Genome-wide scan of healthy human connectome discovers *SPON1* gene variant influencing dementia severity

Neda Jahanshad[a], Priya Rajagopalan[a], Xue Hua[a], Derrek P. Hibar[a], Talia M. Nir[a], Arthur W. Toga[a], Clifford R. Jack, Jr.[b], Andrew J. Saykin[c], Robert C. Green[d], Michael W. Weiner[e,f], Sarah E. Medland[g], Grant W. Montgomery[g], Narelle K. Hansell[g], Katie L. McMahon[h], Greig I. de Zubicaray[i], Nicholas G. Martin[g], Margaret J. Wright[g], Paul M. Thompson[a,1], and the Alzheimer's Disease Neuroimaging Initiative[2]

[a]Imaging Genetics Center, Laboratory of Neuro Imaging, University of California at Los Angeles School of Medicine, Los Angeles, CA 90095; [b]Department of Radiology, Mayo Clinic, Rochester, MN 55905; [c]Center for Neuroimaging, Department of Radiology and Imaging Science, Indiana University School of Medicine, Indianapolis, IN 46202; [d]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115; [e]Department of Radiology, Medicine, and Psychiatry, University of California, San Francisco, CA 94121; [f]Magnetic Resonance Spectroscopy Unit, Department of Veterans Affairs Medical Center, San Francisco, CA 94121; [g]Department of Genetics and Computational Biology, Queensland Institute of Medical Research, Brisbane, QLD 4029, Australia; and [h]Center for Advanced Imaging and [i]School of Psychology, University of Queensland, Brisbane, QLD 4072, Australia

Edited by Marcus E. Raichle, Washington University, St. Louis, MO, and approved January 29, 2013 (received for review September 19, 2012)

Aberrant connectivity is implicated in many neurological and psychiatric disorders, including Alzheimer's disease and schizophrenia. However, other than a few disease-associated candidate genes, we know little about the degree to which genetics play a role in the brain networks; we know even less about specific genes that influence brain connections. Twin and family-based studies can generate estimates of overall genetic influences on a trait, but genome-wide association scans (GWASs) can screen the genome for specific variants influencing the brain or risk for disease. To identify the heritability of various brain connections, we scanned healthy young adult twins with high-field, high-angular resolution diffusion MRI. We adapted GWASs to screen the brain's connectivity pattern, allowing us to discover genetic variants that affect the human brain's wiring. The association of connectivity with the *SPON1* variant at rs2618516 on chromosome 11 (11p15.2) reached connectome-wide, genome-wide significance after stringent statistical corrections were enforced, and it was replicated in an independent subsample. rs2618516 was shown to affect brain structure in an elderly population with varying degrees of dementia. Older people who carried the connectivity variant had significantly milder clinical dementia scores and lower risk of Alzheimer's disease. As a posthoc analysis, we conducted GWASs on several organizational and topological network measures derived from the matrices to discover variants in and around genes associated with autism (*MACROD2*), development (*NEDD4*), and mental retardation (*UBE2A*) significantly associated with connectivity. Connectome-wide, genome-wide screening offers substantial promise to discover genes affecting brain connectivity and risk for brain diseases.

diffusion tensor imaging | neuroimaging genetics | graph theory | HARDI tractography | multiple comparisons correction

Human brain anatomy involves a complex network of structural and functional pathways that connect anatomically distinct regions. These pathways can be visualized, on a gross anatomical scale, with diffusion imaging-based tractography (1). Neural pathways change as our brains develop (2, 3) and are altered in neurodegenerative conditions, including Alzheimer's disease (4). Our individual genetic makeup exerts a strong influence on the functional synchronization of brain regions (5) and the patterning of cortical structure (6, 7), but particular genes that impact the brain's neural connectivity are still largely unknown.

To empower the search for genes that affect the brain's connectivity patterns, we first studied a large twin and family cohort consisting of 366 young adults (ages 20–29 y) from 223 families. We used anatomical brain MRI combined with high-angular resolution diffusion imaging (HARDI) at high magnetic field (4 T) to subdivide cortical regions into areas of known structure and function (8), while also mapping the white matter fiber pathways between them with high-resolution tractography. We defined connectivity maps based on the proportion of the total number of fibers traced that intersect or interconnect cortical regions within and across the brain hemispheres. Such maps of the brain's structural connectome may be represented as graphs or matrices. Here, we use symmetric matrices, in which each matrix element (x,y) shows the proportion of fibers connecting regions x and y of the brain; these connectomes may be compared statistically across subjects to identify consistent patterns or factors that affect them. A flowchart of the image processing steps needed to compute brain connectivity is shown in Fig. S1.

Proportions of the observed variance in any brain trait—such as a network measure—may be ascribed to genes vs. environment by fitting structural equation models to data derived from different kinds of twins: monozygotic (MZ) twins share all their genes, whereas dizygotic (DZ) twins share, on average, one-half. To assess the degree of the genetic contribution to the brain's white matter connectivity pattern, we first fitted a classical structural equation model to connectivity matrices from 46 pairs of MZ and 64 pairs of DZ twins. We included same-sex and mixed-sex twins in our analysis while controlling for sex, age, and intracranial volume. All families were Caucasian and of European ancestry, but families were all unrelated to each other.

Results and Discussion

Fig. 1 shows the proportion of the population variance attributable to additive genetic factors (A); the rest of the variance is attributable to unique individual factors and measurement error (E). We restricted our subsequent genetic analysis to regions

Author contributions: N.J. and P.M.T. designed research; N.J. and P.M.T. performed research; A.W.T., C.R.J., A.J.S., R.C.G., M.W.W., S.E.M., G.W.M., N.K.H., K.L.M., G.I.d.Z., N.G.M., M.J.W., and the A.D.N.I. contributed new reagents/analytic tools; N.J., P.R., X.H., D.P.H., T.M.N., and P.M.T. analyzed data; and N.J. and P.M.T. wrote the paper.

Conflict of interest statement: ADNI is partially funded by public and private agencies, but the authors have no financial interests related to the content of this paper.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: thompson@loni.ucla.edu.

[2]A complete list of investigators from the Alzheimer's Disease Neuroimaging Initiative (ADNI) may be found in the *SI Text*. ADNI investigators contributed to the design and implementation of ADNI and/or provided data, but many of them did not participate in the analysis or writing of this report. For a complete listing of ADNI investigators, please see http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1216206110/-/DCSupplemental.

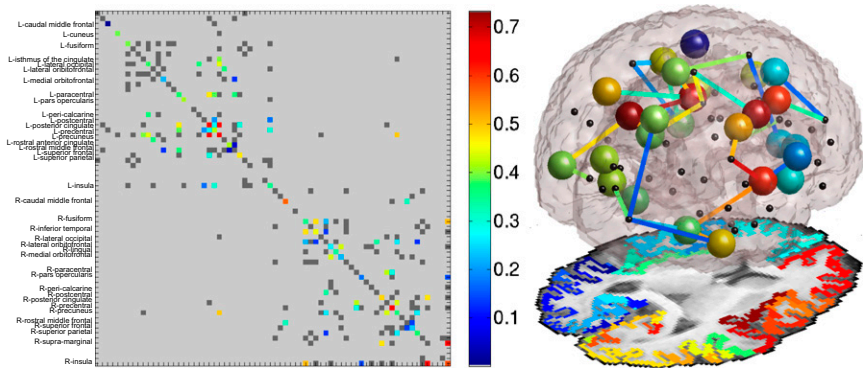

Fig. 1. Heritable brain connections. Quantitative genetic analysis of brain connectivity matrices in 46 MZ and 64 DZ twin pairs based on the A/E model of genetic influence. This model breaks down the observed variance in neural connectivity into components attributable to additive genetic (A) vs. unique environmental effects (E). For all nodes where the A/E model was fitted to the data, the additive genetic component is shown. Regions are displayed only if the additive genetic term exceeded 1% and the model was fitted to the data (χ^2 goodness-of-fit test, $P > 0.05$, denoted a good fit). Adding a shared environmental term did not significantly improve the fit of this model of the factors affecting brain connectivity. Regions listed on the matrix *x* and *y* axes are listed in *SI Text*. The color bar represents the proportion of the variance attributable to additive genetic factors. Warmer colors (e.g., red) represent more heritable connections—where differences in the proportion of fibers depend on genetic differences among individuals.

where this combined A/E model—the best fitting and least complex model that contained a genetic component—was shown to have a good fit and the overall genetic influence on wiring exceeded 1%. Generally, only phenotypes of moderate to high heritability are used for genome-wide scanning, but in our connectome-wide scan, we analyzed all connections with any detectable degree of underlying genetic influence (>1%). For each of 59 heritable regions, the additive genetic component of the variance estimated from the model is listed in Table S1 along with its 95% confidence interval (95% CI).

After heritability is established, the search for specific genetic variants influencing brain connectivity becomes viable. The sample of families was divided into a discovery cohort of 169 individuals and a replication cohort of another 163 individuals (detailed in the *SI Text*); all family members were assigned to the same cohort. The full statistical pipeline for genetic analysis, showing the number of subjects in each analysis, is presented in Fig. S2.

To investigate genetic contributors to brain connectivity, we conducted a genome-wide association scan (GWAS) to detect any association between commonly carried genetic variants and each of 59 connectivity elements considered. An extremely stringent significance threshold of $P < 8.96 \times 10^{-9}$ was enforced. This threshold was determined by assessing the total number of independent connectome-wide, genome-wide tests performed; this threshold value was validated through permutations—which generate datasets with no effect—and used to control the rate of false positives reported. This threshold implies genome-wide significance across all cortical nodes and nodal connections (*Materials and Methods*).

Our GWAS of the connectivity patterns in the discovery cohort showed a genome-wide significant association ($P = 3.23 \times 10^{-9}$) within the *SPON1* gene at rs2618516 (Fig. 2). The contributions of this variant were then assessed in the replication cohort at the same node in the brain's connectivity network. The replication of this association was highly significant ($P = 0.0021$) with increased density (unstandardized regression slope, $\beta_{Group1} = 0.0022$, $\beta_{Group2} = 0.0015$) or a 0.2% increase with respect to the minor T allele for the brain fiber density connecting the left posterior cingulate and left superior parietal lobe.

SPON1 encodes the developmentally regulated protein F-spondin, which is induced in neuronal injury and impairs the binding of cells to the ECM (9). In rats, this protein induces a hippocampal progenitor cell line (and primary cortical neural cells) to differentiate into cells with neuronal features (10). Intriguingly, F-spondin also modulates amyloid-β precursor protein (APP) cleavage by binding to the initial α/β-cleavage site of APP (11). APP has recently been found to bind to cholesterol (12), which makes up much of the myelin composition of white matter fibers. Additionally, many proteins and genes involved in APP processing, including F-spondin (*SPON1*), also interact with the family of receptors for apolipoprotein E, which is coded for by the gene *APOE*—a robust Alzheimer's disease (AD) genetic risk factor (13, 14). Recently, overexpression of F-spondin has even been shown to lead to

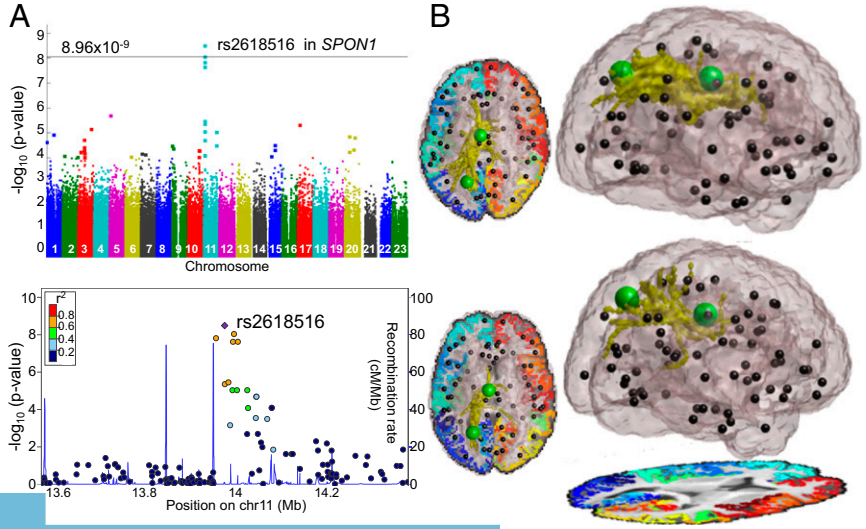

Fig. 2. A significant genome-wide association was found between a common genetic variant on chromosome 11 and anatomical fiber connectivity. This effect was identified first in a discovery subsample and reproduced in a separate replication subsample. (*A*) A Manhattan plot shows the results of this analysis in the discovery cohort; the highest points denote SNPs or common variants on the genome, where genetic variation is associated with fiber connectivity in the brain. Below it, we zoom in on the significant locus using LocusZoom (https://statgen.sph.umich.edu/locuszoom/). (*B*) An association was found between this common genetic variant and the degree of connectivity between the left posterior cingulate gyrus and the left superior parietal lobe. *Upper* shows the raw fiber densities of paths connecting these two regions (centered on the green spheres) in a 26-y-old female with two copies of the minor T allele at rs2618516. *Lower* shows the same connection in an age- and sex-matched noncarrier of the genetic variant. Fiber densities of networks connecting these two regions are displayed in yellow.

cognitive improvements and reduced amyloid-β-plaque deposition in mice (15).

SPON1, a member of the spondin family of genes, is highly expressed in the embryonic neural floor plate, and it is essential for neural growth and cell adhesion (16, 17). Prior studies implicate this same gene in brain structural development and neural connectivity, and its expression declines with aging (18). The significant association of *SPON1* with the fiber density for the connection between the superior parietal cortex and the posterior cingulate is of particular interest. This region is affected early and consistently by AD pathology (19), a neurodegenerative disorder associated with disruptions in neural network connectivity (20, 21) and myelin breakdown (22). AD is partially caused by amyloid-β peptides and τ-proteins that accumulate in the brain and disrupt neural networks (23, 24). The posterior cingulate and superior parietal cortex show early disturbances (25), including volumetric atrophy (26, 27), impaired glucose metabolism (28), altered activity during task-based functional imaging studies (29), and disrupted resting-state functional connectivity (21)—all of these effects have been reported in AD patients or those individuals at heightened risk for AD with mild cognitive impairment. Some AD-associated genes, such as *CLU*, are known to influence the brain's fiber integrity decades before the typical age of onset for symptoms of dementia (28). *SPON1* may affect cortical connectivity in healthy young adults, which may, in turn, influence the risk for and progression of neurodegenerative diseases later in life.

We conducted follow-up tests to assess any association between this *SPON1* variant and brain structure of 738 elderly (mean age = 75.5 y ± 6.8 SD) Caucasian individuals from the Alzheimer's Disease Neuroimaging Initiative (ADNI) who had been evaluated and diagnosed as having AD, having mild cognitive impairment (MCI), or being cognitively normal. Using tensor-based morphometry as described previously (30), we tested whether elderly carriers of the gene variant had different profiles of regional brain volumes.

After controlling for age and sex, the genetic variation at rs2618516 was significantly associated with differences in volume for several regions of the brain (Fig. 3). The maps show the corrected P value after performing searchlight false discovery rate (FDR; $q = 0.05$) to show only the volumetric effect in regions of significance while controlling the false-positive rate at 5%. The minimum P value for association was $P = 2.68 \times 10^{-5}$. Ventricular regions surrounding the temporal lobes and hippocampus were significantly reduced in size, and the gray matter around the posterior cingulate cortex was enlarged with each additional copy of the minor T allele at rs2618516. We also estimated the association between this genetic variant and a widely used measure of clinical severity in dementia, the clinical dementia rating (CDR) (31), and we found a significant association in the same 738 ADNI individuals ($P = 0.0169$). This result remained significant even after controlling for *APOE* genotype ($P = 0.026$). Carriers of the *SPON1* allele showed a decrease in the sum-of-boxes CDR score ($\beta = -0.231$, $\beta_{ApoE4\_control} = -0.208$), suggesting the *SPON1* variant may exert a protective influence on dementia severity regardless of the *APOE* risk genotype. To determine the effect of the gene as a whole rather than the single allele, a gene-based test using versatile gene-based association study (VEGAS; 32) using the top 10% of variants in *SPON1* as associated to CDR showed *SPON1* to associate with CDR in 738 ADNI individuals ($P = 0.01$).

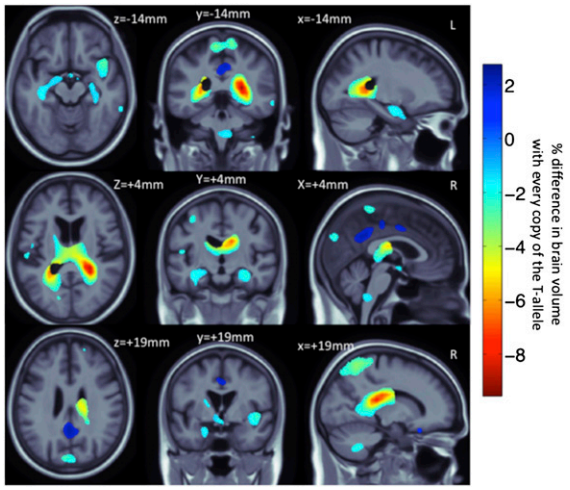

Fig. 3. Connectivity-related gene also affects regional brain volumes. In all 738 Caucasian subjects from the ADNI study, the *SPON1* variant was significantly associated with bilateral differences in regional brain volumes. Ventricular regions, particularly surrounding the temporal lobes, were significantly reduced in size, whereas the gray matter around the posterior cingulate cortex was enlarged with each additional copy of the minor T allele at rs2618516.

The stronger white matter fiber connection associated with the *SPON1* variant in healthy young adults may dampen the severity of dementia in carriers of the protective variant over half a century later. CDR is often used as a clinical diagnostic measure; in a posthoc analysis, we assessed if the discovered *SPON1* variant was associated with disease status in the ADNI cohort. When regressing the additive effect of the allele on the presence or absence of AD in all 738 individuals (AD vs. MCI and controls aggregated together), there was a nominally significant additive (dose-dependent) association to the T allele ($P = 0.0494$, $n = 738$); only comparing AD with controls, the allele was more strongly associated with disease status ($P = 0.0305$, $n = 379$).

To maximize our power for association, for exploratory purposes, we combined our discovery and replication cohorts to perform a GWAS at all nodes as before. With 331 genotyped subjects with similar ancestry (from a total of 366 with matrices computed), the statistical power for genetic association was increased. A stronger association for multiple *SPON1* variants in linkage disequilibrium ($r^2 = 0.6$–0.8) with the original SNP, rs2618516, was additionally found in the full-sample GWAS for the same connection, including rs2697846, rs10832160, rs11023052, and rs7124311. The Manhattan Plot of association statistics across the genome is shown in Fig. S3. Results of this more highly powered association test, including P values and direction of associations for significant SNPs, are presented in Table S2. In the combined analysis, the P value for our initially discovered SNP, rs2618516, was 5.8×10^{-10}. Additionally, a full-scale quantile–quantile plot is shown in Fig. S4, combining all P values from all tested connections, with points along the observed y axis ranging to 9.23.

Because this *SPON1* locus has genome-wide significant effects at the connection between the left posterior cingulate and the superior parietal cortices, its effect on the entire brain network is of interest. To assess the significance of the top *SPON1* variant on structural brain connectivity as a whole, we regressed the additive effect of the replicated variant on each of the tested matrix elements, including all 331 genotyped subject connectivity matrices. This analysis revealed significant associations at multiple cortical connections when correcting for multiple comparisons using the FDR procedure (33). Association with the variant was found with the total number of fibers that cross the left superior parietal cortex altogether ($P = 5.2 \times 10^{-4}$) at the FDR ($q = 0.05$) critical threshold. The single region and single connection associated with this variant suggest that it may have a localized but substantial function in the brain network's organization, although additional brain mapping and investigations of gene expression are required. Interestingly, a prior study of structural connectivity found that the posterior cingulate and the superior parietal cortices, among only a few other regions, play a pivotal role in global network traffic. They were hubs in the rich club of the network (34)—the set of most highly interconnected nodes—despite the fact that over 1,000 nodes were considered. These regions also have some of the highest degrees of centrality in the core of the structural network. They were also found to be among the most heritable nodes locally (35). Our analysis of the

entire heritable structural brain network revealed significant genetic findings in the previously determined rich club of the network (36), and therefore, future genetic studies of the connectome may be able to boost power by focusing on the rich club connections. Rich club elements may be especially promising targets for genetic analysis.

Twin studies of functional brain networks also find that local and global organizational measures of connectivity are remarkably heritable. Smit et al. (37) used EEG-based measures of connectivity to study network clustering and path length and evaluated the two together using the small-worldness metric. Measures of network efficiency (clustering and path length) were found to be heritable, but small worldness was not. Fornito et al. (35) examined local and global measures of efficiency and connection distance along with overall density for binarized resting-state networks. At different thresholds, the degree of heritability varied for different global measures, and heritability was not uniform across all nodes; in fact, they had a wide range of heritability. In a similar investigation of functional connectivity in children, however, van den Heuvel et al. (38) found no heritability on local measures of clustering, but robustly found that the global normalized path length (λ) was heritable. Here, we examined our weighted structural networks from HARDI and performed GWAS on the corresponding network measures previously examined in functional network studies. These measures included local strength (corresponding to density here), local clustering coefficient, local efficiency, global efficiency, λ, γ (normalized global clustering coefficient), and small worldness. To broaden our analysis, we also included another measure of centrality related to brain networks—Eigenvector centrality (39)—to describe a node's importance in a network according to other important nodes. We also assessed global modularity to describe the degree of network segregation.

Table S3 lists the 15 most significant SNPs (with the lowest P values) found for all network measures. Here, local measures of network strength yielded genome-wide significant findings at the strict significance threshold set (described in *Materials and Methods*). These SNPs include rs16997087 near the *MACROD2* gene ($P = 1.11 \times 10^{-10}$), rs17819300 and rs17819282 in linkage disequilibrium (LD) inside the *NEDD4* gene ($P = 1.36 \times 10^{-10}$ and $P = 2.78 \times 10^{-10}$), and rs7879933 in the *UCE2A* gene ($P = 1.83 \times 10^{-10}$) of the X chromosome. All three genes are already strong candidate genes for affecting brain structure and relating to disease. *MACROD2* is expressed in the brain, and variants associate with disorders such as autism (40). *NEDD4*, in the ubiquitin ligase family, is pivotal in neuronal interactions (41) and may be important in Parkinson disease (42). *UBE2A*, another ubiquitin-related gene, may be important for learning and memory (43), and mutations in the gene are associated with mental retardation (44). Local measures of clustering coefficient and efficiency provided suggestive associations (defined here as $5 \times 10^{-8} < P < 3.39 \times 10^{-10}$), where 5×10^{-8} is the classical threshold for a single GWAS with this density of SNPs and 3.39×10^{-10} is the genome-wide significance threshold corrected for the number of tests. We found no suggestive genome-wide significant variants for Eigenvector centrality or any of the global measures, because the lowest P value in all cases was $>5 \times 10^{-8}$.

Although this exploratory analysis requires replication, the results show considerable promise for basic and localized measures of weighted network organization. Global network measures were quite highly heritable in some prior functional imaging studies, but they may not be quite as promising for genetic discovery in sparse structural networks as when using measures of local organization. Alternatively, genetic influences may not be evenly distributed across the connection matrix as seen above, and therefore, heritable regions and connections may be more promising for genetic discovery. Local measures offer the additional advantage of helping to identify specific brain pathways involved with respect to specific genes.

Neuroimaging genetics has mainly focused on mapping structures and regions of the brain associated with candidate genetic markers of disease. More recently, truly vast genome-wide screening of neuroimaging measures (45–48) has uncovered new genetic markers influencing brain structure. This genome-wide discovery and replication of *SPON1* variants associated with brain connectivity in several hundred individuals suggest a new neurogenetic pathway with links to dementia severity. Our posthoc GWASs of topological connectivity measures also suggest that network metrics may be good endophenotypes for genetic discovery, although replication and validation are still required. The power for genetic discovery is greatly increased with added samples, and therefore, future analyses of diffusion-based connectome measures—across many sites worldwide—may implicate many more common and perhaps, also rare genetic determinants influencing brain wiring and dementia severity.

Materials and Methods

Subject Demographics and Image Acquisition. Subjects were recruited as part of an ongoing research project examining healthy young adult Australian twins using structural and functional MRI and HARDI with a projected sample size of ~1,150 at completion (49). Subjects included 92 young adult MZ twins (46 pairs) and 128 DZ twins (64 pairs) along with 146 nontwin siblings and unpaired twins from a total of 223 families. In total, images from 366 right-handed young adults (mean age = 23.5 y, SD = 2.0) were included in this study. Study participants gave informed consent; the institutional ethics committee at the Queensland Institute of Medical Research, University of Queensland, The Wesley Hospital, and the University of California at Los Angeles School of Medicine approved the study. Genomic DNA samples were analyzed on the Human610-Quad BeadChip (Illumina). Additional details on subject exclusion, establishing zygosity, and genotyping can be found in *SI Text*.

Anatomical and HARDI whole-brain MRI scans were acquired from 366 subjects with a high-magnetic field (4 T) Bruker Medspec MRI scanner. Acquisition parameters are detailed in the *SI Text*; briefly, diffusion parameters were optimized to improve the signal to noise ratio (SNR) (56) and included 94 diffusion weighted and 11 non-diffusion weighted scans.

Cortical Extraction and HARDI Tractography. The flowchart in Fig. S1 shows image processing steps to generate a map of brain fiber connectivity based on an individual's anatomical MRI and high-angular resolution diffusion imaging data. To summarize, diffusion-weighted MRI scans are coregistered to a standard anatomical T1-weighted brain image through an image called the average b_0 image. The structural scans undergo automated cortical parcellations, and tractography is performed on the diffusion-weighted MRIs. Cortical labels are uniformly dilated to ensure that they intersect the white matter, where tracts are traced. Tracts are elastically fitted to the labeled structural scan to ensure adequate coregistration. Quality control is performed to ensure accurate alignment (Fig. S5). Finally, connectivity matrices are created—each matrix element shows the proportion of the total number of detected fibers in the brain that crosses or intersects the specific pair of cortical regions at the top and side of the matrix. For each subject, a 70×70 connectivity matrix was created. Each element described the proportion of the total number of fibers connecting each of the regions; diagonal elements of the matrix describe the total number of fibers passing through a certain cortical region of interest. If more than 5% of subjects had no fibers in a matrix element, then that connection was considered invalid or insufficiently consistent in its occurrence in the population; therefore, it was not included in the analysis. Detailed methods can be found in *SI Text*.

Structural Equation Modeling and Heritability Analysis. A covariance matrix S_g was fitted for every measurement of interest (each element in the connection matrix) within the pairs of twins for each of the two types of twins (identical or fraternal). A structural equation model may be fitted to compare the observed and expected covariances to infer the proportion of the variance caused by A, shared environmental (C), and E components of variance (51). Additional details on A/C/E modeling can be found in *SI Text*; 59 connections were found to be sufficiently heritable ($a^2 > 0.01$). The additive genetic component of the variance estimated from the model is listed in Table S2 along with its 95% CI. No CI directly included zero, although some lower bounds were very close (on the order of 10^{-6} or less).

Genome-Wide Associations Across Matrices. Genome-wide associations were performed at each of 59 valid matrix elements using a mixed model approach controlling for age, sex, and intracranial volume (emmaX: http://genetics.cs.ucla.edu/emmax/news.html) (52) to account for the familial relatedness

between subjects through the use of a kinship matrix describing the approximate degree of genetic similarity between subjects. A zero in the kinship matrix represents the relation between unrelated individuals, MZ twins' kinship is denoted by one (with identical genomes), and DZ twins and nontwin siblings within the same family are denoted by 0.5 (because they share approximately one-half). Analysis was limited to those SNPs with a minor allele frequency greater than 0.1. Our sample size used for GWAS is small compared with most genetic analyses, and therefore, associations with SNPs having a low minor allele frequency may show artificially high or low associations driven by few subjects in the homozygous minor allele group. Avoiding these SNPs in our analysis would reduce the chance for such biases. A total of 428,287 SNPs were tested.

Establishing Significance Thresholds. A significance threshold of 8.96×10^{-9} was established for genome-wide significance (i.e., a more stringent threshold than most GWAS studies because of the additional dimension of search across the connectome). We determined significance levels for association tests by first evaluating the total number of independent tests performed. LD among SNPs tested leads to statistical correlation between the values of 428,287 SNPs in the cohort, and therefore, if 2 genotyped SNPs are in high LD, each test is not completely independent of the other. By first estimating the effective number of independent tests, we can avoid using a significance level too conservative for the number of tests performed, while correctly controlling the chance of false-positive associations. Because of this correlation between genotyped SNPs, the effective number (M_{eff}) of SNPs tested as defined in refs. 53 and 54 was 214,578. The same logic may also be applied to the matrix elements tested. Clearly, an off-diagonal element is not independent of the two diagonal elements corresponding to each one of its (x,y) components (i.e., the matrix element representing the total proportion of fibers connecting cortical regions x and y), and $C(x,y)$ is not fully independent of matrix elements $C(x,x)$ and $C(y,y)$ corresponding to the total proportion of tracts crossing each cortical region x and y, respectively. A total of 59 connections was considered (those connections with an a^2 component > 0.01 and an AE model with a good fit). A principal components analysis of 59 matrix elements, using information from the twins in the A/C/E model, reveals that 22 components are sufficient to explain 95% of the variance in the population; 26 of 59 total connections are on the diagonal corresponding to different regions on the cortex rather than connections between cortices. Although several of these 26 regions may be correlated or anticorrelated to some degree, 26 regions serve as the maximal possible number of completely independent regions available, and all other connections are between two of these nodes.

A Bonferroni correction on the effective number of independent samples would be $0.05/(26 \times 214{,}578) = 8.96 \times 10^{-9}$ or as determined by principal components analysis, $0.05/(22 \times 214{,}578) = 1.06 \times 10^{-8}$, respectively. As we set out to establish a method to identify relevant genetic loci, we were less concerned with reducing the number of false-negative associations but more interested in ensuring true-positive findings and reducing the type I error; therefore, we choose the more conservative threshold of $P = 8.96 \times 10^{-9}$ as our threshold for genome-wide significance, which we show is acceptable below through extensive permutations to verify the null distribution under various degrees of heritability.

Other genome-wide association studies of multiple traits have used the FDR procedure over all traits to find the appropriate correction threshold for one analysis across the genome (55). For comparison, we performed a similar analysis of our results using the FDR procedure on the P values obtained from 59 traits (a total of $59 \times 428{,}287 = 25{,}268{,}933$ tests) to obtain a correction threshold of 7.58×10^{-9} for the full-cohort GWAS ($n = 331$), again similar to our estimated $P = 8.96 \times 10^{-9}$ threshold. Note that no true association P values fell in the range from 7.58×10^{-9} (estimated from FDR) to 1.06×10^{-8} (estimated from principal components analysis), such that the upper bound of our FDR threshold is limited to 7.58×10^{-9}. Therefore, the FDR-estimated correction threshold of 7.58×10^{-9} was the highest truly obtained P value below our chosen 8.96×10^{-9}. Therefore, in multiple ways, our threshold of 8.96×10^{-9} is valid. Even with a stringent classical Bonferroni correction over all tests ($0.05/59 \times 428{,}287 = 1.98 \times 10^{-9}$), two SNPs total would have still survived and be declared significant. We additionally ensure that our results are not caused by volumetric differences (*SI Text*).

Modeling Null Distributions for GWAS. At each valid node in the genetic network, a GWAS was performed for 218 of 220 twins used in the A/C/E structural equation model; one pair was removed, because the twins were considered ancestry outliers (*SI Text*). To determine any potential differences in the null distributions with respect to the degree of the additive genetic component, GWAS was performed 1,000 times on permuted matrices. Matrix elements were permuted one by one across individuals, and therefore, each phenotype was permuted independently of others without any disruption of family structure. When conducting the permutations, each subject's covariates (age, sex, and intracranial volume) remained true to their source, whereas the matrix elements were permuted in a manner that ensured preservation of family structure. Values for MZ twin pairs were only permuted with other MZ twin pairs, whereas the DZ twin pairs were permuted separately. Within each permutation, within-twin pair rearrangements were also allowed to maximize the total number of permutation possibilities. Across all 1,000 permutations of all 59 connections, we note an average false-positive rate of 0.008 (summing all P values falling below the 8.96×10^{-9} genome-wide significance threshold) per connection or $0.008 \times 59 = 0.472$ false positives connectome-wide, which takes into account all correlations. Five significant SNPs were found in our full GWAS in the *SPON1* gene alone, and one was found in our discovery half sample. For specificity, in 1,000 permutations of the node where *SPON1* was found to be significant, a total of six P values fell below the threshold (i.e., the procedure has a false-positive rate of 0.6%). In Fig. S6, we show the null distribution of GWAS statistics at a subset of seven valid connections at various levels of heritability, increasing from 1% to 63%.

Tensor-Based Morphometry Study of the ADNI Cohort. A well-characterized cohort of elderly subjects with neuroimaging and genetic data from ADNI was analyzed to confirm effects of *SPON1* on brain structure. The ADNI is a large 5-y study launched in 2004 by public and private institutions, and it is described in detail elsewhere (50). Overview information summarizing the ADNI and previously described analyses can be found in *SI Text*. All ADNI data are publicly available (http://www.loni.ucla.edu/ADNI/). Our analysis included 738 individuals (average age ± SD = 75.52 ± 6.78 y; 438 men/300 women) including 178 AD, 354 MCI, and 206 healthy participants. Effect sizes for individual genetic variants are expected to be small; hence the analysis was not split into diagnostic groups since analyzing the full phenotypic continuum is thought to yield greater power to detect genetic associations (56).

To investigate effects of our replicated *SPON1* variant on regional brain volumes in ADNI, univariate linear regression was used to associate the number of minor alleles (zero, one, or two) with the Jacobian values (describing the amount of brain tissue deficit or excess relative to the standard template) at each voxel in the brain after covarying for age and sex. Computing thousands of tests of associations on a voxelwise level can introduce a high type I (false positive) error rate in neuroimaging studies. To control these errors, we used a searchlight method for FDR correction as described in ref. 57, which ensures a regional control over the FDR in any reported findings. We implemented this searchlight method to correct the associations between the neuroimaging phenotype (morphometry) and the rs2618516 genotype. The maps shown in Fig. 3 are thresholded at the appropriate corrected P value after performing searchlight FDR ($q = 0.05$) to show only the volumetric effect in regions of significance.

ADNI—Assessment of Clinical Dementia Rating. Six clinical scores are assessed to determine CDR: Personal Care Score, Community Affairs Score, Home and Hobbies Score, Judgment and Problem Solving Score, Memory Score, and Orientation Score (58). Each score is rated from 0 to 0.5 to 1 to 2 to 3 in order of increasing risk. The average score is used to help distinguish between healthy controls, people with MCI, and AD patients. The sum of all scores, also known as the sum-of-boxes CDR, was used to determine an association with the rs2618516 locus using a linear regression, as described above, controlling for age and sex.

Posthoc GWAS of Network Measures. Although the initial aim of our investigation was to focus on the genetic association of specific edges and connection densities in the connectivity matrix and be able to effectively map these genetic pathways, prior studies of functional brain networks found high levels of heritability for organizational and topological measures of the brain's network, including measures of efficiency and small worldness (35, 37, 38). To understand the genetics involved in topological measures of brain connectivity in a hierarchical manner, we explored network measures at individual nodes, including strength, local efficiency, local clustering coefficients, and eigenvalue centrality; on a global topological scale, we examined modularity, global efficiency, normalized mean clustering coefficient (γ), normalized characteristic path length (λ), and small worldness, where the normalized measures (and hence, small worldness) were compared with 50 random networks. Network measures were calculated using the Brain Connectivity Toolbox (59) at https://sites.google.com/site/bctnet/.

GWASs were performed for all four local measures on all 70 nodes as well as all five global metrics. Testing these 285 phenotypes with 428,287 SNPs leads to

a Bonferroni corrected significance threshold of $0.05/(285 \times 428{,}287) = 4.096 \times 10^{-10}$ [using the effective number of SNPs as described above, the correction threshold would be $0.05/(285 \times 214{,}578) = 8.176 \times 10^{-10}$]; however, because this analysis was posthoc and exploratory, to be especially stringent with correction, we do not take phenotypic or genetic correlation into account, and we also include the previous 59 connections tested as phenotypes for GWAS and correct for the exploratory analysis at $P = 0.05/(344 \times 428{,}287) = 3.394 \times 10^{-10}$.

ACKNOWLEDGMENTS. This study was supported by National Institute of Child Health and Human Development Grant R01 HD050735 and National Health and Medical Research Council (NHMRC), Australia Grant 486682. Genotyping was supported by NHMRC Grant 389875. N.J. was supported in part by National Institutes of Health (NIH) Grant T15 LM07356. D.P.H. is partially supported by National Science Foundation Graduate Research Fellowship Program Grant DGE-0707424. G.I.d.Z. is supported by Australian Research Council Future Fellowship FT0991634. R.C.G. is supported by NIH Grants AG027841, HG002213, and HG006500. Additional support for algorithm development was provided by NIH Grants R01 NS080655, R01 MH097268, R01 AG040060, R01 EB008432, R01 EB008281, R01 EB007813, and P41 EB015922 (to P.M.T.). The Alzheimer's Disease Neuroimaging Initiative (ADNI) was supported by public and private funding sources, listed in the *SI Text*.

1. Wedeen VJ, et al. (2012) The geometric structure of the brain fiber pathways. *Science* 335(6076):1628–1634.
2. Dosenbach NU, et al. (2010) Prediction of individual brain maturity using fMRI. *Science* 329(5997):1358–1361.
3. Kochunov P, et al. (2012) Fractional anisotropy of water diffusion in cerebral white matter across the lifespan. *Neurobiol Aging* 33(1):9–20.
4. Buckner RL, et al. (2009) Cortical hubs revealed by intrinsic functional connectivity: Mapping, assessment of stability, and relation to Alzheimer's disease. *J Neurosci* 29(6): 1860–1873.
5. Glahn DC, et al. (2010) Genetic control over the resting brain. *Proc Natl Acad Sci USA* 107(3):1223–1228.
6. Thompson PM, et al. (2001) Genetic influences on brain structure. *Nat Neurosci* 4(12): 1253–1258.
7. Chen CH, et al. (2012) Hierarchical genetic organization of human cortical surface area. *Science* 335(6076):1634–1636.
8. Desikan RS, et al. (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31(3): 968–980.
9. Tzarfati-Majar V, Burstyn-Cohen T, Klar A (2001) F-spondin is a contact-repellent molecule for embryonic motor neurons. *Proc Natl Acad Sci USA* 98(8):4722–4727.
10. Schubert D, Klar A, Park M, Dargusch R, Fischer WH (2006) F-spondin promotes nerve precursor differentiation. *J Neurochem* 96(2):444–453.
11. Ho A, Südhof TC (2004) Binding of F-spondin to amyloid-beta precursor protein: A candidate amyloid-beta precursor protein ligand that modulates amyloid-beta precursor protein cleavage. *Proc Natl Acad Sci USA* 101(8):2548–2553.
12. Barrett PJ, et al. (2012) The amyloid precursor protein has a flexible transmembrane domain and binds cholesterol. *Science* 336(6085):1168–1171.
13. Hoe HS, Rebeck GW (2008) Functional interactions of APP with the apoE receptor family. *J Neurochem* 106(6):2263–2271.
14. Perreau VM, et al. (2010) A domain level interaction network of amyloid precursor protein and Abeta of Alzheimer's disease. *Proteomics* 10(12):2377–2395.
15. Hafez DM, et al. (2012) F-spondin gene transfer improves memory performance and reduces amyloid-β levels in mice. *Neuroscience* 223:465–472.
16. Klar A, Baldassare M, Jessell TM (1992) F-spondin: A gene expressed at high levels in the floor plate encodes a secreted protein that promotes neural cell adhesion and neurite extension. *Cell* 69(1):95–110.
17. Burstyn-Cohen T, et al. (1999) F-Spondin is required for accurate pathfinding of commissural axons at the floor plate. *Neuron* 23(2):233–246.
18. Lu T, et al. (2004) Gene regulation and DNA damage in the ageing human brain. *Nature* 429(6994):883–891.
19. Jacobs HI, Van Boxtel MP, Jolles J, Verhey FR, Uylings HB (2012) Parietal cortex matters in Alzheimer's disease: An overview of structural, functional and metabolic findings. *Neurosci Biobehav Rev* 36(1):297–309.
20. Villain N, et al. (2010) Sequential relationships between grey matter and white matter atrophy and brain metabolic abnormalities in early Alzheimer's disease. *Brain* 133(11): 3301–3314.
21. Sorg C, et al. (2007) Selective changes of resting-state networks in individuals at risk for Alzheimer's disease. *Proc Natl Acad Sci USA* 104(47):18760–18765.
22. Bartzokis G (2004) Age-related myelin breakdown: A developmental model of cognitive decline and Alzheimer's disease. *Neurobiol Aging* 25(1):5–18.
23. Palop JJ, Mucke L (2010) Amyloid-beta-induced neuronal dysfunction in Alzheimer's disease: From synapses toward neural networks. *Nat Neurosci* 13(7):812–818.
24. Braak H, Braak E (1991) Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* 82(4):239–259.
25. Thompson PM, et al. (2001) Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cereb Cortex* 11(1):1–16.
26. Buckner RL, et al. (2005) Molecular, structural, and functional characterization of Alzheimer's disease: Evidence for a relationship between default activity, amyloid, and memory. *J Neurosci* 25(34):7709–7717.
27. Thompson PM, et al. (2003) Dynamics of gray matter loss in Alzheimer's disease. *J Neurosci* 23(3):994–1005.
28. Braskie MN, et al. (2011) Common Alzheimer's disease risk variant within the *CLU* gene affects white matter microstructure in young adults. *J Neurosci* 31(18):6764–6770.
29. Celone KA, et al. (2006) Alterations in memory networks in mild cognitive impairment and Alzheimer's disease: An independent component analysis. *J Neurosci* 26(40): 10222–10231.
30. Hua X, et al. (2008) Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: An MRI study of 676 AD, MCI, and normal subjects. *Neuroimage* 43(3):458–469.
31. Morris JC, et al. (1997) Clinical dementia rating training and reliability in multicenter studies: The Alzheimer's Disease Cooperative Study experience. *Neurology* 48(6): 1508–1510.
32. Liu JZ, et al. (2010) A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 87(1):139–145.
33. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57(1): 289–300.
34. van den Heuvel MP, Kahn RS, Goñi J, Sporns O (2012) High-cost, high-capacity backbone for global brain communication. *Proc Natl Acad Sci USA* 109(28):11372–11377.
35. Fornito A, et al. (2011) Genetic influences on cost-efficient organization of human cortical functional networks. *J Neurosci* 31(9):3261–3270.
36. van den Heuvel MP, Sporns O (2011) Rich-club organization of the human connectome. *J Neurosci* 31(44):15775–15786.
37. Smit DJ, et al. (2010) Endophenotypes in a dynamically connected brain. *Behav Genet* 40(2):167–177.
38. van den Heuvel MP, et al. (2012) Genetic control of functional brain network efficiency in children. *Eur Neuropsychopharmacol*, 10.1016/j.euroneuro.2012.06.007.
39. Hardmeier M, et al. (2012) Cognitive dysfunction in early multiple sclerosis: Altered centrality derived from resting-state functional connectivity using magneto-encephalography. *PLoS One* 7(7):e42087.
40. Anney R, et al. (2010) A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 19(20):4072–4082.
41. Donovan P, Poronnik P (2012) Nedd4 and Nedd4-2: Ubiquitin ligases at work in the neuron. *Int J Biochem Cell Biol* 45(3):706–710.
42. Muñoz-Soriano V, Nieto-Arellano R, Paricio N (2012) Septin 4, the Drosophila ortholog of human CDCrel-1, accumulates in parkin mutant brains and is functionally related to the Nedd4 E3 ubiquitin ligase. *J Mol Neurosci* 48(1):136–143.
43. Park CC, et al. (2011) Gene networks associated with conditional fear in mice identified using a systems genetics approach. *BMC Syst Biol* 5:43.
44. Budny B, et al. (2010) Novel missense mutations in the ubiquitination-related gene UBE2A cause a recognizable X-linked mental retardation syndrome. *Clin Genet* 77(6): 541–551.
45. Ikram MA, et al. (2012) Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat Genet* 44(5):539–544.
46. Taal HR, et al. (2012) Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nat Genet* 44(5):532–538.
47. Stein JL, et al. (2012) Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* 44(5):552–561.
48. Bis JC, et al. (2012) Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nat Genet* 44(5):545–551.
49. de Zubicaray GI, et al. (2008) Meeting the challenges of neuroimaging genetics. *Brain Imaging Behav* 2(4):258–263.
50. Mueller SG, et al. (2005) The Alzheimer's Disease Neuroimaging Initiative. *Neuroimaging Clin N Am* 15(4):869–877.
51. Rijsdijk FV, Sham PC (2002) Analytic approaches to twin data using structural equation models. *Brief Bioinform* 3(2):119–133.
52. Kang HM, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42(4):348–354.
53. Gao X, Starmer J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol* 32(4):361–369.
54. Gao X, Becker LC, Becker DM, Starmer JD, Province MA (2010) Avoiding the high Bonferroni penalty in genome-wide association studies. *Genet Epidemiol* 34(1):100–105.
55. Sabatti C, et al. (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet* 41(1):35–46.
56. Stein JL, et al. (2010) Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. *Neuroimage* 51(2):542–554.
57. Langers DR, Jansen JF, Backes WH (2007) Enhanced signal detection in neuroimaging by means of regional control of the global false discovery rate. *Neuroimage* 38(1): 43–56.
58. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL (1982) A new clinical scale for the staging of dementia. *Br J Psychiatry* 140:566–572.
59. Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* 52(3):1059–1069.