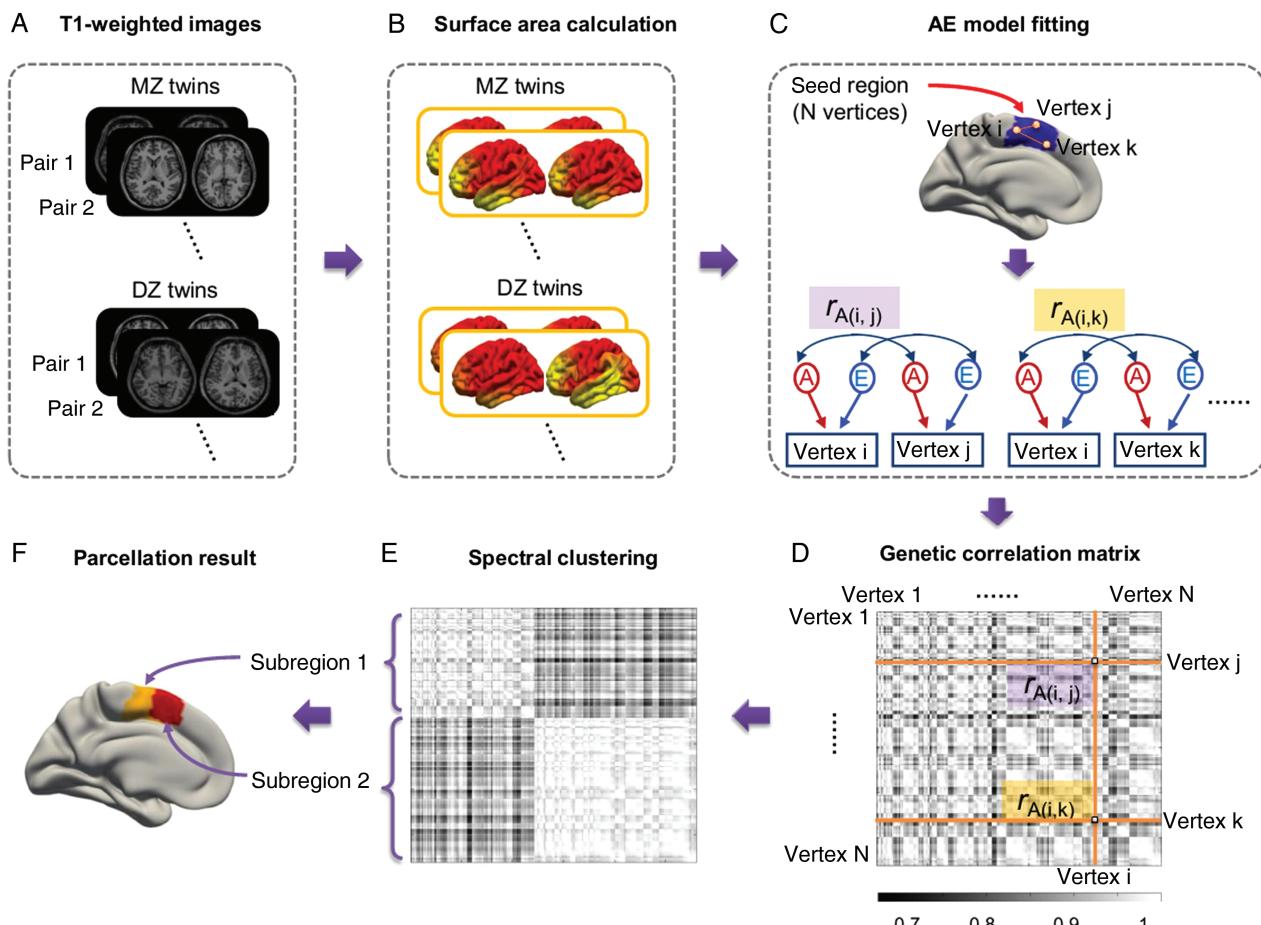


Figure 1. Genetic correlation-based parcellation pipeline. We used T1-weighted imaging data from human twins (A) to explore the detailed division of the seed regions. The cortical surface area of each vertex on the cortical surface was calculated (B). Each hemisphere was down-sampled from 163 842 to 10 242 vertices. The left SMFC, which was used as an example of a seed region (C), consisted of 280 vertices. Subsequently, bivariate AE models (C) were used to estimate the pairwise genetic correlations (r_A) between the surface area of the vertices within the seed regions. This stage generated 1 square genetic correlation matrix (D) using all the twins data, including those for both monozygotic (MZ) and dizygotic (DZ) twins. Spectral clustering (E) was applied to the genetic correlation matrix, generating the parcellation results for the seed region (F).

A, C, and E influences on the variance of the cortical surface area. However, based on previous findings, the cortical surface area shows little evidence of influences from C (Chen et al. 2011; Fyler et al. 2011). Therefore, we used a twin model that only contained A and E. The univariate AE model can be extended to a bivariate model, which can explain the sources of genetic and environmental covariance. Specifically, in addition to examining the genetic and environmental influences on the surface area of each vertex in the seed region, the bivariate correlated-factors model allows for estimates of the genetic (r_A) and environmental (r_E) correlations between the surface area of every pair of vertices (Neale and Cardon 1992). Here, r_A represents the extent of overlap of the genetic factors influencing the surface area, and r_E indicates the extent of the overlap of the environmental factors as well as measurement errors. The analyses were performed using the OpenMx package (version 1.3), a free structural equation modeling software integrated into the R environment (Boker et al. 2011). Before the model fitting, the vertex-wise surface area data were adjusted for age, sex, and global effects (i.e., the vertex-wise surface area at each vertex was divided by the total surface area).

Genetic Correlation-Based Parcellation

A genetic correlation map was generated by pairwise correlations between the vertices within the seed regions. These were first down-sampled from the original 163 842 to 10 242 vertices per hemisphere in order to reduce the computation time for genetic estimation. Given the level of smoothing of the cortical surface area, this down-sampling does not involve any significant loss of information. After obtaining a genetic correlation map, which consisted of the r_A between the surface area measurements for each pair of vertices, a spectral clustering algorithm was used for automatic clustering (Ng et al. 2001) of the left and right hemispheres, separately. Spectral clustering is an unsupervised machine learning algorithm that groups vertices that share similar genetic correlations. The optimal number of clusters was evaluated by the Calinski–Harabasz (CH) index (Caliński and Harabasz 1974), which calculates the ratio of the between-cluster (A) and the within-cluster variance (B): $CH = (A/B) \times (N - k)/(k - 1)$, where N is the number of observations and k is the number of clusters. Generally, the optimal cluster number was chosen as the one that maximizes the CH index, that is, when A is large and B is small. Here, we synthesized the CH indices of the discovery and replication datasets and previous phenotypically based parcellation numbers to determine the optimal genetically based number of clusters for each seed region.

Comparisons of Whole Hemisphere Genetic Correlations of Subregions

To explore the specific genetic correlation profiles of the various subregions, comparisons of the genetic correlations between the subregions and the whole hemisphere were performed. Specifically, for each vertex in a subregion, we calculated its genetic correlations to all the vertices in the entire hemisphere, yielding a genetic correlation matrix $M_{v \times h}$, where v is the number of vertices in a given subregion and h is number of vertices in the corresponding hemisphere. We averaged each column of the matrix and obtained the subregion's mean genetic correlation to every vertex in the hemisphere. Two-sample t-tests between every 2 subregions' genetic correlations to all the vertices in the hemisphere were then performed to obtain the contrast maps. Random field theory was then used for a multiple comparisons

correction (Hayasaka et al. 2004). The statistical analysis was performed using the SurfStat package (<http://www.math.mcgill.ca/keith/surfstat/>).

Relationships between the Similarities in Genetic Correlation Profiles and Brain Connectivity Patterns

We measured the consistency between 2 kinds of pairwise similarities among the vertices within the SMFC. The 2 kinds of similarities reflected genetic correlation profiles and functional connectivity patterns, both of which were calculated based on the correlations between the SMFC and the entire hemisphere. Diffusion tensor imaging was not carried out in the discovery dataset, which limits our measure of anatomical connectivity. However, functional connectivity with resting-state fMRI is also an important aspect of brain connectivity, which reflects brain regions sharing functional profiles. The processing pipeline is shown in Fig. 7A. For the genetic correlation analysis, the number of surface area measurements was down-sampled to 2562 vertices per hemisphere (as targets), of which 70 were in the left SMFC (as seeds). A genetic correlation matrix was generated by calculating the correlations between each seed and target ($G_{70 \times 2562}$). A genetic similarity matrix (dimensions: 70×70) was then obtained by calculating the similarity between each pair of seeds, that is, $W_g = G \times G^T$. In order to maintain the same numbers of seeds and targets for both the genetic correlation and the brain connectivity calculation, seed and target masks were generated from the projection of the surface-based vertices to the volume-based voxels using FreeSurfer. The preprocessing of the resting-state fMRI data was carried out using the DPARSF toolbox (<http://www.restfmri.net/forum/DPARSF>). The preprocessing steps included: (1) discarding the first 10 volumes of each functional time series to allow for magnetization equilibrium, (2) correcting the slice timing for the remaining 230 images and realigning them to the first volume for head motion correction (2 subjects were excluded either because the maximum displacement in the cardinal direction was >3 mm or because the maximum spin was $>3^\circ$), (3) spatial normalizing of all the volumes to the MNI EPI template, (4) spatial smoothing with a Gaussian kernel of 6 mm full-width at half maximum, (5) removing linear trends, (6) temporal band pass filtering (0.01–0.1 Hz), and (7) regressing out nuisance signals, such as those from the white matter and cerebrospinal fluid, as well as global signals and 6 motion parameters. Pearson's correlations were performed between each seed and the target, generating a functional connectivity matrix ($F_{70 \times 2562}$). The similarity between each pair of seeds was then calculated as $W_f = F \times F^T$, resulting in a 70×70 functional cross-correlation matrix for each subject. We randomly selected 1 twin from each of the 101 pairs of twins with minimal head motion and averaged their functional similarity matrices (in total 101 matrices) to obtain an entry-wise mean functional similarity matrix. Finally, the Mantel test was used to quantify the relationship between the average functional similarity matrix and the genetic correlation similarity matrix. The Mantel test was conducted using the vegan package (Oksanen et al. 2015) in R.

Results

Parcellation of the SMFC and Subregional Genetic Correlation Comparison

As shown in Supplementary Figure 1, a cluster number of 2 was optimal for the SMFC. In agreement with existing structural-and functional-based parcellations, we found that the genetic

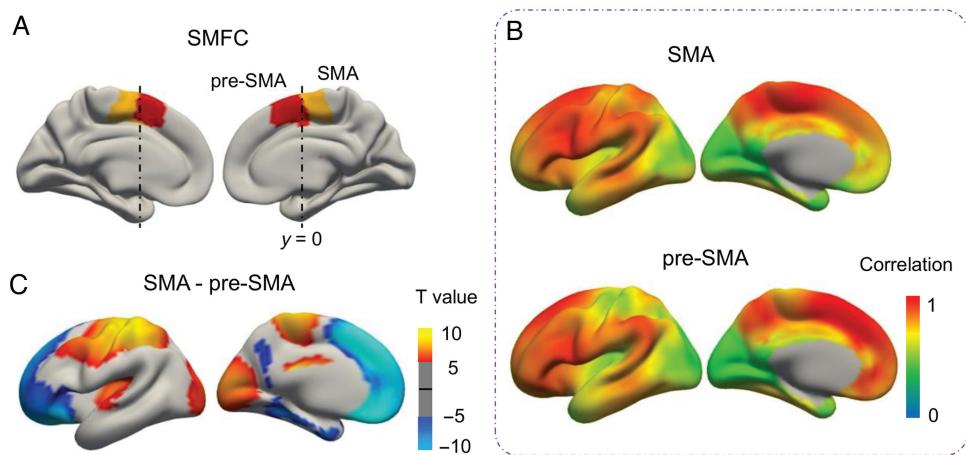


Figure 2. Genetic correlation-based parcellation of the SMFC using the discovery dataset. The SMFC was able to be reproducibly subdivided into the supplementary motor area (SMA, yellow) and pre-SMA (red) subregions, as shown in the medial views of the left and right SMFC (A). (B) Lateral and medial views of the genetic correlations between the left SMFC subregions and the entire left hemisphere. The color bar indicates the genetic correlation strength. (C) Lateral and medial views of the contrast maps between 2 subregions for the left hemisphere obtained using two-sample t-tests. Random field theory was used for multiple comparisons correction with a threshold of $P < 0.05$. The color bar indicates T-values.

clustering of the SMFC showed anterior and posterior clusters that correspond to the presupplementary motor area (pre-SMA) and the SMA (Fig. 2A; for the re-arranged genetic correlation matrices of the seed regions, see *Supplementary Fig. 2*). Specifically, the anterior cluster was located just rostral to the vertical line extending from the anterior commissure ($y \approx 0$), whereas the posterior cluster was located just caudal to the vertical line. The genetic architecture of the SMA and pre-SMA is in line with cytoarchitectonic (Zilles et al. 1996), anatomical (Johansen-Berg et al. 2004), and functional (Kim et al. 2010) connectivity-based parcellations. The parcellation topology was consistent between the replicate datasets (*Supplementary Fig. 3*).

We further investigated the genetic correlation profiles, which represent the genetic correlation distribution of the various subregions across the entire brain hemisphere (Fig. 2B). A two-sample independent t-test revealed significant differences between subregions (Fig. 2C). We found that the genetic correlations for the pre-SMA were significantly stronger with regions in the lateral prefrontal cortex and medial frontal cortex. In contrast, the significant correlations for the SMA were with regions in the motor and visual cortices. Fiber tractography and functional connectivity studies have shown that the pre-SMA has many projections to and from the prefrontal cortex and the anterior cingulate cortex (ACC), whereas the SMA is mainly connected to the nearby M1 (Johansen-Berg et al. 2004; Kim et al. 2010). The subregional genetic correlation profiles were therefore largely consistent with their anatomical and functional connectivity patterns.

Parcellation of the FP and Subregional Genetic Correlation Comparisons

In the case of the FP, the CH index indicated 3 or 4 solutions (*Supplementary Fig. 1*). The optimal number of clusters was previously investigated using a cross-validation technique in a connectivity-based parcellation scheme (Liu et al. 2013), and a cluster number of 3 was found to be the most appropriate value. This cluster number was therefore selected in the present study, which enabling a direct comparison of the genetic correlation- and anatomical connectivity-based divisions. We identified 3

separable subregions, FPO, FPM, and FPI, from the regional maps of the bilateral FP (Fig. 3A) using the genetic correlations within the FP. The left and right FP subregions presented similar patterns, which were consistent with the maximum probability maps provided by a connectivity-based parcellation with diffusion tensor imaging (Liu et al. 2013). The parcellation results were consistent between the replicate datasets (*Supplementary Fig. 3*).

Three subregions showed similar patterns with the highest correlation being with the frontal lobe, the lowest with the occipital lobe, and the median with the temporal and parietal lobes (Fig. 3B). Comparisons of the genetic correlation profiles between subregions (Fig. 3C) showed that the FPO had a stronger correlation with the temporal pole (TP) and the orbitofrontal cortex (OFC) compared with the FPI, as well as an increased trend with the supramarginal gyrus and the OFC compared with the FPM. The FPM showed a significantly increased correlation with the medial prefrontal cortex (MPFC) compared with the FPO. The FPI was found to have a stronger correlation with the dorsal prefrontal cortex (DPFC) compared with either the FPO or the FPM. A previous investigation based on white matter and functional connectivity found that the FP subregions are involved in distinct functional networks (Liu et al. 2013). In brief, the FPO anatomically connects with regions of the social emotion network including the OFC and TP, the FPM connects with areas of the default mode network including the ACC and MPFC, and the FPI connects with regions of the cognitive processing networks, including the DPFC. Therefore, the genetic correlation profiles of the various subregions were consistent with the connectivity patterns of the corresponding subregions.

Parcellation of the IFG and Subregional Genetic Correlation Comparison

As shown in *Supplementary Figure 1*, a cluster number of 5 reached a local peak for the IFG. We also used a parcellation number of 3 in order to test the resemblance to the cytoarchitectonic division. The IFG can be divided into BA 44 and 45 and the frontal operculum (FOP) in a three-cluster solution (Fig. 4A). Again, this finding was

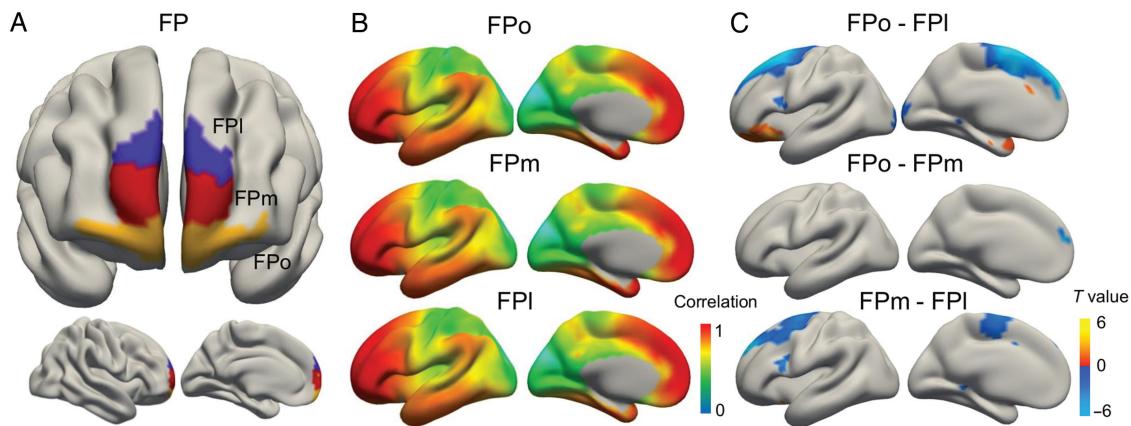


Figure 3. Genetic correlation-based parcellation of the FP using the discovery dataset. The FP was able to be reproducibly subdivided into orbital (FPO, yellow), medial (FPM, red), and lateral (FPI, blue) subregions, as shown in the multiple views of the left and right FP (A). (B) Lateral and medial views of the genetic correlations between the left FP subregions and the entire left hemisphere. The color bar indicates the genetic correlation strength. (C) Lateral and medial views of the contrast maps between 2 subregions for the left hemisphere obtained using two-sample t-tests. Random field theory was used for multiple comparisons correction with a threshold of $P < 0.05$. The color bar indicates T-values.

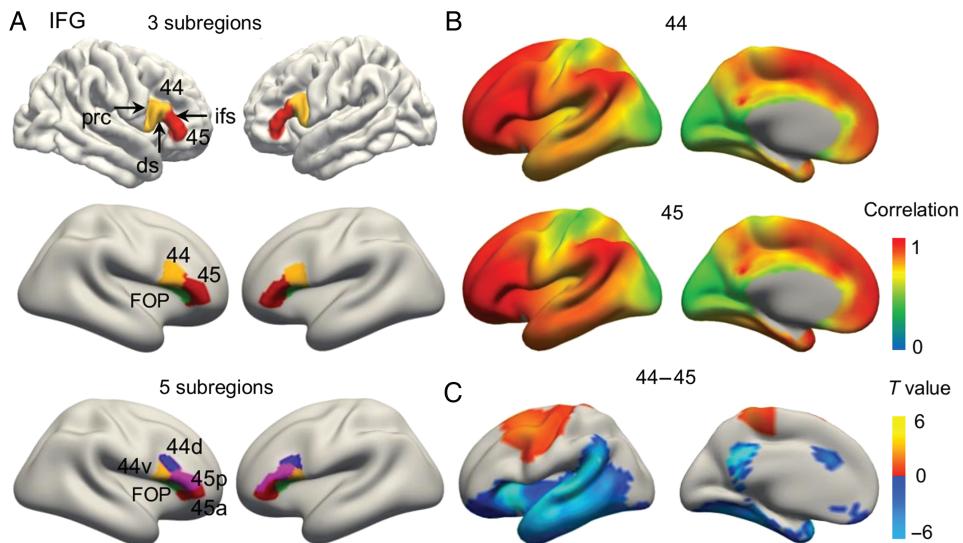


Figure 4. Genetic correlation-based parcellation of the IFG using the discovery dataset. The IFG was able to be reproducibly subdivided into BA44 (yellow), BA45 (red), and the FOP (green) in a three-cluster solution (A, displayed on an uninflated [top row] and inflated [middle row] brain surface, respectively) and into areas 44d, 44v, 45a, 45p, and FOP in a five-cluster solution (A, bottom row), as shown in the lateral views of the left and right IFG. (B) Lateral and medial views of the genetic correlations between the left BA44 and BA45 subregions and the entire left hemisphere. The color bar indicates the genetic correlation strength. (C) Lateral and medial views of the contrast maps between BA44 and BA45 for the left hemisphere obtained using two-sample t-tests. Random field theory was used for multiple comparisons correction with a threshold of $P < 0.05$. The color bar indicates T-values. prc, precentral sulcus; ifs, inferior frontal sulcus; ds, diagonal sulcus.

largely in accordance with classical cytoarchitectonics, with a boundary that aligned with the diagonal sulcus (Nishitani et al. 2005). BA44 was subdivided into dorsal and ventral areas, 44d and 44v, and BA45 was subdivided into anterior and posterior areas, 45a and 45p, in the five-cluster solution (Fig. 4A), which corresponds to the subdivisions identified using transmitter receptor distribution data (Amunts et al. 2010).

BA44 and BA45 constitute the classical Broca's area. Comparisons of the BA44 and BA45 subregions revealed that BA45 exhibited higher correlations with regions in the temporal lobe and prefrontal cortex compared with BA44. In contrast, for BA44, the genetic correlations were stronger with regions in the motor areas (Fig. 4C). The genetic correlation profiles of BA44 and 45 corresponded to the functional roles of language

comprehension and speech production, respectively (Clos et al. 2013).

Parcellation of M1

For M1, a cluster number of 5 was found to be preferable because of the higher regional CH values in both datasets (Supplementary Fig. 1). M1 was able to be subdivided into 5 subregions, 5 of which corresponded to motor representations of body parts: the face, hand and arm, trunk, hip, and leg and foot (from ventrolateral to dorsomedial). The single remaining subregion in the anterior medial part of M1 was the SMA (Fig. 5). This is in line with widely recognized topographic organization. A connectivity-based parcellation of the lateral precentral gyrus resulted in 4 distinct

regions (Schubotz et al. 2010), in accordance with our parcellation of the lateral convexity of M1.

Ruling out Potential Underestimation from the Twin Models

To rule out the possibility that the power to detect genetic correlations might have been limited because of the fairly modest MRI twin samples, we combined the discovery and replication datasets to obtain a total of 204 pairs of healthy young twins. We used the same processing procedure as that used for the discovery dataset except that the site information was regressed out as

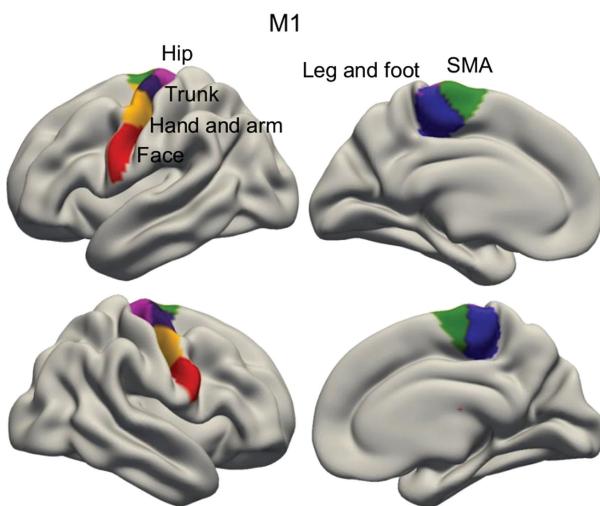


Figure 5. Genetic correlation-based parcellation of the primary motor cortex (M1) using the discovery dataset. M1 was able to be reproducibly subdivided into 6 subregions, 5 of which corresponded to motor representations of the face, hand and arm, trunk, hip, leg, and foot. The remaining subregion was the SMA. All are shown in the lateral and medial views of the left and right M1.

another nuisance variable before the model fitting. The division of the seed regions displayed the same delineations as those obtained using the discovery and replication datasets separately (Fig. 6), demonstrating that our models are sufficient to detect genetic correlations from bivariate twin analyses.

Relationships Between the Similarities in Genetic Correlation Profiles and Brain Connectivity Patterns

To further characterize the quantitative relationships between the similarity in genetic correlation profile and the similarity in brain connectivity patterns, we measured resting state functional connectivity (using resting-state fMRI) in the discovery dataset. We found that the genetic correlation and functional connectivity similarity matrices were significantly correlated ($r = 0.344$, $P < 0.001$, Fig. 7B).

Discussion

To the best of our knowledge, this is the first study to parcellate fine-scale, functionally distinct subregions noninvasively based on intrinsic genetic information obtained by twin analysis. We found that the genetic clustering of the human SMFC, FP, IFG, and M1 was predominantly bilaterally symmetric and reproducible between 2 independent datasets. The resulting clusters closely resembled the subregions identified by other approaches using structural and functional features. Furthermore, we observed subregional genetic correlation profiles across each hemisphere which were generally consistent with previously identified structural and functional connectivity patterns. These results establish the presence of a close relationship between genetically driven and phenotypically driven cortical parcellations and further indicate that cortical functional segregation is primarily genetically determined.

Previous attempts at human brain parcellation have focused on the use of neuroimaging measures of brain structure and function. In the present study, we used a data-driven clustering

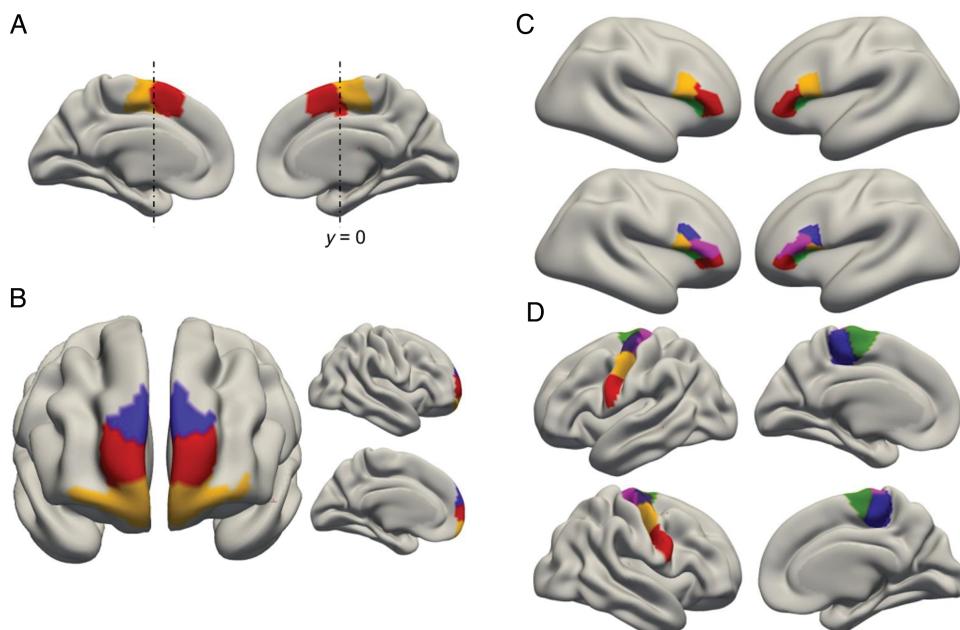


Figure 6. Genetic correlation-based parcellation using a combination of the discovery and replication datasets. The seed regions were the SMFC (A), FP (B), IFG (C), and M1 (D).

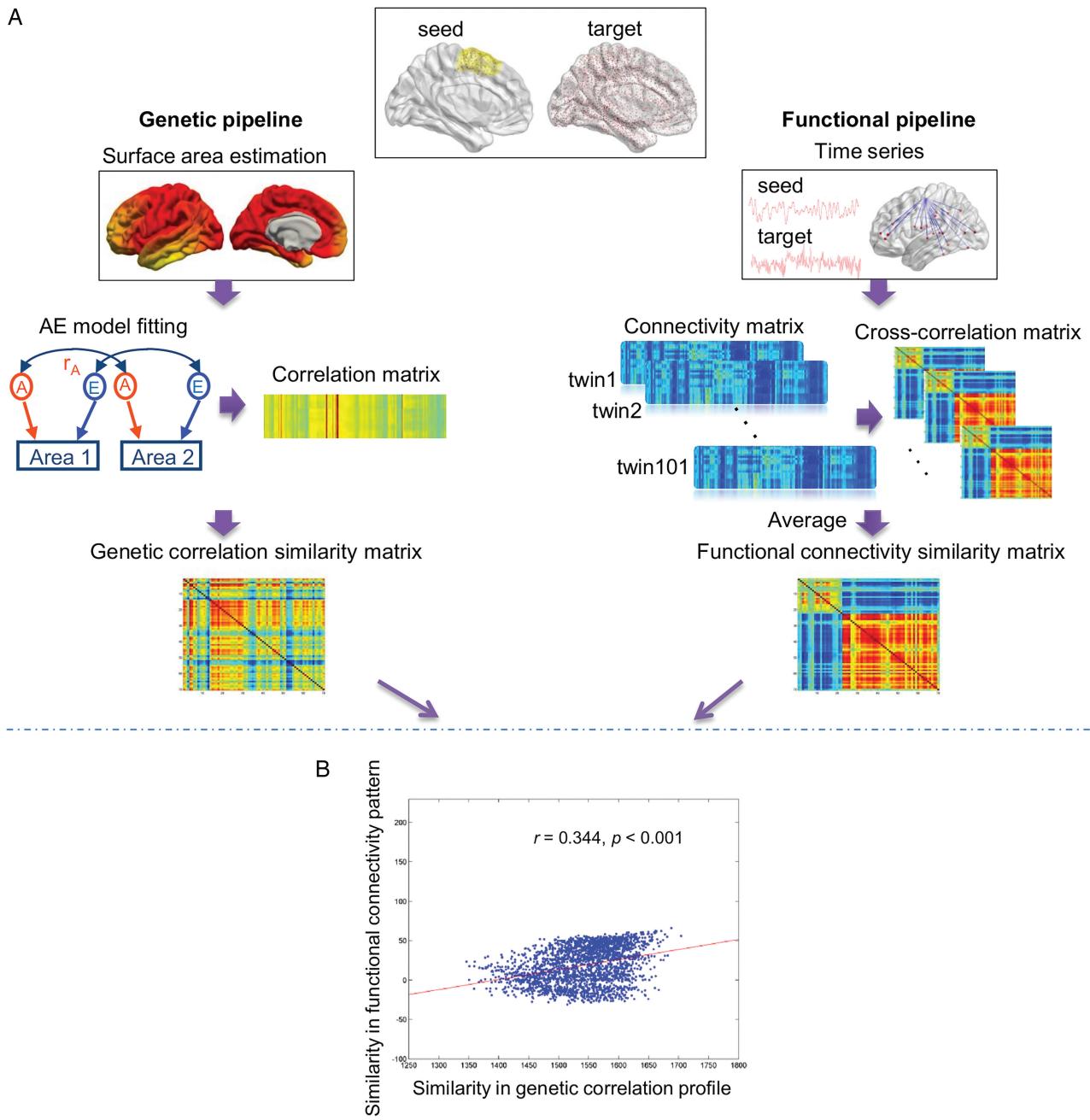


Figure 7. Relationships between the similarities in functional connectivity patterns and in genetic correlation profiles seeded from the left SMFC (B) and the processing pipelines (A).

algorithm that automatically grouped cortical locations that shared homogeneous genetic factors. Our identifications of the genetic borders were largely in agreement with previously defined structural and functional borders, indicating the validity of our parcellation scheme. Genetic factors participate in many aspects of cortical development, including neuron induction, polarization, migration, differentiation, and the intra- and inter-areal connections of nerve cells in the brain (Sanes et al. 2011), all of which are important for establishing and maintaining the regional identity of cells and tissues. For instance, reelin has been identified as a multifunctional protein that controls the positioning of neurons, together with their growth, maturation, and synaptic activity in the developing and adult brain

(Lee et al. 2014). A number of transcription factors, such as Pax6 and Emx2, which are expressed in opposing gradients across the cortical surface, have also been shown to play an important role in the regional identities of cortical areas at the whole brain level (Bishop et al. 2000; O’Leary et al. 2007; Rakic et al. 2009). At the lobar level, evidence from animal studies using mutant mice supports the concept that genetic mechanisms govern the process of subdividing the frontal cortex (Cholfin and Rubenstein 2007, 2008). Specifically, frontal cortex subdivision patterning is regulated by *Fgf8*, *Fgf17*, and *Emx2*, which play distinct roles in molecular regionalization (Cholfin and Rubenstein 2008). Although many genes are likely to be involved in cortical regionalization, a particular set of genes may influence a specific

- ventral subregions with anatomical and functional specializations. *J Neurosci.* 27:10259–10269.
- Van Essen DC. 2005. A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. *NeuroImage.* 28:635–662.
- Van Essen DC, Dierker DL. 2007. Surface-based and probabilistic atlases of primate cerebral cortex. *Neuron.* 56:209–225.
- Van Essen DC, Glasser MF. 2014. In vivo architectonics: a cortico-centric perspective. *NeuroImage.* 93(Pt 2):157–164.
- Van Essen DC, Glasser MF, Dierker DL, Harwell J, Coalson T. 2012. Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cereb Cortex.* 22:2241–2262.
- Wandell BA, Winawer J. 2011. Imaging retinotopic maps in the human brain. *Vision Res.* 51:718–737.
- Zilles K, Schlaug G, Geyer S, Luppino G, Matelli M, Qu M, Schleicher A, Schormann T. 1996. Anatomy and transmitter receptors of the supplementary motor areas in the human and nonhuman primate brain. *Advan Neurol.* 70:29–43.