Behavioral/Systems/Cognitive

Gene Network Effects on Brain Microstructure and Intellectual Performance Identified in 472 Twins

Ming-Chang Chiang,[1,2] **Marina Barysheva,**[2] **Katie L. McMahon,**[3] **Greig I. de Zubicaray,**[4] **Kori Johnson,**[3] **Grant W. Montgomery,**[5] **Nicholas G. Martin,**[5] **Arthur W. Toga,**[2] **Margaret J. Wright,**[5] **Paul Shapshak,**[6] **and Paul M. Thompson**[2]

[1]Department of Biomedical Engineering, National Yang-Ming University, Taipei, 112, Taiwan, [2]Laboratory of Neuro Imaging, Department of Neurology, University of California, Los Angeles School of Medicine, Los Angeles, California 90095-7334, [3]Centre for Advanced Imaging and [4]School of Psychology, University of Queensland, Brisbane, St. Lucia, Queensland 407, Australia, [5]Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia, and [6]Division of Infectious Disease and International Medicine, Department of Medicine and Department of Psychiatry and Behavioral Medicine, University of South Florida Morsani College of Medicine, Tampa General Hospital, Tampa, Florida 33606

A major challenge in neuroscience is finding which genes affect brain integrity, connectivity, and intellectual function. Discovering influential genes holds vast promise for neuroscience, but typical genome-wide searches assess approximately one million genetic variants one-by-one, leading to intractable false positive rates, even with vast samples of subjects. Even more intractable is the question of which genes interact and how they work together to affect brain connectivity. Here, we report a novel approach that discovers which genes contribute to brain wiring and fiber integrity at all pairs of points in a brain scan. We studied genetic correlations between thousands of points in human brain images from 472 twins and their nontwin siblings (mean age: 23.7 ± 2.1 SD years; 193 male/279 female). We combined clustering with genome-wide scanning to find brain systems with common genetic determination. We then filtered the image in a new way to boost power to find causal genes. Using network analysis, we found a network of genes that affect brain wiring in healthy young adults. Our new strategy makes it computationally more tractable to discover genes that affect brain integrity. The gene network showed small-world and scale-free topologies, suggesting efficiency in genetic interactions and resilience to network disruption. Genetic variants at hubs of the network influence intellectual performance by modulating associations between performance intelligence quotient and the integrity of major white matter tracts, such as the callosal genu and splenium, cingulum, optic radiations, and the superior longitudinal fasciculus.

Introduction

A major goal in neuroscience is to identify specific genes that affect brain structure and function, as well as genes that modulate risk for neuropsychiatric disease (Gandhi and Wood, 2010). Using genome-wide association (GWA) scans, scientists can now link genetic markers across the whole genome with biological measures in human populations, to search for common genetic variants that associate with higher risk for disease, e.g., abnormalities in the *ApoE* gene (Schuff et al., 2009) or the *CLU* gene (Braskie et al., 2011) associated with late-onset Alzheimer's disease. GWA can help to identify contributors to the mechanisms of neuropsychiatric disease, which can be pursued to identify molecular targets for treatment. Even so, standard GWA studies ignore interactions between genes (Manolio et al., 2009; Becker et al., 2011); the power of GWA is also rather low as one needs to correct for approximately a million hypotheses tested across the genome, so vast samples of subject need to be assessed.

Recently, the cDNA microarray method has been widely used to identify gene coexpression networks where the component genes are "clustered" into groups that jointly affect complex traits (Ghazalpour et al., 2006). The use of clustering may reveal evidence of collective effects of networks of genes on brain pathology. For example, some gene sets associated with neuronal integrity may show altered expression patterns in schizophrenia patients (Torkamani et al., 2010). Nevertheless, *in situ* DNA sampling and analysis with microarrays may only be feasible in the *postmortem* brain; moreover, the spatial scope and resolution in tissue sampling across the brain are limited.

In this study, we developed a novel approach to gene discovery in the living brain, outlined in Figure 1. We examined networks

Figure 1. Flowchart of the genetic clustering strategy to boost power to discover genetic variants influencing the brain. Our imaging-genetic analysis method was designed to reduce the complexity of the search for significant associations between the genotype of every SNP across the genome and the imaging phenotype (white matter integrity here) at every location across the brain. We first reduced the dimensionality of the image data by selecting ROIs where white matter integrity was under strong genetic control, using structural equation modeling and hierarchical clustering. We then searched along the genome to detect every SNP that was significantly associated with FA in these ROIs. To account for possible interactions between these significant SNPs, we connected them to form a network, based on correlations between their effects on white matter integrity. We then assessed (1) their topological configurations, specifically the small-world and scale-free properties of the network and (2) the functional roles of the hub SNPs in the network, i.e., how these hub SNPs interacted to affect white matter integrity and intellectual performance. See Materials and Methods for details of each individual step in the analysis.

Table 1. Demographic data and family composition of the participants whose image and genome-wide SNP data were both available

Data	
Age, years	23.7 ± 2.1[a]
Sex (M/F)	193/279[a]
FIQ	113.9 ± 12.6[b]
PIQ	114.0 ± 15.9[b]
VIQ	111.8 ± 10.9[b]
Number of families	
MZ pairs	77
MZ pair plus one nontwin sibling	8
DZ pairs	87
Trizygotic triplets	2
DZ pair plus one nontwin sibling	12
DZ pair plus two nontwin siblings	1
Single participants	37
Two nontwin siblings	17
Three nontwin siblings	1

[a]Number of participants, $N = 472$; [b]$N = 456$. Age and IQ data are displayed as mean $\pm$ SD. M, Male; F, female.

of genes that influence brain integrity measured by diffusion tensor images (DTI). DTI is a variant of standard MRI, measuring the 3D directionality of water diffusion at each location in the brain; fractional anisotropy (FA), derived from DTI, is a widely accepted measure of fiber tract coherence and integrity (Basser et al., 1994). By adapting classical twin methods and hierarchical clustering to DTI scans from 472 twins and their nontwin siblings, we identified patterns of brain regions where white matter integrity, measured by FA, was under strong and relatively homogeneous genetic control. We then performed GWA to identify single nucleotide polymorphisms (SNPs) associated with FA in these regions. We discovered SNP pairs that tended to jointly affect brain regions. By applying network topology analysis to this SNP association matrix, we isolated an SNP network influencing white matter integrity in the living brain. This gene network showed small-world and scale-free properties, with several dom-

inant hubs and structured subnetworks. These configurations may make the network biologically economical and more resistant to disruption than random networks (Albert et al., 2000; Rubinov and Sporns, 2010). Hub SNPs in the network interacted with each other in their effects on white matter integrity. They may also influence intellectual performance by modulating the associations between white matter integrity and the performance intelligence quotient (IQ).

Materials and Methods

Participants. Five-hundred and thirty-one healthy adult subjects (age: 23.7 $\pm$ 2.1 years, mean $\pm$SD; age range: 20–29 years; sex: 217 male/314 female)— consisting of pairs of twins and their nontwin siblings—were recruited from 271 different families. Twin subjects were recruited to study genetic effects that underlie different traits, by comparing the similarity of monozygotic (MZ) and dizygotic (DZ) twins/ siblings. All subjects received high-resolution brain MRI and neurocognitive evaluations as part of a 5-year research project evaluating healthy Australian twins. The projected sample size for the adult study is 1150 at completion (de Zubicaray et al., 2008). As described previously (Chiang et al., 2009b), zygosity was established objectively by typing nine independent DNA microsatellite polymorphisms (polymorphism information content >0.7), using standard PCR methods and genotyping. Results were cross-checked with blood group (ABO, MNS, and Rh), and phenotypic data (hair, skin, and eye color), giving an estimated overall probability of correct zygosity assignment >99.99%. Zygosity was subsequently confirmed by GWA. All subjects were screened to exclude left-handers and cases of pathology known to affect brain structure, a history of significant head injury, a neurological or psychiatric illness, substance abuse or dependence, or a psychiatric disorder in any first-degree relative. The study was formally approved by the Institutional Review Boards of the Queensland Institute for Medical Research and the University of California, Los Angeles. Written informed consent was obtained from all participants before any experimental procedures were performed.

SNP genotyping. Genotyping for subjects in this study is part of a larger genotyping project that includes >4400 twins and their siblings, designed to identify melanoma risk factors (Wright and Martin, 2004; Medland et al., 2009). DNA was isolated from blood cells using standard protocols. Genomic DNA samples were analyzed on the Human610-Quad BeadChip (Illumina). Of single nucleotide polymorphisms (SNPs), 592,392 were genotyped. SNPs that did not satisfy the following quality control criteria were excluded: genotype call rate <95%, significant deviation from Hardy–Weinberg equilibrium $p < 10^{-6}$, minor allele frequency <0.01, or only one allele was observed, and a platform-specific recommended quality control score (BeadStudio GenCall score) <0.7. After quality control, 529,497 SNPs remained. Genotype data were collected for 484 of the 531 subjects who received brain MRI scanning. We excluded 12 subjects who were identified to be ancestry outliers (Medland et al., 2009) from the subsequent genome-wide association study (GWAS), so there were 472 subjects with available image and genome-wide SNP data (age: 23.7 $\pm$ 2.1 years, mean $\pm$ SD; age range: 20–29 years; sex: 193 M/279 F). The participants' demographic information and the twin/sibling family composition are summarized in Table 1.

Evaluation of psychometric intelligence. General intellectual ability was assessed at age 16 [as in the study by Luciano et al. (2003)] using the Multidimensional Aptitude Battery (MAB) (Jackson, 1984), a measure highly correlated with the Wechsler Adult Intelligence Scale. MAB is designed for assessment of adults and adolescents aged 16 and older. In
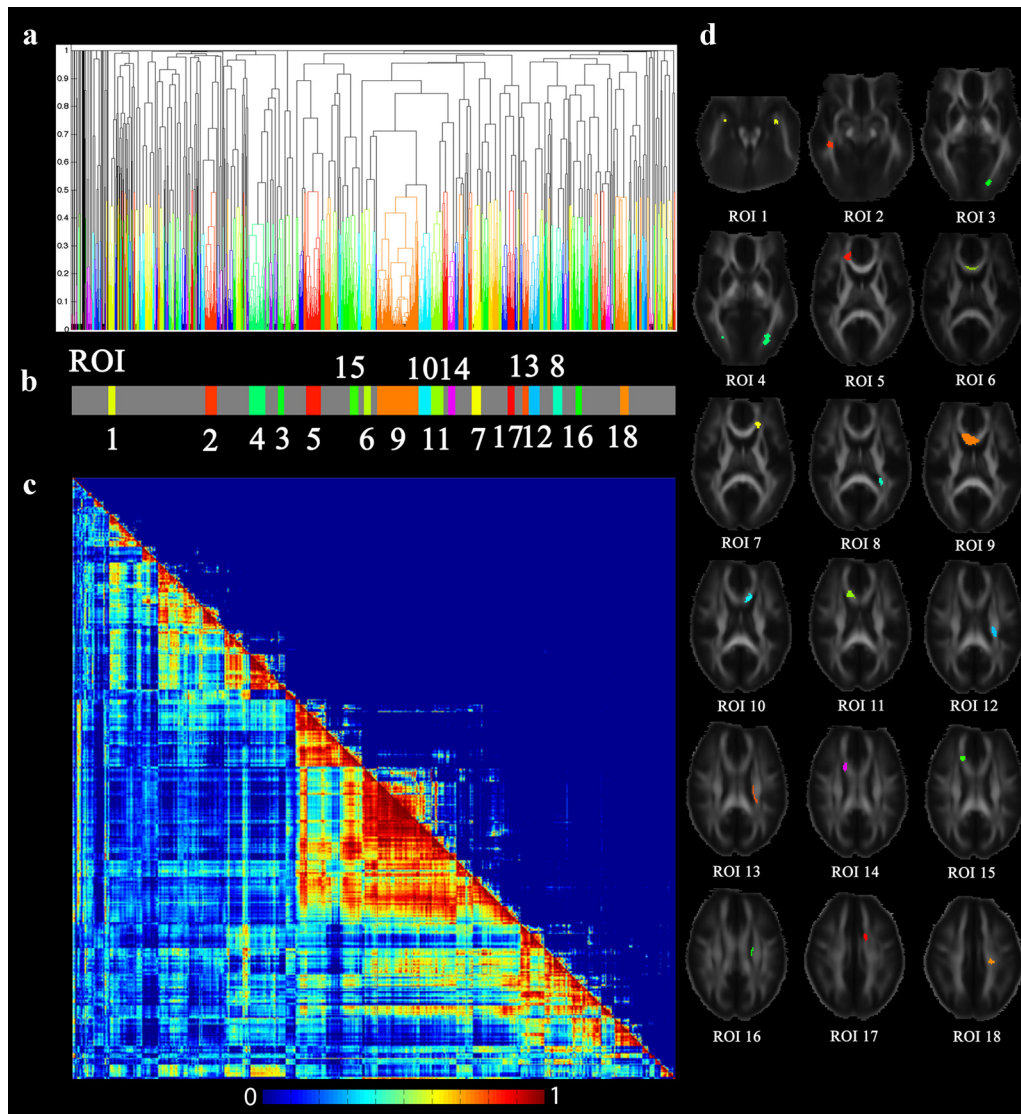
Figure 2. ROI selection using hierarchical clustering. ***a***, The dendrogram shows how the hierarchical clustering method aggregates the 4876 candidate voxels into clusters based on the TO index between each pair of voxels. Each voxel was represented by the most terminal branch, i.e., leaf, of the dendrogram. Voxels for which the pairwise distance $d \leq 0.5$ were grouped to form a cluster, where $d = 1 - TO$ index (the topological overlap index), and were demonstrated as leaves of the same color in the dendrogram. ***b***, Eighteen clusters composed of no less than 50 voxels each were selected as ROIs; the color of dendrogram leaves in these ROI-clusters was retained in the bar, while leaves of the non-ROI clusters were shown in gray. ***c***, Each leaf of the dendrogram coincides with the corresponding column in the color-coded matrix, where the TO index matrix is shown in the upper triangle, and the genetic correlation (r_g) matrix in the lower triangle. Compared with r_g, the TO index identified clusters that were smaller in size, but contained voxels of higher connectedness. ***d***, Selected ROIs are shown overlapped on the FA maps that are displayed in radiological orientation. These ROIs were rearranged as sequentially numbered from the inferior to superior level of the brain. Therefore, the color and numbering of ROIs in ***d*** match those on the bar in ***b***, but the numbering of the ROIs in ***b*** is not in numerical order.

this study, we examined three verbal (information, arithmetic, and vocabulary) and two performance (spatial and object assembly) sub tests. Each subtest gave a raw score, and verbal (VIQ), performance (PIQ), and full-scale (FIQ) intelligence quotient standardized scores were derived from these sub tests. IQ data were available for 456 participants (Table 1).

Image processing and registration. All MR images were collected using a 4 tesla Bruker Medspec MRI scanner (Bruker Medical), with a transverse electromagnetic (TEM) headcoil, at the Center for Advanced Imaging (University of Queensland, Australia). Diffusion-weighted scans were acquired using single-shot echo planar imaging with a twice-refocused spin echo sequence, to reduce eddy-current induced distortions. Imaging parameters were as follows: 21 axial slices (5 mm thick), FOV = 23 cm, TR/TE 6090/91.7 ms, 0.5 mm gap, with a 128×100 acquisition matrix. Thirty images were acquired: three with no diffusion sensitization (i.e., T_2-weighted images) and 27 diffusion-weighted images ($b = 1145.7$ s/mm^2) with gradient directions evenly distributed on an imaginary hemisphere. The reconstruction matrix was 128×128, yielding a 1.8×1.8 mm^2 in-plane resolution. Total scan time was 3.05 min. We used the

FMRIB software library (FSL, http://www.fmrib.ox.ac.uk/fsl/) to preprocess and linearly align the diffusion images. For each participant, motion artifacts were corrected by linearly registering all the T_2-weighted and diffusion-weighted images to one of the T_2-weighted images (the "eddy_correct" command). Then the three T_2-weighted images were averaged and stripped of nonbrain tissues to yield a binary brain extraction mask (cerebellum included), using the Brain Extraction Tool (BET) (Smith, 2002), followed by expert manual editing, if necessary. The masked T_2-weighted image was then registered to a standardized high-resolution brain MRI template defined in the International Consortium for Brain Mapping space (ICBM) (Holmes et al., 1998) with a nine-parameter linear transformation using the software FLIRT (Jenkinson and Smith, 2001). The resulting transformation parameters were used to rotationally reorient the diffusion tensors (computed from diffusion-weighted images using the "DTIFIT" command) at each voxel (Alexander et al., 2001). The tensor-valued images were linearly realigned based on trilinear interpolation of the log-transformed tensors (Arsigny et al., 2005) and resampled to isotropic voxel resolution (with dimensions: $128 \times 128 \times$

93 voxels, resolution: $1.7 \times 1.7 \times 1.7$ mm^3). A fractional anisotropy (FA) map was then constructed from each affine-registered diffusion tensor image. FA is defined as the ratio of the SD to the root mean square of the eigenvalues of a diffusion tensor (Basser and Pierpaoli, 1996). Each participant's FA image was then registered to a mean FA image computed for the first 258 subjects scanned [a subset of the study sample in this paper; (Chiang et al., 2009a)]. For this alignment step, we used a validated fluid registration algorithm that maximizes the Jensen–Rényi divergence of the joint intensity histogram of the two FA images (Chiang et al., 2007).

We averaged the fluidly registered FA images across all subjects and restricted subsequent data analysis to regions with average FA > 0.2, as recommended by Smith et al. (2006). This focused our regions of interest on major white matter fiber structures. Each participant's FA map was smoothed using an isotropic Gaussian filter with full-width at half maximum (FWHM) = 6 mm (Smith et al., 2006).

Defining regions of interest for GWA. We defined regions of interest (ROIs) for GWA from brain regions where genetic influences on white matter integrity were high, based on a standard structural equation model (SEM) to model the observed variation of FA, which is widely used in twin studies. The SEM partitions the observed variance into components (proportions of the overall variance) due to additive genetic factors (A), shared environment (C), and unique environment (E). In our sample, where nontwin siblings are also included, we used the extended twin design (Posthuma et al., 2000; Lenroot et al., 2009) where the covariance matrix for FA was modeled for each family based on the known genetic similarity between relatives. The parameters of the SEM were estimated using a maximum likelihood scheme to maximize their fit (Neale et al., 1992). The significance map for the genetic factors, p(A), was determined as minus two times the difference between the log-likelihood of the full (ACE) and the restricted (CE) model at each voxel, which is asymptotically distributed approximately as a χ^2 distribution with one degree of freedom. p(A) was further assessed using the false discovery rate method by Benjamini and Hochberg (1995) (BH-FDR) to correct for multiple comparisons. We selected voxels where the genetic component A contributed to at least 60% of the total variation in FA and its significance passed the BH-FDR $\leq$ 0.05 threshold. This led to 4876 candidate voxels for the ROIs (5% of the total 99,480 voxels). By doing this, we ensured that the main effect (i.e., genetic influences on FA) for the genetic correlation analysis below was significant, and no more than 5% of the voxels declared as genetically influenced are likely to be false positive findings.

We estimated the genetic correlation coefficient r_g between any two voxels, using the cross-trait cross-twin method (Chiang et al., 2009b), yielding a 4876-by-4876 genetic correlation matrix. r_g is the correlation between the genetic effects influencing two traits x and y, i.e., $\text{Cov}(G_x, G_y)/\sqrt{\text{Var}(G_x) \cdot \text{Var}(G_y)}$, where G_x and G_y are the genetic effects that influence traits x and y, respectively. $\text{Var}(G_x)$ and $\text{Cov}(G_x, G_y)$ denote the variance of G_x and the covariance of G_x and G_y. Note that r_g is not the proportion of total covariance due to genes, i.e., $\text{Cov}(G_x, G_y)/[\text{Cov}(G_x, G_y) + \text{Cov}(E_x, E_y)]$, where E_x and E_y are respectively the environmental effects that influence traits x and y; this term can be quite small even if the genetic correlation is very large (even unity). We then identified clusters of voxels that were highly similar in their genetic determination, and defined these as our ROIs, using the hierarchical clustering method in the Matlab Statistics Toolbox (The MathWorks). We assumed that it would be more likely to detect associations between SNPs identified across the genome and FA in these ROIs than in other regions, as the FA of voxels in an individual ROI was strongly determined by a common set of genes. Hierarchical clustering initially treated each of the 4876 voxels as a node in a network. It began with each node as a single cluster, and merged every two nodes with the highest similarity, or the shortest distance, into a successively larger cluster. Genetic similarity between any two nodes was quantified by the topological overlap (TO) index, which was based on r_g and measured the interconnectivity between the two nodes (Ravasz et al., 2002; Zhang and Horvath, 2005), as follows:

$$\text{TO}_{ij} = \frac{l_{ij} + a_{ij}}{\min\{k_i, k_j\} + 1 - a_{ij}}.$$

Table 2. SNPs whose associations with FA reach overall significance, at levels that correct for the genome-wide search, and correct for the number of ROIs assessed

SNP	ROIs (β, P)[a]	Chromosome (position in base pairs)	Name of gene containing the SNP[b]
rs12410007	6 (0.03, 3.0×10^{-8})	1 (15136857)	KAZN
	9 (0.03, 1.0×10^{-8})		
rs12725686	12 (0.09, 2.0×10^{-11})	1 (15481005)	FHAD1
	13 (0.07, 9.0×10^{-10})		
rs850880	11 (-0.02, 9.0×10^{-8})	2 (185913434)	
	15 (-0.02, 6.0×10^{-8})		
rs13097297	12 (0.05, 8.0×10^{-8})	3 (29150402)	
	13 (0.04, 2.0×10^{-8})		
rs17038289	13 (0.06, 1.0×10^{-8})	4 (109131688)	HADH
rs6915645	8 (-0.01, 6.0×10^{-10})	6 (143882733)	
rs2710234	13 (0.03, 2.0×10^{-8})	7 (152432764)	
rs10091460	12 (0.09, 1.0×10^{-9})	8 (24590168)	
	13 (0.09, 2.0×10^{-11})		
rs10088359	12 (0.09, 8.0×10^{-10})	8 (24592338)	
	13 (0.09, 1.0×10^{-11})		
rs11135816	12 (0.09, 8.0×10^{-10})	8 (24603960)	
	13 (0.09, 1.0×10^{-11})		
rs12680982	13 (0.05, 3.0×10^{-8})	8 (24680599)	
r4872262	13 (0.05, 3.0×10^{-8})	8 (24681608)	
rs10990619	12 (0.04, 1.0×10^{-8})	9 (105005484)	
	13 (0.03, 6.0×10^{-11})		
rs10742552	12 (0.08, 3.0×10^{-8})	11 (5145403)	
	13 (0.07, 1.0×10^{-8})		
rs897102	1 (0.02, 8.0×10^{-9})	11 (8288070)	
rs11238009	8 (0.02, 4.0×10^{-8})	11 (79237030)	
rs10894606	13 (0.02, 2.0×10^{-8})	11 (132176821)	OPCML
rs10844932	12 (0.05, 2.0×10^{-8})	12 (34564711)	
rs11829528	12 (0.06, 2.0×10^{-9})	12 (34669982)	
	13 (0.05, 2.0×10^{-8})		
rs16946541	13 (0.03, 8.0×10^{-8})	12 (115142545)	
rs1009981	13 (0.07, 2.0×10^{-8})	12 (115588318)	
rs17126072	8 (0.02, 2.0×10^{-8})	14 (52577434)	DDHD1
rs6498964	12 (0.05, 7.0×10^{-8})	16 (63502473)	
	13 (0.05, 9.0×10^{-9})		
rs678156	13 (0.04, 6.0×10^{-8})	18 (2992436)	LPIN2

[a]β is the partial regression coefficient that represents change in the average FA of the ROI due to the unit additive effect of the major allele of the SNP; p is the significance of β (uncorrected for multiple comparisons).

[b]See Table 5 for the full name and function of these genes.

Here $a_{ij} = 0$ or 1 is the adjacency function between nodes i and j. $a_{ij} = 1$ if the genetic correlation r_g between them is not <0.95, and indicates that the two nodes are connected. $l_{ij} = \sum_u a_{iu} a_{uj}$ is the number of nodes connected to both nodes i and j, and $k_i = \sum_u a_{iu}$ is the degree of node i (Rubinov and Sporns, 2010). TO_{ij} lies between 0 and 1; a higher value of TO_{ij} indicates that nodes i and j are linked to more common neighbors. We used the TO index for network clustering, as it has been shown to provide more distinct aggregations of network nodes than direct correlation (r_g) (Zhang and Horvath, 2005). We defined the distance between nodes i and j as $d_{ij} = 1 - \text{TO}_{ij}$. The distance between two clusters that contain more than one node was given by the average distance across all pairs of nodes in the two clusters. Finally, we created a hierarchical tree of clusters at different distance levels (Fig. 2). Here, a cluster at level d means that the distance between every pair of all the subclusters it contains is not greater than d. By setting a threshold $d = 0.5$, we obtained 233 clusters. We then selected those of no less than 50 voxels (246 mm^3) in size as our ROIs; 18 ROIs were identified (Fig. 2). Setting a high threshold on ROI size might bias gene hunting in favor of SNPs with extensive effects on white matter integrity but ignore those with strong associations with FA in small regions only. Even so, we assumed that effects detected in larger ROIs are more likely to be neurophysiologically meaningful, especially when detected in the presence of imaging noise.

GWA for white matter integrity. The mean FA for each ROI was associated respectively with each SNP across the whole genome ($n = 529,497$)

**Table 3. The topological properties of the genetic network**

ρ_T	Small-world					Scale-free[a]		
	C	L	C_{rand}	L_{rand}	S	λ	Signed R^2	p value
0.4	0.71	2.05	0.41	1.72	1.44	0.23	−0.07	0.004
0.5	0.70	2.44	0.37	1.81	1.39	0.06	−0.01	0.29
0.6	0.63	2.92	0.31	1.92	1.36	−0.21	0.10	0.001
0.7	0.58	3.86	0.23	2.11	1.40	−0.49	0.38	1.31×10^{-9}
0.8	0.44	2.87	0.12	2.45	3.00	−0.77	0.68	2.39×10^{-12}
0.9	0.14	2.70	0.03	3.44	6.70	−1.58	0.93	3.26×10^{-11}

[a]To test the scale-free property of the network, the frequency of the nodes that have degree k, denoted by $f(k)$, was fitted to a power function of k, given by $f(k) = \text{constant} \cdot k^{-\lambda}$, where λ is also the slope of the fitted line of the log-log plot for $\log(f(k))$ regressed against $\log(k)$. Signed $R^2 = -\text{sgn}(\lambda) \cdot R^2$. Here sgn represents the sign function; R^2 is the Pearson's coefficient of determination for the fitted line, whose significance is given by the p value in the last column.

using the Quantitative Transmission Disequilibrium Test (qTDT) model, where the phenotype (mean FA of the ROI here) was regressed against the SNP genotypic value, with variance components that accounted for within-family phenotypic correlations included in the model (Abecasis et al., 2000). If an SNP has two alleles A and a, and allele A is the major allele that appears more frequent than allele a, then the SNP genotypic value equals 1 for AA, 0 for Aa, and −1 for aa. Subjects' age and sex were included as covariates. qTDT tests the significance of the association between FA and an SNP genotype effect by comparing the difference in log-likelihood functions between the full model (the SNP effect was included) and the restricted (the SNP effect was excluded) model. To increase power for detecting associations, we used the whole genotypic value of the SNPs for the association analysis (total association), instead of dividing the genotypic value into between-family and within-family components (Abecasis et al., 2000). Since the ethnic outliers had been removed from GWA, the risk for detecting spurious associations due to population stratification was minimized.

Corrections for multiple comparisons. We adopted a two-stage approach to correct for multiple comparisons across the genome and across the image. For associations between FA and SNPs, we first used the traditional Bonferroni method to correct for multiple comparisons across the genome (Lander and Kruglyak, 1995): SNPs with a p value <0.05/(number of SNPs across the genome) = 0.05/529,497 = 9.4×10^{-8} were declared to be significantly associated with FA. For each of these significant SNPs, we then further corrected for multiple comparisons across the ROIs using the bootstrap method (Rhodes et al., 2002). The p value of the SNP was converted to a Z value using the inverse normal cumulative distribution function. Then the p value adjusted for multiple comparisons, denoted by q, was derived by comparing the sum of Z values across all the 18 ROIs for that SNP with the bootstrap distribution of 100,000 sums of Z values randomly selected from the 529,497 SNPs. An SNP with $q < 0.05$ was considered to reach overall significance.

Genetic interconnection network and network topology. We then extended our analysis from detecting the effects of individual genetic loci at single locations in the image, to constructing a genetic network that underlies white matter integrity over multiple brain regions. Our methods were inspired by previous work deriving gene coexpression networks (Zhang and Horvath, 2005; Ghazalpour et al., 2006) but modified to deal with the case where SNPs in genes may influence multiple phenotypes (different parts of a brain image). In the network scheme, the connectedness between SNPs was derived from the correlations between their phenotypic effects. We selected SNPs whose association with FA in at least one ROI was with a significance p value $< 10^{-5}$; this criterion was somewhat relaxed (compared with the very strict GWA-significant threshold with $p < 9.4 \times 10^{-8}$) to include more SNPs into the network analysis and might therefore boost sensitivity for detecting interactions of gene effects. Three-hundred and twenty-nine SNPs were selected. We then computed the Spearman's rank correlation coefficient ρ for every pair of SNPs, based on their partial regression coefficients with respect to FA across all the 18 ROIs. The partial regression coefficient β for an SNP represents change in FA due to the unit additive effect of the major allele. β could be positive, zero, or negative, depending on the direction of the SNP major allele effect on FA. We used $\mathbf{B}_i$ to denote the vector that contains β values across all the 18 ROIs for SNP i. SNPs i and j were considered to be connected (i.e., $a_{ij} = 1$) if (1) the correlation between $\mathbf{B}_i$ and $\mathbf{B}_j$, denoted by ρ_{ij} was no less than a threshold ρ_T, and, (2) the two
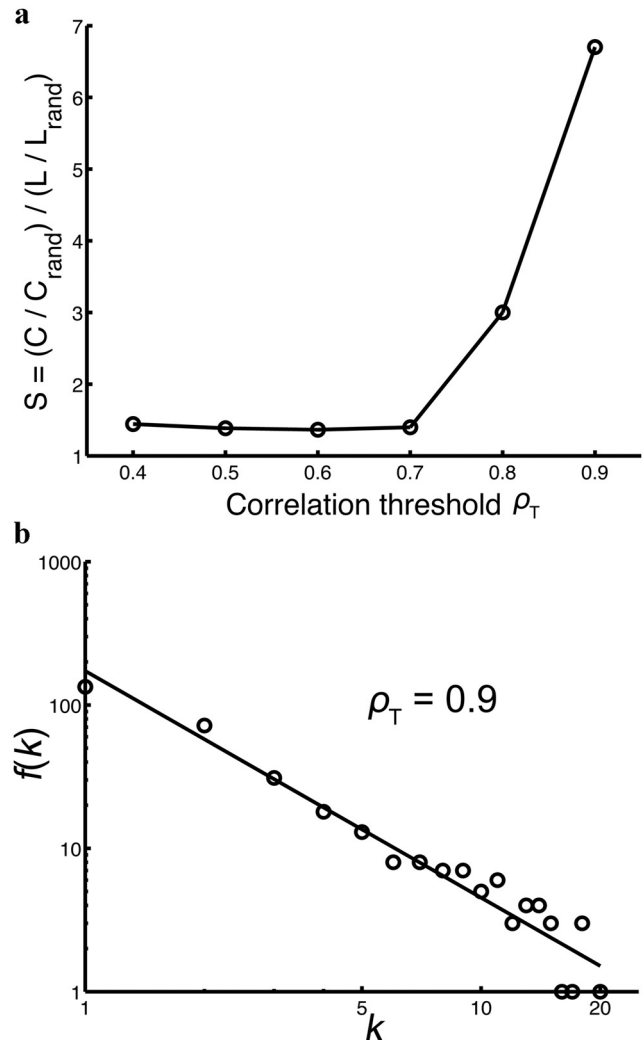
Figure 3. Network topology at different correlation thresholds ρ_T. ***a***, We measured the small-world topology of the genetic network by estimating the small-worldness index, S (see Materials and Methods for definition) at different correlation thresholds ρ_T. The nodes of the network were SNPs associated with white matter integrity with a significance p value $< 10^{-5}$ in at least one ROI. Two nodes (SNPs) are defined as "connected," if they have strongly correlated effects on FA (the Spearman rank correlation coefficient ρ between their respective association with FA is no less than ρ_T) and are not in linkage disequilibrium (linkage disequilibrium $r^2 \leq 0.2$). Every ρ_T yielded a small-world network ($S > 1$). However, the value of S at $\rho_T \geq 0.8$ is much higher than that with respect to a smaller ρ_T (0.4–0.7), suggesting that the efficiency of the genetic network is determined by SNPs with strong correlations in their effects. ***b***, When $\rho_T = 0.9$, the genetic network has a scale-free topology, as shown by the log–log plot (the network is visualized in Fig. 4 below). k is the number of connections that an SNP has, and $k = 1$ means that the SNPs are isolated and not connected to any other SNPs. $f(k)$ is the number of SNPs with k connections, and $f(k) = 144.00 \times k^{-1.58}$ (Pearson R^2 for line fitting = 0.93, $p = 3.26 \times 10^{-11}$).

a

b
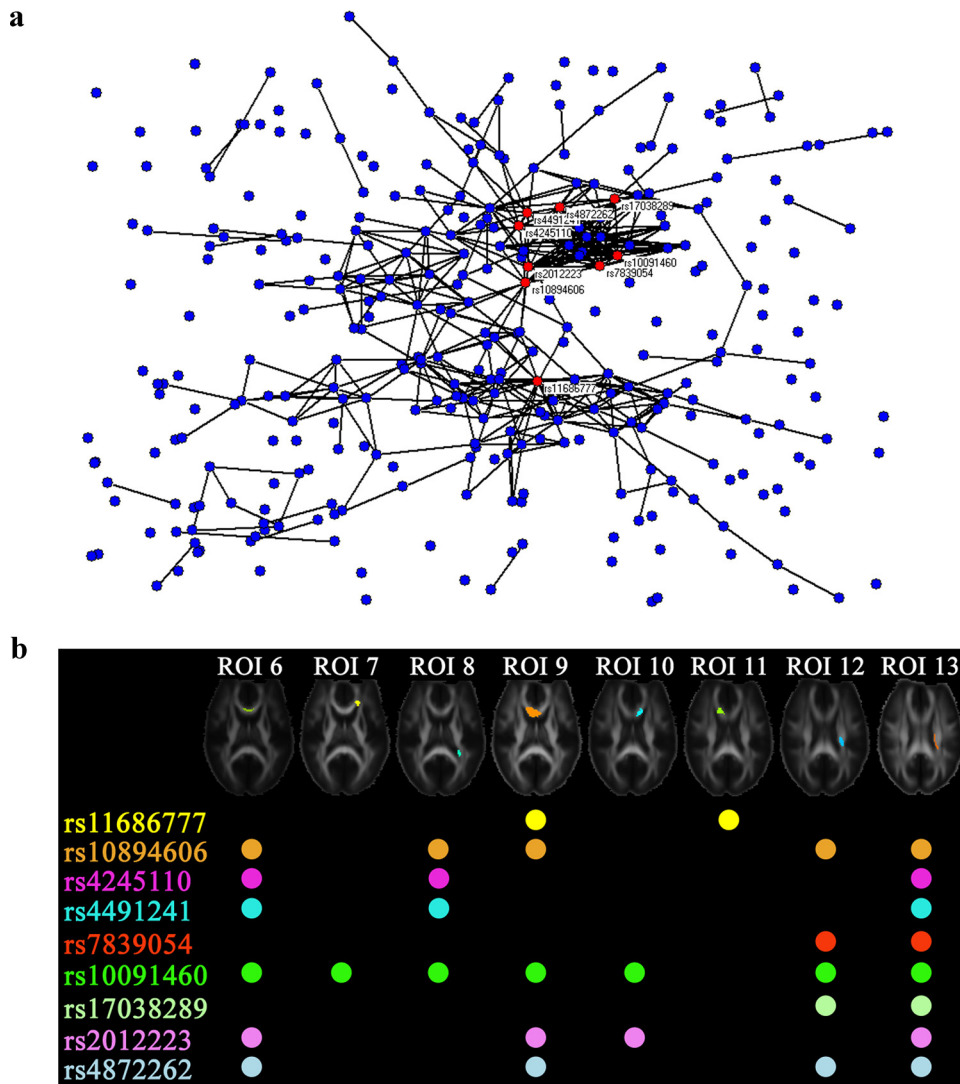
Figure 4. Visualization of the genetic network at $\rho_T = 0.9$. ***a***, The nodes of the network are displayed as blue circles, and two connected nodes are linked by a black line. See Materials and Methods or Figure 3 for the definition of network nodes and their connections. SNPs that have no less than 15 connections ($k \geq 15$) are defined as the hubs of the network and are colored in red, with their names labeled (see ***b*** for the names of these SNPs). To identify white matter regions that the genetic network is most likely to influence, associations between these hub SNPs and the white matter ROIs (see Fig. 2) are demonstrated by a matrix of circles in ***b***. Each circle is the same color as the corresponding hub SNP on the same row, to ease interpretation. When a circle is present, it indicates that the hub SNP was associated with the white matter integrity of the ROI (shown in the same column as the circle) with a significance p value $< 10^{-5}$. ROIs that did not show significant associations with any hub SNP (e.g., ROIs 1–5 and 14–18) are not shown.

SNPs were not in linkage disequilibrium (LD), defined as a pairwise LD r^2 for the SNPs ≤ 0.2 (Hill and Robertson, 1968). The latter criterion ensured that the connectedness between these two SNPs was not simply due to their positional proximity in the chromosomes, which was reflected by a higher LD between them. The topological configuration of the genetic network was a function of ρ_T. Here, we tested different ρ_T values from 0.4 to 0.9, with increments of 0.1. For each ρ_T, we computed three common measures of the network: the degree of each node (SNP), the clustering coefficient C, and the characteristic path length L (Watts and Strogatz, 1998; Bullmore and Sporns, 2009; Rubinov and Sporns, 2010). The clustering coefficient C is a measure of segregation in the network; a higher value of C indicates that the nodes of the network tend to form isolated clusters (Watts and Strogatz, 1998). On the other hand, the characteristic path length L is an index of integration in the network. It measures the average shortest path length between all pairs of nodes in the network; a smaller L indicates that the topological distance between the segregated clusters in the network is shorter, or the integration between these segregated clusters is more efficient. When a network is simultaneously highly segregated and integrated, it is in a small-world topology (Rubinov and Sporns, 2010). This small-world property is quantitatively estimated by the "small-worldness" of the network

Table 4. Hub SNPs of the genetic network

SNP	Number of connections, n	Chromosome (location at the chromosome)	Name of gene containing the SNP[b]
rs11686777	20	2 (165673614)	SCN3A
rs10894606[a]	18	11 (132176821)	OPCML
rs4245110	18	11 (132203925)	OPCML
rs4491241	18	11 (132205441)	OPCML
rs7839054	17	8 (24565907)	
rs10091460[a]	16	8 (24590168)	
rs17038289[a]	15	4 (109131688)	HADH
rs2012223	15	11 (132130097)	OPCML
rs4872262[a]	15	8 (24681608)	

[a]These SNPs are also genome-wide significant.

[b]See Table 5 for the full name and function of these genes.

(Humphries and Gurney, 2008): $S = (C/C_{rand})/(L/L_{rand})$. Here, C_{rand} and L_{rand} are the mean clustering coefficient and the characteristic path length of N_r random networks, where each random network was generated by randomly rewiring the links of the original network but keeping the distribution of the degree of the nodes unchanged (Maslov and Snep-

pen, 2002). N_r was set to 1000. A network is deemed a small-world if $S > 1$ (Humphries and Gurney, 2008).

We further tested whether the genetic network was also in a scale-free topology. The degree distribution of a scale-free network follows a power law, i.e., $f(k) \propto k^{-\lambda}$, where $f(k)$ is the frequency of the nodes that have degree k (Barabasi and Albert, 1999). A scale-free network is known to be more tolerant of random errors (Albert et al., 2000).

Gene ontology enrichment analysis. To identify possible functional roles for the genes in the network, we performed gene ontology (GO) enrichment analysis using the web-based Gene Set Enrichment Analysis Toolkit (Web-Gestalt2) (Zhang et al., 2005; Duncan et al., 2010). The 329 SNPs selected for network analysis were first annotated to corresponding genes, where possible, by searching the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/batchquery.html), yielding a total of 101 genes. These genes were then compared with all genes in the human genome using the hypergeometric test, and the genes were assigned, where possible, to enriched GO categories under three main domains, cellular components, molecular functions, and biological processes. Each GO category contained no less than two genes. Multiple comparisons across these enriched categories were corrected using the BH-FDR method (Benjamini and Hochberg, 1995), with the significance level for the FDR-adjusted p value set to 0.05. The results were represented by an enriched directed acyclic graph (DAG).

Detecting genetic interactions between SNPs and interaction effects of SNPs on FA-IQ associations. Finding connections between SNPs defined in the above section may help to search for epistasis (genetic interactions) between SNPs that affect white matter integrity, i.e., where the presence of one SNP may modify or interact with the effect of other SNPs. To explore this, we estimated the interaction between SNPs whose associations with FA were with a significance p value $< 10^{-5}$ in at least one ROI ($n = 329$ SNPs), by adding an interaction term to the qTDT model: $y = \mu + \mathbf{x}_N\beta_N + g_1\beta_1 + g_2\beta_2 + (g_1 \cdot g_2) \beta_{int}$.

Here, y is the mean FA within an ROI. For voxel-based analysis, y is the FA of an individual voxel in the white matter mask (defined as FA > 0.2). μ is the overall mean of FA, and $\mathbf{x}_N$ is a vector of covariates that includes the subjects' age and sex. g_1 and g_2 indicate the genotypic value of SNP1 and SNP2 respectively, and $(g_1 \cdot g_2)$ represents the interaction between the two SNPs.

We also tested whether the hub SNPs in the scale-free genetic network (see Results for the 9 hub SNPs) may affect IQ via its modulatory effect on the association between white matter integrity and IQ. Here, we estimated the interaction of a hub SNP with FA and the PIQ scale at every voxel within the white matter mask; we selected PIQ, as in our previous study (Chiang et al., 2009b) we found that the cross-trait correlation between FA and PIQ was strongly mediated by overlapping genetic factors. This was modeled by: $y = \mu + \mathbf{x}_N\beta_N + g \cdot \beta_1 + \text{PIQ} \cdot \beta_2 + (g \cdot \text{PIQ}) \beta_{int}$.

Here, g is the genotypic value of a hub SNP, PIQ indicates the participant's PIQ scale, and $(g \cdot \text{PIQ})$ represents the interaction between the
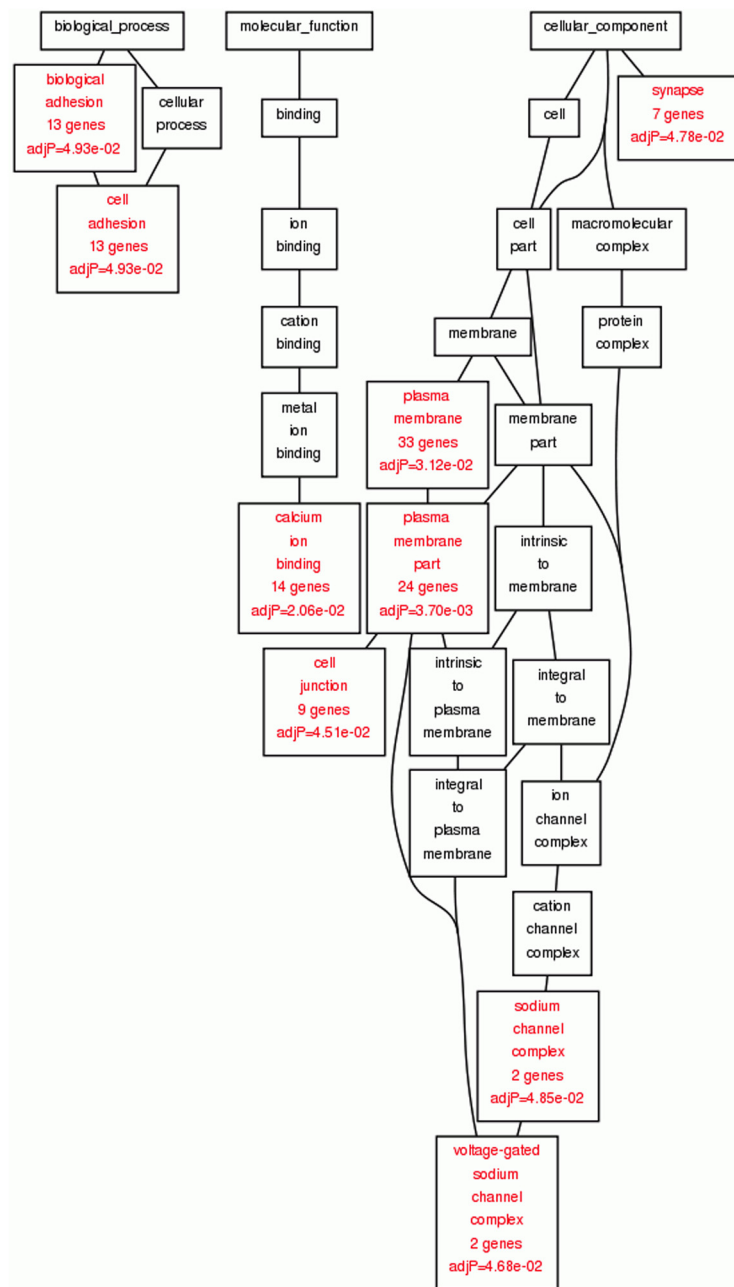
Figure 5. This directed acyclic graph (DAG) shows the gene ontology categories that are enriched in SNPs that are found in the genetic network described here. SNPs from the genetic network in Figure 4 were annotated as belonging to a set of 101 genes. These genes were further annotated and organized into enriched GO categories using WebGestalt2 (Zhang et al., 2005; Duncan et al., 2010), with all genes in the human genome as a reference (see Materials and Methods). These enriched GO categories were hierarchically organized into a DAG tree, where each box in the tree lists the name of the GO category, the number of genes in that category, as well as the FDR-adjusted p value (adjP) if the enrichment is significant. Categories in red are enriched ones (adjusted p value $<$0.05), while those in black are non-enriched.

hub SNP and PIQ. The definition of y, μ, and $\mathbf{x}_N$ are the same as in the previous regression equation for genetic interactions between SNPs.

$\beta_{int} \neq 0$ indicates a significant interaction effect, which was determined as minus two times the difference between the log-likelihood of the full model and the restricted (β_{int} was excluded) model, which is asymptotically distributed approximately as a χ^2 distribution with one degree of freedom. We then used the Bonferroni method to correct for multiple comparisons across SNPs, with the p threshold set equal to $0.05/n_s = 1.7 \times 10^{-6}$. Here, $n_s = 29,180$ is the number of all possible pairs of the SNPs ($= 329 \times 328/2$), minus those in linkage disequilibrium (LD r^2 for the SNPs ≤ 0.2). To correct for multiple comparisons across
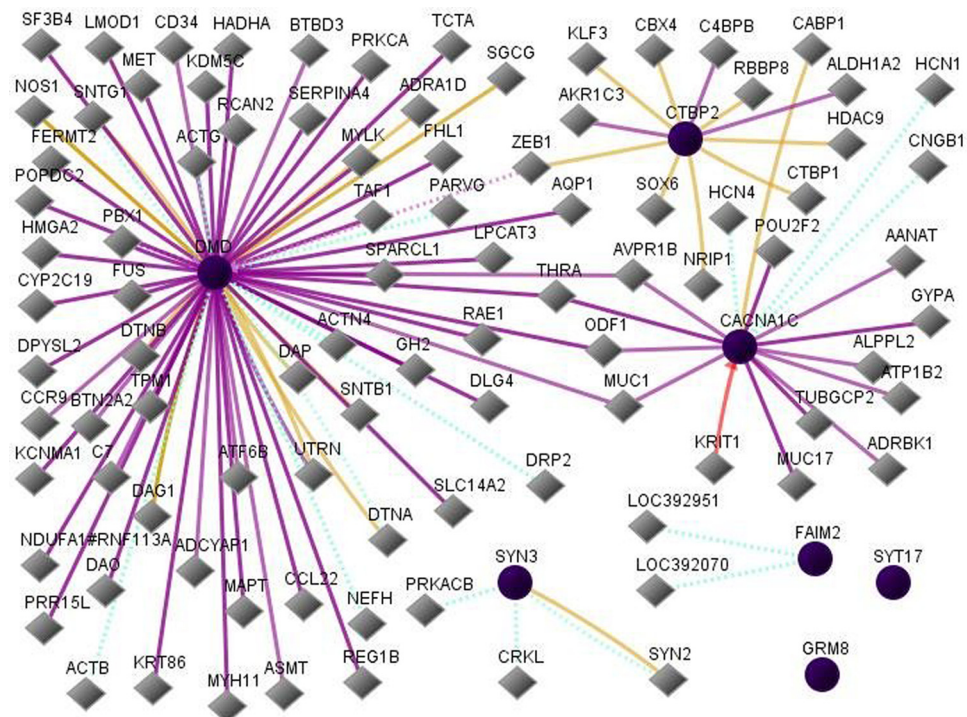
Figure 6. An interaction network involving some of the genes we recovered, i.e., those influencing synaptic integrity in prior studies. We used the program, GNCPro (SABiosciences), to analyze the expression relationships between the seven genes belonging to the enriched gene ontology category "synapse" (see Fig. 5), visualized as a gene interaction network (Liu et al., 2010). These seven genes are represented by solid dark purple circles; they are as follows: *DMD, CACNA1C, CTBP2, FAIM2, SYT17, SYN3*, and *GRM8*. See Table 5 for the full name and function of these genes. To detect possible indirect interactions among these genes, e.g., where a gene influences the expression of another gene via some intermediate genes, any other genes found to interact with any of the seven synapse genes were included in the analysis as neighboring genes, represented by gray lozenge nodes. These neighboring genes do not necessarily contain SNPs of the genetic network in Figure 4. The edges of the network indicate various types of interactions among the genes, represented as solid and dotted lines in color as follows: solid red lines, downregulation (the product of a gene downregulates the other one); solid purple lines, coexpression (the two genes are expressed at the same time); solid brown lines, physical interaction (products of the two genes are known to bind one another); dotted aquamarine lines, predicted protein interaction (products of the two genes are predicted to bind one another); dotted purple lines, predicted transcription factor regulation (a transcription factor has a predicted binding site for the promoter of another gene). See http://www.genecards.org/ for the full name of the genes represented by lozenge nodes.

the ROIs or the image volume, for the ROI analysis (for genetic interactions between SNPs), the interaction between two SNPs was regarded as significant, if the significance *p* value of β_{int} in any of the 18 ROIs was less than the threshold that controlled the BH-FDR within 5%. For the voxel-based analysis (for genetic interactions between SNPs, and interactions of the hub SNP with FA and PIQ), as the statistical comparisons were performed in ~100,000 voxels, we used a more stringent but accepted method, known as topological FDR (Chumbley and Friston, 2009), to control the expected proportion of false positive findings over all connected sets of voxels, i.e., clusters. We controlled the topological FDR of clusters within 5%, which means that only 5% of the clusters that are declared to be significant tend, on average, to be false positive findings.

Results

To perform GWA for white matter integrity measured by FA, we first identified 18 ROIs using hierarchical clustering based on the 4876-by-4876 topological overlap (TO) matrix (Fig. 2). We assumed that there should be more power to detect associations between FA in these ROIs and SNPs identified from the whole genome, as the FA of voxels within an individual ROI was determined by a common set of genes. Twenty-four SNPs were detected in GWA (Table 2), which showed strong "genome-wide significant" associations with white matter integrity (FA), with an uncorrected *p* value $<9.4 \times 10^{-8}$, and an overall significance (across all 18 ROIs) *q* value <0.05.

Next, we were interested in how individual SNPs were correlated under a network configuration to affect brain integrity, or in other words, how the genetic network drives the brain network. We defined the nodes of the network as SNPs that were associated with

white matter integrity with a significance *p* value $< 10^{-5}$ in at least one ROI (number of nodes = 329). We then examined the topological properties of the genetic network, by increasing the correlation threshold ρ_T from 0.4 to 0.9 (Table 3). Every ρ_T yielded a small-world network ($S > 1$). However, when the network consisted of nodes with stronger connections only ($\rho_T \geq 0.8$), S increased abruptly, as also shown in Figure 3*a*, suggesting that the efficiency of the genetic network was determined by SNPs with stronger correlations in their effects. For the scale-free property, given that the slope of the fitted line for $\log(f(k))$ versus $\log(k)$ should be negative in a typical scale-free configuration, the genetic network can be considered to be scale-free only when $\rho_T \geq 0.7$, where *p* values for the fit are also very small. If we further considered the effect size for the above correlation fitted by requiring a signed $R^2 > 0.8$ (Zhang and Horvath, 2005), only $\rho_T = 0.9$ yielded a scale-free network by this definition, where $f(k) \propto k^{-1.58}$ (Fig. 3*b*, signed $R^2 = 0.93$, $p = 3.26 \times 10^{-11}$). Signed R^2 here was derived by adjusting the sign of the Pearson's coefficient of determination R^2, or signed $R^2 = -\text{sgn}(\lambda) \cdot R^2$, where sgn represents the sign function. In other words, if the fitted line for $\log(f(k))$ versus $\log(k)$ has a negative (positive) slope λ, it would have a positive (negative) signed R^2. Note that $\rho_T = 0.9$ also gave the highest small-worldness index ($S = 6.70$), which corroborated the theoretical prediction that scale-free networks are ultra-small (Cohen and Havlin, 2003).

Figure 4 visualizes the genetic network at $\rho_T = 0.9$. The visual layout of these SNPs is based on the Kamada–Kawai algorithm (Kamada and Kawai, 1989) implemented in the Pajek software

[http://vlado.fmf.uni-lj.si/pub/networks/pajek/ (Batagelj and Mrvar, 1998)]. This SNP network obeys a scale-free topology. This means that most SNPs have no or only scarce connections with other SNPs. However, a small proportion of SNPs is highly connected with each other and they are therefore the hubs of the network. For example, 134 SNPs (41% of all 329 SNPs) are isolated and not connected with any other SNPs ($k = 1$). On the other hand, 9 SNPs ($< 3\%$ of all SNPs) are highly connected with other SNPs, and thus the hubs of the network. They are (Fig. 4a, labeled as red circles) rs11686777 ($k = 20$), rs10894606, rs4245110, and rs4491241 ($k = 18$), rs7839054 ($k = 17$), rs10091460 ($k = 16$), rs17038289, rs2012223, and rs4872262 ($k = 15$) (Table 4). rs10894606, rs10091460, rs17038289, and rs4872262 are also genome-wide significant. These hub SNPs were associated with FA in the ROIs identified in major white matter tracts (Fig. 4b). Specifically, almost all (8 of 9) hub SNPs were associated with white matter integrity in ROI 13 in the superior *corona radiata*, a white matter structure containing axons projecting to and from the cerebral cortex.

To determine the functional significance of the genetic network, we further used the program, WebGestalt2, to assess gene ontology (GO) enrichment for the 101 genes to which the nodes (SNPs) of the network were annotated. Figure 5 shows the directed acyclic graph showing patterns of genetic enrichment. Thirteen genes are enriched in "biological processes," with their function specified as "cell adhesion." Fourteen genes belong to the "molecular functions" domain, particularly for calcium ion binding. Forty-two genes are enriched in cellular components relevant to the integrity of nerve fibers: 7 of them are implicated in the integrity of synapses, 33 for the plasma membrane, and 2 for voltage-gated sodium channel complexes. We further analyzed the seven genes in our network that influence the integrity of synapses, using the GNCPro program (http://gncpro.sabiosciences.com/gncpro/gncpro.php). GNCPro is a visualization/data integration system that integrates collective biological knowledge and databases to visualize gene expression interactions as a network (Liu et al., 2010). Figure 6 visualizes various interaction patterns between these seven genes and their neighboring genes in a network configuration. For example, the *DMD* (dystrophin) and the *CACNA1C* (calcium channel, voltage-dependent, L type, α 1C subunit) genes, both belong to the enriched GO category "synapse. " They interact with each other by coexpressing with *AVPR1B* (arginine vasopressin receptor 1B) and *THRA* (thyroid hormone receptor, α)—two hormone-related genes that influence cerebral metabolism and cognitive function (Esaki et al., 2003; DeVito et al., 2009).

We found that network connectivity between SNPs determined their genetic interactions. Figure 7a shows that SNPs with more connections in the network had on average greater number of interactions with other SNPs. In other words, hub SNPs in the network are more likely to interact with other SNPs in their effects on FA than nonhub ones. This trend is clear in Figure 7b, where the number of interactions for an SNP with degree k is weighted by the probability of finding such an SNP. The fitted curve in Figure 7b shows that hub SNPs in the network are very promising candidates for genome-wide searches for genetic interactions (GWIA), as the likelihood for detecting interactions for these SNPs with others is raised to nearly the second power of their connectivity degree. Figure 8 displays a map of interactions between two hub SNPs at different chromosomes, rs10091460 and rs10894606 (the former is on chromosome 8 and the latter on chromosome 11; both are genome-wide significant). The voxel-based analysis shows that interactions between these two SNPs are highly significant across major white matter
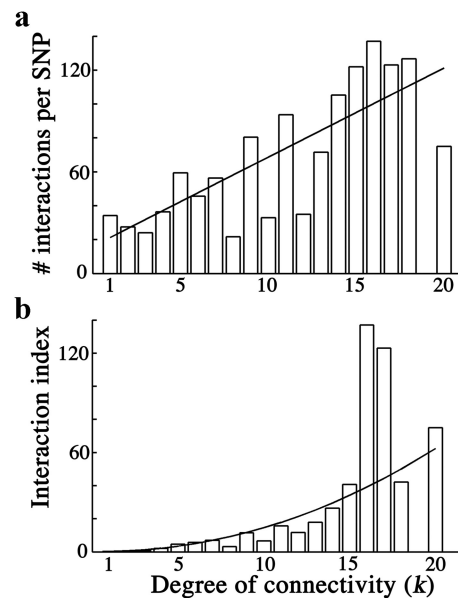

Figure 7. Connectivity between SNPs predicts their interactions. Each bar in *a* indicates how many other SNPs, on average, interact with a single SNP with degree k, in terms of their statistical influences on white matter integrity. Two SNPs are defined as statistically interacting with each other if the significance p value of β_{int} in any of the 18 ROIs is less than the threshold that controls the BH-FDR within 5%, as described in Materials and Methods. The fitted line in the bar graph shows that if an SNP has more connections, it is more likely to interact with other SNPs on affecting FA (y-coordinate $= 15.91 + 5.26k$; Pearson R^2 for line fitting $= 60\%$; Spearman's rank correlation coefficient $= 0.76$, $p = 1.4 \times 10^{-4}$). Even so, considering that the frequency of SNPs with k connections follows a power-law distribution with respect to k (Fig. 3), in (*b*), we developed an interaction index, which was the ratio of the average number of SNPs that interacted with SNPs with k connections (i.e., the y-coordinate in plot *a*), over the number of SNPs with k connections. This interaction index may be thought of as the frequency of interactions for an SNP with k connections, conditioned on the occurrence of that SNP. The interaction index is approximately proportional to the square of k (y-coordinate $= 0.11 \times k^{2.10}$; Pearson R^2 for the power curve fitting $= 90\%$, while fitting to a line gave $R^2 = 53\%$ only; Spearman's rank correlation coefficient $= 0.96$, $p < 10^{-6}$), indicating that interactions between SNPs are much more likely to be detected for SNPs with higher connectivity.

tracts, especially in the superior *corona radiata* and superior longitudinal fasciculus bilaterally.

We next investigated the functional roles of the hub SNPs by testing their interaction effects on FA-IQ associations. Six of the nine hub SNPs indirectly influenced performance intellectual performance by modulating the associations between white matter anisotropy and performance IQ (Fig. 9). This modulatory effect was significant in white matter tracts related to IQ (Chiang et al., 2009b), such as the fornix (rs10091460 and rs4872262), genu and splenium of the corpus callosum (rs17038289), cingulum (rs17038289), the anterior limb of the internal capsule (rs10091460 and rs4872262), optic radiation (rs2012223), superior longitudinal fasciculus (rs7839054), and superior *corona radiata* (rs4245110 and rs17038289). The sign of β_{int} indicates the direction of the modulatory effects of these SNPs. We found $\beta_{\text{int}} > 0$ for rs4245110 and rs2012223, meaning that the addition of the major allele of these SNPs was associated with an increase in slope for the PIQ (on the x-axis) regressed against FA (on the y-axis). By contrast, rs7839054, rs10091460, and rs4872262 had a value of $\beta_{\text{int}} < 0$, indicating that the addition of the major allele of these SNPs was associated with decrease in slope of the FA-PIQ regression. By contrast with the other SNPs, rs17038289 showed both directions of genetic modulation, with $\beta_{\text{int}} > 0$ in the cingulum and the genu and splenium of the corpus callosum, and with $\beta_{\text{int}} < 0$ in the superior *corona radiata*.
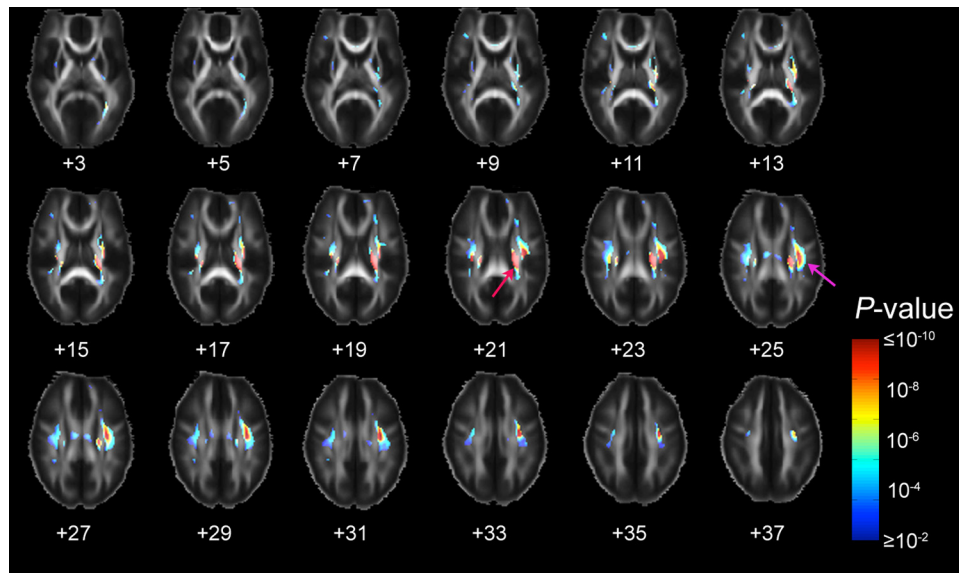
Figure 8. Interactions between two example SNPs, rs10091460 and rs10894606. These maps show regions where the two SNPs, rs10091460 and rs10894606, significantly interact with each other in affecting white matter integrity. Voxelwise significance *p* values were color-coded and displayed on a log₁₀ scale. We used the topological FDR method to correct for multiple comparisons across the significant clusters; these were defined as connected sets of voxels with the standard Normal variate $Z > 3$, where Z corresponds to the voxelwise *p* value. Only clusters that passed the topological FDR ≤ 0.05 threshold are displayed. MNI coordinates (in mm) of the slices, in radiological orientation, are shown below each slice. Interactions between these two SNPs are highly significant across major white matter tracts, especially in the superior *corona radiata* (red arrow) and superior longitudinal fasciculus (magenta arrow) bilaterally.
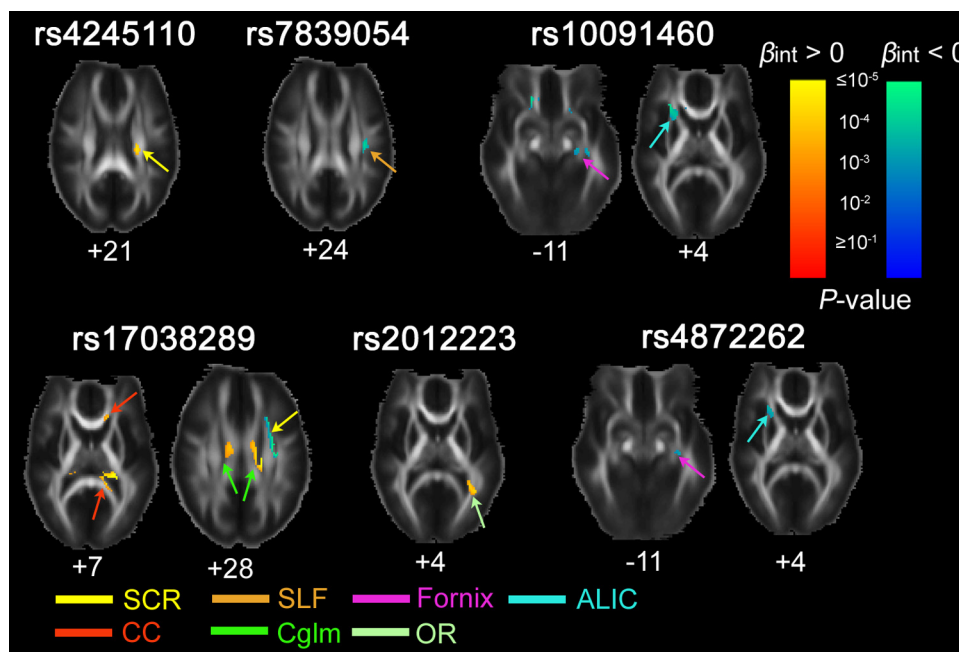
Figure 9. Interaction effects of the hub SNPs on the correlation between white matter integrity and performance IQ. These maps show regions where the hub SNPs significantly interact with associations between white matter integrity and performance IQ, displayed as clusters that passed the topological FDR ≤ 0.05 threshold. Voxels in these clusters were color-coded by their *p* value (on a log₁₀ scale) and also the direction of the interaction effect (red-yellow for $\beta_{int} > 0$, and blue-green for $\beta_{int} < 0$). MNI coordinates (in mm) of the slices, in radiological orientation, are shown below each slice. SCR, Superior *corona radiata*; SLF, superior longitudinal fasciculus; ALIC, anterior limb of the internal capsule; CC, corpus callosum; Cglm, cingulum; OR, optic radiation.

Discussion

In this study, we developed a new approach to identify networks of genetic variants that influence brain fiber integrity and intellectual performance in 472 young healthy adults. Voxel-based analysis is popular in imaging: when applied to genetic questions, it has identified genetic effects on regions of the brain without the need for an a priori hypothesis (Hulshoff Pol et al., 2006; Chiang et al., 2009b). Even so, directly performing GWA for imaging measures voxel-by-voxel may lose considerable power due to the sheer number of multiple tests—across both the image and the genome. Here we combined voxel-based and ROI approaches for GWA. Using genetic correlation to find voxels with common genetic influences, we defined ROIs, whose component voxels had FA values influenced by a highly overlapping set of genes (quantified by the genetic correlation matrix, $r_g(x,y)$, that compares all pairs of voxels). We detected 24 SNPs in 6 genes from the

Table 5. The full name and function (if known) of the gene symbols referred in this paper

Gene symbol	Name	Chromosome	Function
CACNA1C	Calcium channel, voltage-dependent, L type, alpha 1C subunit	12	Voltage-sensitive calcium channels
CTBP2	C-terminal binding protein 2	10	Encoding a major component of synaptic ribbons
DDHD1	DDHD domain containing 1	14	A probable phospholipase that hydrolyzes phosphatidic acid
DMD	Dystrophin	X	Anchoring the extracellular matrix to the cytoskeleton
FAIM2	Fas apoptotic inhibitory molecule 2	12	Protecting cells uniquely from Fas-induced apoptosis
FHAD1[a]	Forkhead-associated (FHA) phosphopeptide binding domain 1	1	
GRM8	Glutamate receptor, metabotropic 8	7	Encoding receptors for glutamate
HADH	Hydroxyacyl-CoA dehydrogenase	4	Mitochondrial beta-oxidation of short chain fatty acids
KAZN	Kazrin	1	Cell adhesion and cytoskeletal organization
LPIN2	lipin 2	18	Controlling the metabolism of fatty acids
OPCML	Opioid-binding protein/cell adhesion molecule-like	11	Binding opioids in the presence of acidic lipids; probably involved in cell contact
SCN3A	Sodium channel, voltage-gated, type III, alpha subunit	2	Mediating the voltage-dependent sodium ion permeability of excitable membranes
SYN3	Synapsin III	22	May be involved in the regulation of neurotransmitter release and synaptogenesis
SYT17[a]	Synaptotagmin XVII	16	

Information in this table is derived from the NCBI gene database (http://www.ncbi.nlm.nih.gov/gene) and the GeneCards database from the Weizmann Institute of Science (http://www.genecards.org/).
[a]Information about the function of this gene is not available in either of the two databases.

whole genome with significant effects on white matter integrity, after a highly conservative Bonferroni-type correction for multiple comparisons across the genome and across all ROIs.

The genetic network underlying white matter integrity
We then recovered a network of SNPs that influence the brain by linking SNPs with correlated effects on white matter integrity. The correlation coefficient between the phenotypic effects of two genetic loci measures the similarity of their expression patterns (Eisen et al., 1998). We further assumed that if two SNPs are connected in a network to affect the brain, the spatial distribution of their influences on brain integrity will also be similar, i.e., their partial regression coefficients (β) with respect to FA will be correlated. This correlation-based approach has been used in gene expression profiling, where genes with correlated expression levels in cDNA microarrays have been linked to form gene coexpression networks (Zhang and Horvath, 2005; Ghazalpour et al., 2006). We note that this approach to gene expression networks inspired our work here on SNP networks that influence brain images. Such gene coexpression networks in various organisms, e.g., mouse or yeast, have been shown to exhibit scale-free topology (Carlson et al., 2006; Ghazalpour et al., 2006). We note that a genetic network based on correlations between gene expression levels may better describe genetic influence on phenotypic traits than networks based on SNPs. Essentially, genes are functional units of heredity, and it is not always clear how SNPs impact gene expression. Nevertheless, the multivariate nature of a gene-wide association analysis makes it computationally intensive, especially when applied to imaging brain phenotypes (Hibar et al., 2011).

The genetic network influencing white matter integrity is both small-world and scale-free
The genetic network underlying white matter integrity exhibited a small-world topology. In a small-world network, interactions among network nodes (SNPs) are routed through "hub" nodes of the network, which makes the mean path length of the network smaller. Small-world networks are thus more efficient than random networks (Barabási and Oltvai, 2004). When the genetic network was defined by including only SNPs with strong correlations in their effects on the brain, the network became ultrasmall, with a so-called "scale-free" topology, with connection probabilities approximately following a power law. Evolutionarily, hub genes (SNPs) in the scale-free genetic network may

originate from gene duplication that follows the "preferential attachment," or "rich-gets-richer" principle (Barabási and Oltvai, 2004). For example, when any gene linked to a hub gene is duplicated, since the duplicated gene shares the same function as the original gene to some degree, the connectivity degree (k) of the hub gene will increase by 1 (i.e., rich-gets-richer). On the contrary, an isolated gene ($k = 1$) may increase its number of connections with other genes only when it is duplicated itself, which is a low probability event. Scale-free genetic networks may be evolutionarily advantageous: they are relatively error tolerant, and may remain stable even if their nodes are randomly damaged (Albert et al., 2000). As white matter integrity is critical for normal brain function, its underlying genetic network may be quite robust, offering resistance to environmental stressors. The scale-free property observed in our genetic network has been reported for several other kinds of biological networks, e.g., in gene or protein interactions, and even in the internet and in other social networks (Ravasz et al., 2002). Brain connectivity, assessed using functional MRI, may also follow a scale-free distribution (Eguíluz et al., 2005).

Functional significance of the genetic network
The gene ontology enrichment analysis indicates that the genetic network for white matter integrity is also critical for neural transmission, affecting the integrity of synapses and plasma membranes, and the function of sodium-channel complexes. Moreover, hub SNPs of the network, i.e., SNPs that were highly connected with other SNPs, belong to genes crucial for normal brain function (Tables 4, 5): the *OPCML* gene is a member of the IgLONs glycoprotein gene family, modulating cell—cell recognition and axonal migration (Reed et al., 2007). *SCN3A* encodes the α subunit of the voltage-gated sodium channel. Sodium channels mediate the voltage-dependent permeability of excitable membranes to sodium ions, facilitating the generation and propagation of action potentials in neurons (Malo et al., 1994). The *HADH* protein is one of the enzymes in the fatty acid β-oxidation pathway, which catalyzes the conversion of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA, a key step in fatty acid metabolism during brain development (Reichmann et al., 1988). This gene is also involved in amyloid β-peptide binding in Alzheimer's disease (He et al., 1998).

These hub SNPs influenced intellectual performance by modulating the associations between performance IQ and the integrity of the fiber systems that are part of visuospatial circuits

(Chiang et al., 2009b; Kravitz et al., 2011), including the optic radiations (Tuch et al., 2005), fornix (Kwok and Buckley, 2010), corpus callosum (Begré et al., 2007), cingulum bundle (Bush et al., 2000), and the superior longitudinal fasciculus, which is associated with spatial working memory performance in children (Vestergaard et al., 2011). Human intelligence is highly heritable; a substantial proportion (40–50%) of the variation in IQ is genetically determined (Davies et al., 2011). However, effects of individual SNPs on IQ are very small (~1% of variation in IQ), indicating that genetic variants may influence intellectual performance via the accumulated effects of multiple SNPs (Davies et al., 2011), or, as found here, as a modulatory factor that interacts with both white matter integrity and IQ. We also found that the same genotype of an SNP, e.g., rs17038289, could impose both positive and negative modulatory effects on FA–IQ associations in different brain regions. This may be attributed to differences in the microenvironment of the brain area where a gene is expressed (Posner and Rothbart, 2005; Lein et al., 2007; Liang et al., 2007). A similar example of a region-specific gene effect has also been found for the Val66Met polymorphism of the brain-derived neurotrophic factor (*BDNF*), a gene crucially related to cognitive performance. Dominant homozygote genotypes (Val/Val) were associated with greater volume in the hippocampus and the prefrontal cortex (Pezawas et al., 2004), but with lower integrity in major white matter tracts such as the corpus callosum and the optic radiation (Chiang et al., 2011).

Network connectivity predicts genetic interactions

The SNP network detected in Figure 4 may help to guide the search for genes that interact to influence white matter integrity. The connected SNPs, especially those at network hubs, strongly interact with each other in their effects on white matter integrity. Finding interactions between genetic variants may help in explaining their effects on normal phenotypes or disease expression (Prata et al., 2009); higher-order interactions between genetic variants are typically overlooked in GWA designs, which typically seek associations between traits and individual SNPs one at a time (Becker et al., 2011). To address this, genome-wide interaction analysis (GWIA) has become popular in recent years, to search for possible interactions between pairs of SNPs on traits of interest (Marchini et al., 2005; Becker et al., 2011). Even so, the computational requirements to apply GWIA to imaging studies are currently prohibitive. For a single voxel, a 500,000-SNP GWA requires 1.25×10^{11} statistical tests to detect SNP interactions. There is a concomitant loss in power as more stringent corrections are needed for the vast number of tests involved in GWIA versus GWA (Becker et al., 2011). Computational effort may be wasted in such a brute-force search, as most genomic SNPs have no detectable phenotypic effect, both here and in other GWA (Visscher, 2008; Stein et al., 2010). In our analyses, the network connectivity of a given SNP predicts its likelihood in interacting with other SNPs. Our method offers a simpler approach to GWIA and selects candidate SNPs by detecting correlated effects on FA; computing correlations between SNPs' partial regression coefficients (β) with respect to FA is much easier than testing how every SNP pair might interact, using a second-order regression. This may boost power to detect interactions, as promising interactions are prioritized based on advantageous search criteria.

Our method will likely empower work in imaging genomics and connectomics (e.g., the Human Connectome Project). Specifically, associations between the SNPs detected here and FA, and the modulatory effects of the hub SNPs on associations between FA and IQ, require replication. While our study was under

review, a related study was published that used clustering of the genetic correlation matrix, in twins, to define brain regions with common genetic determination (Chen et al., 2012). In that paper, the authors partitioned the cortical surface, modeled using anatomical MRI, into several regions that had partially overlapping genetic determinants. It was speculated that using these genetically clustered regions may boost the power for subsequent genome-wide analysis using GWAS. However, that study did not perform GWAS on the clustered cortical regions. In our current study, we have shown that a related method for genetic clustering can indeed be applied to diffusion images of the brain, and when GWAS was performed on the resulting clusters, the method yielded promising genome-wide associations. We further showed that the resulting SNP network showed a characteristic network topology, with scale-free and small-world properties, and implicated specific molecular pathways in driving brain integrity. Rare DNA sequence variations that contribute to genetic influences on FA may go undetected in GWA even when sample sizes are very large (Schork et al., 2009). When data from large independent cohorts become more widely available, for example via the Enigma project (http://enigma.loni.ucla.edu), two advantages of this genetic network analysis will be apparent: (1) genetic correlation can tap into the natural latent structure of gene action in a brain image; (2) voxel clustering by genetic affinity should boost power to find SNPs with correlated effects in genome-wide scans. SNP aggregation can be used for dimension reduction, so the resulting networks are strong candidates for verification and replication, without the computational and statistical burden of a full image-wide genome-wide search.

References

Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66:279–292.

Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks. Nature 406:378–382.

Alexander DC, Pierpaoli C, Basser PJ, Gee JC (2001) Spatial transformations of diffusion tensor magnetic resonance. IEEE Trans Med Imaging 20:1131–1139.

Arsigny V, Fillard P, Pennec X, Ayache N (2005) Fast and simple calculus on tensors in the log-Euclidean framework. Med Image Comput Assist Interv 8:115–122.

Barabasi AL, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512.

Barabási AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5:101–113.

Basser PJ, Pierpaoli C (1996) Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. J Magn Reson B 111:209–219.

Basser PJ, Mattiello J, LeBihan D (1994) MR diffusion tensor spectroscopy and imaging. Biophys J 66:259–267.

Batagelj V, Mrvar A (1998) Pajek—program for large network analysis. Connections 21:47–57.

Becker T, Herold C, Meesters C, Mattheisen M, Baur MP (2011) Significance levels in genome-wide interaction analysis (GWIA). Ann Hum Genet 75:29–35.

Begré S, Frommer A, von Känel R, Kiefer C, Federspiel A (2007) Relation of white matter anisotropy to visual memory in 17 healthy subjects. Brain Res 1168:60–66.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 57:289–300.

Braskie MN, Jahanshad N, Stein JL, Barysheva M, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Ringman JM, Toga AW, Thompson PM (2011) Common Alzheimer's disease risk variant within the *CLU* gene affects white matter microstructure in young adults. J Neurosci 31:6764–6770.

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical

analysis of structural and functional systems. Nat Rev Neurosci 10:186–198.

Bush G, Luu P, Posner MI (2000) Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn Sci 4:215–222.

Carlson MR, Zhang B, Fang Z, Mischel PS, Horvath S, Nelson SF (2006) Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. BMC Genomics 7:40.

Chen CH, Gutierrez ED, Thompson W, Panizzon MS, Jernigan TL, Eyler LT, Fennema-Notestine C, Jak AJ, Neale MC, Franz CE, Lyons MJ, Grant MD, Fischl B, Seidman LJ, Tsuang MT, Kremen WS, Dale AM (2012) Hierarchical genetic organization of human cortical surface area. Science 335:1634–1636.

Chiang MC, Dutton RA, Hayashi KM, Lopez OL, Aizenstein HJ, Toga AW, Becker JT, Thompson PM (2007) 3D pattern of brain atrophy in HIV/AIDS visualized using tensor-based morphometry. Neuroimage 34:44–60.

Chiang MC, Avedissian C, Barysheva M, Toga AW, McMahon KL, de Zubicaray GI, Wright MJ, Thompson PM (2009a) Extending genetic linkage analysis to diffusion tensor images to map single gene effects on brain fiber architecture. Med Image Comput Comput Assist Interv 12:506–513.

Chiang MC, Barysheva M, Shattuck DW, Lee AD, Madsen SK, Avedissian C, Klunder AD, Toga AW, McMahon KL, de Zubicaray GI, Wright MJ, Srivastava A, Balov N, Thompson PM (2009b) Genetics of brain fiber architecture and intellectual performance. J Neurosci 29:2212–2224.

Chiang MC, Barysheva M, Toga AW, Medland SE, Hansell NK, James MR, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM (2011) BDNF gene effects on brain circuitry replicated in 455 twins. Neuroimage 55:448–454.

Chumbley JR, Friston KJ (2009) False discovery rate revisited: FDR and topological inference using Gaussian random fields. Neuroimage 44:62–70.

Cohen R, Havlin S (2003) Scale-free networks are ultrasmall. Phys Rev Lett 90:058701.

Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, Ke X, Le Hellard S, Christoforou A, Luciano M, McGhee K, Lopez L, Gow AJ, Corley J, Redmond P, Fox HC, Haggarty P, Whalley LJ, McNeill G, Goddard ME, et al. (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry 16:996–1005.

DeVito LM, Konigsberg R, Lykken C, Sauvage M, Young WS 3rd, Eichenbaum H (2009) Vasopressin 1b receptor knock-out impairs memory for temporal order. J Neurosci 29:2676–2683.

de Zubicaray G, Chiang MC, McMahon KL, Shattuck DW, Toga AW, Martin NG, Wright MJ, Thompson PM (2008) Meeting the challenges of neuroimaging genetics. Brain Imaging Behav 2:258–263.

Duncan D, Prodduturi N, Zhang B (2010) WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. BMC Bioinformatics 11 [Suppl 4]:10.

Eguíluz VM, Chialvo DR, Cecchi GA, Baliki M, Apkarian AV (2005) Scale-free brain functional networks. Phys Rev Lett 94:018102.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863–14868.

Esaki T, Suzuki H, Cook M, Shimoji K, Cheng SY, Sokoloff L, Nunez J (2003) Functional activation of cerebral metabolism in mice with mutated thyroid hormone nuclear receptors. Endocrinology 144:4117–4122.

Gandhi S, Wood NW (2010) Genome-wide association studies: the key to unlocking neurodegeneration? Nat Neurosci 13:789–794.

Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, Brozell A, Schadt EE, Drake TA, Lusis AJ, Horvath S (2006) Integrating genetic and network analysis to characterize genes related to mouse weight. PLoS Genet 2:e130.

He XY, Schulz H, Yang SY (1998) A human brain L-3-hydroxyacyl-coenzyme A dehydrogenase is identical to an amyloid beta-peptide-binding protein involved in Alzheimer's disease. J Biol Chem 273:10741–10746.

Hibar DP, Stein JL, Kohannim O, Jahanshad N, Saykin AJ, Shen L, Kim S, Pankratz N, Foroud T, Huentelman MJ, Potkin SG, Jack CR Jr, Weiner MW, Toga AW, Thompson PM (2011) Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. Neuroimage 56:1875–1891.

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231.

Holmes CJ, Hoge R, Collins L, Woods R, Toga AW, Evans AC (1998) En-

hancement of MR images using registration for signal averaging. J Comput Assist Tomogr 22:324–333.

Hulshoff Pol HE, Schnack HG, Posthuma D, Mandl RC, Baaré WF, van Oel C, van Haren NE, Collins DL, Evans AC, Amunts K, Bürgel U, Zilles K, de Geus E, Boomsma DI, Kahn RS (2006) Genetic contributions to human brain morphology and intelligence. J Neurosci 26:10235–10242.

Humphries MD, Gurney K (2008) Network 'small-world-ness': a quantitative method for determining canonical network equivalence. PLoS One 3:e0002051.

Jackson DN (1984) MAB, multidimensional aptitude battery: manual. Port Hurton, MI: Research Psychologists.

Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. Med Image Anal 5:143–156.

Kamada T, Kawai S (1989) An algorithm for drawing general undirected graphs. Inform Process Lett 31:7–15.

Kravitz DJ, Saleem KS, Baker CI, Mishkin M (2011) A new neural framework for visuospatial processing. Nat Rev Neurosci 12:217–230.

Kwok SC, Buckley MJ (2010) Fornix transection selectively impairs fast learning of conditional visuospatial discriminations. Hippocampus 20:413–422.

Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11:241–247.

Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. Nature 445:168–176.

Lenroot RK, Schmitt JE, Ordaz SJ, Wallace GL, Neale MC, Lerch JP, Kendler KS, Evans AC, Giedd JN (2009) Differences in genetic and environmental influences on the human cerebral cortex associated with development during childhood and adolescence. Hum Brain Mapp 30:163–174.

Liang WS, Dunckley T, Beach TG, Grover A, Mastroeni D, Walker DG, Caselli RJ, Kukull WA, McKeel D, Morris JC, Hulette C, Schmechel D, Alexander GE, Reiman EM, Rogers J, Stephan DA (2007) Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain. Physiol Genomics 28:311–322.

Liu GG, Fong E, Zeng X (2010) GNCPro: navigate human genes and relationships through net-walking. Adv Exp Med Biol 680:253–259.

Luciano M, Wright MJ, Geffen GM, Geffen LB, Smith GA, Evans DM, Martin NG (2003) A genetic two-factor model of the covariation among a subset of Multidimensional Aptitude Battery and Wechsler Adult Intelligence Scale–Revised subtests. Intelligence 31:589–605.

Malo MS, Srivastava K, Andresen JM, Chen XN, Korenberg JR, Ingram VM (1994) Targeted gene walking by low stringency polymerase chain reaction: assignment of a putative human brain sodium channel gene (SCN3A) to chromosome 2q24-31. Proc Natl Acad Sci USA 91:2975–2979.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, et al. (2009) Finding the missing heritability of complex diseases. Nature 461:747–753.

Marchini J, Donnelly P, Cardon LR (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. Nat Genet 37:413–417.

Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. Science 296:910–913.

Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, Gordon SD, Ferreira MA, Wright MJ, Henders AK, Campbell MJ, Duffy DL, Hansell NK, Macgregor S, Slutske WS, Heath AC, Montgomery GW, Martin NG (2009) Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 85:750–755.

Neale MC, Cardon LR, the NATO Scientific Affairs Division (1992) Methodology for genetic studies of twins and families. Dordrecht; Boston: Kluwer Academic Publishers.

Pezawas L, Verchinski BA, Mattay VS, Callicott JH, Kolachana BS, Straub RE, Egan MF, Meyer-Lindenberg A, Weinberger DR (2004) The brain-derived neurotrophic factor val66met polymorphism and variation in human cortical morphology. J Neurosci 24:10099–10102.

Posner MI, Rothbart MK (2005) Influencing brain networks: implications for education. Trends Cogn Sci 9:99–103.

Posthuma D, de Geus EJ, Neale MC, Hulshoff Pol HE, Baaré WEC, Kahn RS,

Boomsma D (2000) Multivariate genetic analysis of brain structure in an extended twin design. Behav Genet 30:311–319.

Prata DP, Mechelli A, Fu CH, Picchioni M, Toulopoulou T, Bramon E, Walshe M, Murray RM, Collier DA, McGuire P (2009) Epistasis between the *DAT* 3′ UTR VNTR and the *COMT* Val158Met SNP on cortical function in healthy subjects and patients with schizophrenia. Proc Natl Acad Sci USA 106:13600–13605.

Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297:1551–1555.

Reed JE, Dunn JR, du Plessis DG, Shaw EJ, Reeves P, Gee AL, Warnke PC, Sellar GC, Moss DJ, Walker C (2007) Expression of cellular adhesion molecule 'OPCML' is down-regulated in gliomas and other brain tumours. Neuropathol Appl Neurobiol 33:77–85.

Reichmann H, Maltese WA, DeVivo DC (1988) Enzymes of fatty acid beta-oxidation in developing brain. J Neurochem 51:339–344.

Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM (2002) Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. Cancer Res 62:4427–4433.

Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52:1059–1069.

Schork NJ, Murray SS, Frazer KA, Topol EJ (2009) Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev 19:212–219.

Schuff N, Woerner N, Boreta L, Kornfield T, Shaw LM, Trojanowski JQ, Thompson PM, Jack CR Jr, Weiner MW (2009) MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain 132:1067–1077.

Smith SM (2002) Fast robust automated brain extraction. Hum Brain Mapp 17:143–155.

Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TE (2006) Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31:1487–1505.

Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, Saykin AJ, Shen L, Foroud T, Pankratz N, Huentelman MJ, Craig DW, Gerber JD, Allen AN, Corneveaux JJ, Dechairo BM, Potkin SG, Weiner MW, Thompson P (2010) Voxelwise genome-wide association study (vGWAS). Neuroimage 53:1160–1174.

Torkamani A, Dean B, Schork NJ, Thomas EA (2010) Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. Genome Res 20:403–412.

Tuch DS, Salat DH, Wisco JJ, Zaleta AK, Hevelone ND, Rosas HD (2005) Choice reaction time performance correlates with diffusion anisotropy in white matter pathways supporting visuospatial attention. Proc Natl Acad Sci USA 102:12212–12217.

Vestergaard M, Madsen KS, Baare WF, Skimminge A, Ejersbo LR, Ramsøy TZ, Gerlach C, Akeson P, Paulson OB, Jernigan TL (2011) White matter microstructure in superior longitudinal fasciculus associated with spatial working memory performance in children. J Cogn Neurosci 23:2135–2146.

Visscher PM (2008) Sizing up human height variation. Nat Genet 40:489–490.

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393:440–442.

Wright MJ, Martin NG (2004) Brisbane adolescent twin study: outline of study methods and research projects. Aust J Psychol 56:65–78.

Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol 4:Article17.

Zhang B, Kirov S, Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res 33: W741–W748.