

Published in final edited form as:

Neuroimage. 2014 August 15; 97: 252–261. doi:10.1016/j.neuroimage.2014.04.021.

## Common genetic variants and gene expression associated with white matter microstructure in the human brain

Emma Sprooten, PhD<sup>1,2</sup>, Emma E Knowles, PhD<sup>1,2</sup>, D Reese McKay, PhD<sup>1,2</sup>, Harald HH Göring, PhD<sup>3</sup>, Joanne E Curran, PhD<sup>3</sup>, Jack W Kent Jr, PhD<sup>3</sup>, Melanie A Carless, PhD<sup>3</sup>, Thomas D Dyer, PhD<sup>3</sup>, Eugene I Drigalenko<sup>3</sup>, Rene L Olvera, MD<sup>4</sup>, Peter T Fox, MD<sup>5,6</sup>, Laura Almasy, PhD<sup>3</sup>, Ravi Duggirala, PhD<sup>3</sup>, Peter Kochunov, PhD<sup>7</sup>, John Blangero, PhD<sup>3</sup>, and David C Glahn, PhD<sup>1,2</sup>

<sup>1</sup>Department of Psychiatry, Yale University School of Medicine, 300 George Street, New Haven, CT, USA

<sup>2</sup>Olin Neuropsychiatry Research Center, Institute of Living, Hartford Hospital, 200 Retreat Avenue, CT, USA

<sup>3</sup>Department of Genetics, Texas Biomedical Research Institute, PO Box 760549, San Antonio, TX, USA

<sup>4</sup>Department of Psychiatry, University of Texas Health Science Center San Antonio, 7703 Floyd Curl Drive, San Antonio, TX, USA

<sup>5</sup>Research Imaging Institute, University of Texas Health Science Center San Antonio, 8403 Floyd Curl Drive, San Antonio, TX, USA

<sup>6</sup>South Texas Veterans Health System, 7400 Merton Minter, San Antonio, Texas 78229

<sup>7</sup>Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, USA

### Abstract

Identifying genes that contribute to white matter microstructure should provide insights into the neurobiological processes that regulate white matter development, plasticity and pathology. We detected five significant SNPs using genome-wide association analysis on a global measure of fractional anisotropy in 776 individuals from large extended pedigrees. Genetic correlations and genome-wide association results indicated that the genetic signal was largely homogeneous across white matter regions. Using RNA transcripts derived from lymphocytes in the same individuals, we identified two genes (*GNAI3* and *CCDC91*) that are likely to be *cis*-regulated by top SNPs, and whose expression levels were also genetically correlated with fractional anisotropy. A

---

© Published by Elsevier Inc. All rights reserved.

**Corresponding Author:** Emma Sprooten Ph.D., Olin Neuropsychiatry Research Center, Institute of Living/Hartford Hospital, 200 Retreat Avenue, CT, 06106, USA; emma.sprooten@yale.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Conflicts of Interest:** The authors declare no competing financial interests.

transcript of *HTR7* was phenotypically associated with FA, and was associated with an intronic genome-wide significant SNP. These results encourage further research in the mechanisms by which *GNA13*, *HTR7* and *CCDC91* influence brain structure, and emphasize a role for g-protein signaling in the development and maintenance of white matter microstructure in health and disease.

## Keywords

Diffusion tensor imaging; white matter; genome-wide association; gene expression; transcripts; extended pedigrees

---

## 1. Introduction

Studies in twins and extended pedigrees have established that white matter microstructure, as measured *in vivo* by diffusion tensor imaging (DTI), is heritable (Chiang et al., 2011b; Jahanshad et al., 2013a; Kochunov et al., 2010). However, the genetic variants contributing to this heritability are unknown and little is understood about the mechanisms that govern the development, maintenance, plasticity and pathology of white matter microstructure. White matter plays an important role in several neurological diseases (Stebbins and Murphy, 2009) and psychiatric disorders (Kubicki et al., 2007; Mahon et al., 2010), which are phenotypes that also have substantial but poorly characterized, genetic components. There is increasing evidence that compromised white matter microstructure is part of the inherited risk for these disorders, as indicated by reduced FA in unaffected relatives (Gold et al., 2012; Hoptman et al., 2008; Sprooten et al., 2013a; Sprooten et al., 2011a), and polygenic risk score analysis (Whalley et al., 2013). Therefore, identifying genes that influence white-matter microstructure could provide a biological anchor for disentangling basic molecular mechanisms that predispose to these debilitating disorders, potentially leading to novel treatment agents and prevention strategies.

DTI is a magnetic resonance imaging technique that is based on the orientation and magnitude of the motion of water molecules, and its restriction by surrounding tissue. Because of the parallel alignment of white matter fibers that restrict motion primarily in directions perpendicular to the fibers, DTI is ideally suited to measure properties of white matter microstructure (Beaulieu, 2002). Fractional anisotropy (FA) is an index of the extent to which this motion is directionally constrained and, as validated in animal (Li et al., 2011) and post-mortem research (Schmierer et al., 2007), it reflects a combination of myelin thickness, fiber coherence and axon integrity. Studies using *a priori* selected candidate genes and SNPs have associated FA with genetic variation in *NRG1* (McIntosh et al., 2008; Sprooten et al., 2009; Winterer et al., 2008), *ErbB4* (Konrad et al., 2009; Zuliani et al., 2011), *DISC1* (Sprooten et al., 2011b), *NTRK1* (Braskie et al., 2012), *BDNF* (Chiang et al., 2011a) and *APOE* (Jahanshad et al., 2012), amongst others. However, FA is a complex, polygenic phenotype and for most complex phenotypes data-driven GWA have not implicated *a priori* candidate variants in their top results (Flint and Munafò, 2013; Stein et al., 2012), hence many more novel SNP-associations contributing to variation in FA could be discovered using GWA.

Numerous common variants correlated with complex disease risks have been reported using GWA (Hindorff et al., 2009; Hirschhorn and Daly, 2005; Ripke et al., 2013), but the effect size of individual common variants on complex phenotypes tend to be small (Flint and Munafo, 2013; Hindorff et al., 2009). Significant genome-wide association reflects the presence of a relevant functional variant in the surrounding genomic region and thus is indicative of causal gene localization but not the identification of underlying biological mechanism, which is the ultimate goal of complex disease genetics. It is difficult to infer a specific gene's involvement in trait variance solely based upon a statistically significant association, since the polymorphisms tagged in GWA rarely influence gene function directly and the effect of a tagging SNP reflects, in addition to the effect that it exerts, the effects of all SNPs within the surrounding linkage disequilibrium (LD) block, which may span many genes any one (or combination) of which could be driving the observed association. Examining complementary biological information, such as RNA expression, can refine inferences made from GWA and identify potential genes through which the associated SNPs are likely to exert their effect.

In the present paper, we aim to characterize the common variation contributing to the genetics of white matter microstructure. Firstly, to identify common variants affecting white matter microstructure we performed GWA of a global FA measure in a sample of 776 Mexican-American members of extended pedigrees. Secondly, to propose genes that may be responsible for the effects of these common variants, we correlated lymphocyte-derived RNA transcripts of nearby genes both with genetic variation in the genome-wide significant SNPs and with white matter microstructure.

We chose a global index of white matter microstructure, namely average FA across the white matter skeleton derived from tract-based spatial statistics (TBSS), as our primary phenotype. This phenotype was previously shown to be heritable in multiple cohorts (Jahanshad et al., 2013a), and relevant to genetic risk for bipolar disorder (Sprooten et al., 2013a; Sprooten et al., 2011a) and major depressive disorder (Whalley et al., 2013). To examine the neuroanatomical specificity of genetic effects on FA, we performed voxel-wise analyses for our top SNPs, and used the family structure and additional GWA results of regional FA to investigate the degree to which additive genetic effects across the genome are anatomically homogeneous throughout the brain.

## 2. Materials and Methods

### 2.1 Participants

Participants were individuals of Mexican American ancestry who took part in the Genetics of Brain Structure and Function Study (GOBS) (Olvera et al., 2011), which is an extension of the San Antonio Family study (Mitchell et al., 1996; Puppala et al., 2006). Individuals in this cohort have actively participated in research for over 18 years and were randomly selected from the community with the only constraints that they are of Mexican-American ancestry, part of a large family and live within the San Antonio region. For the present analysis, subjects were excluded if they had MRI contraindications, documented history of neurological illness, or any major neurological event visible on the structural T1 or DTI scans (see also section 2.3 below). Of the participants in the final analysis, 16 self-reported

history of a neurological event, but excluding these individuals did not change the pattern of results presented. Individuals with a history of psychiatric illness were included in the main analyses (psychosis: N=18, bipolar disorder: N=8, anxiety disorder: N=162, major depression: N=249), but all results remained significant at similar levels when covarying for history of these psychiatric diagnoses. Additional recruitment and demographic details of the sample are available elsewhere (Glahn et al., 2012; Olvera et al., 2011).

After thorough data quality checks (see section 2.3), 776 individuals contributed both high-quality DTI scans and blood samples. This sample included 65 pedigrees, varying in size from 2 to 82 individuals, and 49 unrelated individuals. Sixty percent (N = 466) of the participants were female and their age ranged from 18 to 85 years (mean 44.37). All participants provided written informed consent on form approved by the IRBs at Yale University and the University of Texas Health Science Center in San Antonio (UTHSCSA).

## 2.2 Diffusion tensor imaging acquisition

Diffusion imaging data were acquired on a Siemens 3T Trio scanner located at the Research Imaging Institute, UTHSCSA (Kochunov et al., 2010; Kochunov et al., 2007). A single-shot, echo-planar, single refocusing spin-echo weighted sequence with TE/TR=87/8000ms was applied to obtain 50 contiguous axial slices field-of-view=200 × 200mm, thickness 3mm, resolution 1.7mm × 1.7mm) in 55 isotropically distributed diffusion weighted directions ( $b=0$  and  $b=700\text{s/mm}^2$ ). Three non-diffusion weighted ( $b=0$ ) images were acquired at the start of each sequence.

## 2.3 DTI processing and extraction of FA values

Diffusion imaging data were converted to NIFTI format using MRICron and pre-processed using standard tools available in FSL (<http://www.fmrib.ox.ac.uk/fsl>) (Behrens et al., 2003). First, for each subject, the 55 diffusion-weighted images were corrected for subject motion and eddy currents by linearly aligning them to the first  $b_0$  volume within each individual. Next, non-brain tissue was removed using the brain extraction tool (“bet”) with a fractional intensity threshold of 0.1. Then, diffusion eigenvectors, eigenvalues and FA were calculated for each voxel using “dtifit”.

TBSS (Smith et al., 2006), a standardized and validated procedure for DTI analysis tailored to white matter anatomy, was applied to extract a global measure of whole-brain white matter microstructure. First FA images were slightly eroded and non-linearly registered to standard MNI space. FA maps were masked with a study-specific mask, which was created such that only voxels with  $\text{FA} > 0$  in every participant were included.

The average FA map was created and fed into the TBSS skeletonization algorithm, and thresholded at  $\text{FA} > 0.2$ , resulting in a sample-specific skeleton template (Fig. 1). Next, within each individual FA map the nearest maximum FA voxel was projected onto the skeleton template, to obtain one skeleton image per subject, representing the central structure of each subject’s white matter. The mean FA within this skeleton was extracted for each individual, hereafter referred to as “global FA”.

The FA maps were visually inspected before and after registration to the MNI template. From the total of 796 DTI scans, 20 subjects were excluded from further analysis (7 for anatomical abnormalities, 5 for scanner and 8 for registration artifacts).

In addition to global FA, 31 regional FA values were derived from the TBSS skeletons based on the John Hopkins University White Matter Labels Atlas, which is integrated in FSL. FA was extracted for all regions within this atlas, apart from for the uncinate fasciculi, the fornix and the cortico-spinal tracts because of incomplete coverage by the atlas and/or previous indications of low and unreliable heritability estimates in these regions (Jahanshad et al.; 2013a). Masks derived from the atlas were smoothed with a kernel of 1.1 mm FWHM, thresholded at  $>0.2$  and binarized. Accurate overlap with the TBSS skeleton was visually verified for each region. Each mask was multiplied by the TBSS skeleton mask to obtain a parcellation of the center of each region-of-interest and mean FA values within each region were extracted for each subject.

## 2.4 Heritability estimation

All statistics were calculated in SOLAR (Almasy and Blangero, 1998), which applies maximum likelihood methods to decompose the variance of a trait into genetic and environmental components by modeling the covariance between individuals as a function of their genetic proximity. More specifically, to estimate heritability the covariance matrix was decomposed into the following: the observed additive coefficient of relationship matrix (reflecting genetic proximity), a set of covariates (sex, age, age<sup>2</sup>, age x sex, age<sup>2</sup> x sex), and a random environment component. The significance of each variance component is assessed by a chi-squared test comparing the final model to the model without the component of interest. This method was applied to calculate heritability for global FA and for each of the regions of interests in separate models. As SOLAR assumes a multivariate normal distribution, FA measures were transformed using an inverse normalization transformation prior to running variance decomposition analysis.

## 2.5 Genotyping and genome-wide association

DNA was extracted from lymphocytes and genotyped using either the Illumina Human1M-Duo Beadchip or the Illumina HumanHap550 BeadChip in tandem with the supplemental HumanHap450S Beadchip. Both methods result in  $\sim 1$  million tagging SNPs, capturing approximately 90% of the common variation in humans. Genotype data were checked for accordance with Mendelian consistency using SimWalk2 (Sobel and Lange, 1996). Missing genotypes were imputed using a procedure specifically developed for extended pedigrees in MERLIN (Burdick et al., 2006). After excluding the 2407 SNPs that deviated significantly from the Hardy-Weinberg equilibrium ( $X^2 > 10$ ), SNPs with a call rate  $< 95\%$  and 820 SNPs with a minor allele frequency  $< 0.001$ , a total of 929,187 SNPs remained for further analysis.

For genome-wide association analysis of global FA, each SNP dosage was coded as 0, 1 or 2, or for missing genotypes as the weighted value based on the imputation. Each SNP is added as a “covariate” to the heritability model described above, one at a time and assuming an additive effect of alleles. The significance of each SNP is adjusted for family structure by comparing the null model without the SNP, but with the coefficient of relationship matrix

and the covariates, to the model of interest with the SNP. To account for population stratification, the first four principal components of the genome-wide SNP data were also added to this model. The fast exact likelihood calculation method described by Blangero *et al.* (Blangero et al., 2013) was employed for all analyses. P-values  $< 5 \times 10^{-8}$  were considered genome-wide significant (Pe'er et al., 2008). The genotypes of the five top SNPs were checked in the original Illumina cluster files, and there were no obvious artifacts at these specific loci.

## 2.6 Transcript processing and analysis

Whole-genome expression profiles were derived from untransformed lymphocytes which were collected during the clinic visit when the MRI scans were taken. The methods used to generate gene expression levels are similar to those described in detail elsewhere (Goring et al., 2007). In brief, lymphocytes were isolated from whole-blood samples that were collected in the morning after a night of fasting, and stored in liquid nitrogen. Lymphocyte RNA was isolated and anti-sense RNA was synthesized, amplified, purified and hybridized to Illumina BeadChips version 3 according to standard protocols. 647 samples (including controls and duplicates) were processed in a single batch and analyzed using Illumina BeadScan and BeadStudio software. The samples were evaluated based on the average number of probes with significant expression (at a detection p-value = 0.05; see below), mean raw “average signal” across probes, and average correlation (across probes) against all other samples. 595 high quality, unique samples were identified. Of these, 360 individuals also had high-quality DTI and GWA data (i.e. overlapped with the sample on which we did the GWA). Significantly expressed probes were identified as follows: (1) for each probe, samples were classified as significantly or not significantly expressed at a “detection p-value” (which is defined by the Illumina software based on comparisons with null probes) of  $p \leq 0.05$ ; (2) for each probe, binomial tests were conducted on the number of samples in which the probes was “detected” or “not detected”; and (3) finally, considering all p-values resulting from the binomial tests for all probes, probes were accepted to be significantly expressed at a 5% false-discovery rate (FDR). As a result, 25,663 probes were significantly expressed above the baseline levels at 5% FDR. Subsequently, expression levels were processed by background noise subtraction (shifting values upwards so that the minimum raw “average signal” for a probe in a sample was 1), log<sub>2</sub> transformation, following by quantile normalization. Outliers of expression levels ( $> 3$  standard deviations from the mean) were removed before further analyses (between 0 and 9 cases per transcript).

To investigate whether any of the top SNPs (or SNPs in LD with them) could potentially have an effect on the function of a nearby gene, we entered the standardized, normalized expression levels of genes within a 500kb window (according to the NCBI SNP database) as traits in a polygenic model in SOLAR, with SNP as a covariate. The resulting p-values were corrected for multiple comparisons across transcripts at a false-discovery rate of 5%, and those that were significant were further investigated for any associations with FA. To quantify this relationship between FA and expression levels, the global FA measure and the gene’s transcript levels were both entered as traits in bivariate polygenic models in SOLAR to calculate their genetic correlation. In these models, the pedigree structure is used to

decompose the phenotypic correlation ( $\rho_p$ ) between two traits (i and j) into genetic ( $\rho_g$ ) and environmental correlations ( $\rho_e$ ) so that

$$\rho_p = \rho_g \sqrt{(h^2_j * h^2_i)} + \rho_e \sqrt{((1 - h^2_j)(1 - h^2_i))}$$

while the  $\rho_p$  quantifies the overall relationship between two traits,  $\rho_g$  represents the amount of genetic variance shared between two traits, or pleiotropy (Almasy et al., 1997; Williams et al., 1999). These parameters were estimated in polygenic models in SOLAR using all available pedigree information, and their significance was tested by comparing the log-likelihood of this polygenic model against the log-likelihood of the null model (where  $\rho_p$  and  $\rho_g$  are respectively constrained equal to 0). The resulting likelihood ratio test is distributed as a chi-square with a single degree of freedom. As before, prior to running these models, FA and transcriptional measures were normalized using an inverse Gaussian transformation, and for all polygenic models, age, sex, age squared and their interactions were included as (additional) covariates.

The Allen Human Brain Atlas (Hawrylycz et al., 2012) (<http://human.brain-map.org/>) was utilized to examine the expression patterns of genes that were significantly genetically correlated with FA in human white matter. For each gene, we downloaded the z-scores of all probes, of all available individuals for the corpus callosum, the cingulum and the central glial system. Since the transcript data in the Allen Brain Atlas are normalized within each batch, and within each subject across the brain, this z-score reflects the number of standard deviations away from the overall mean expression across probes, across brain regions, thus indicating the relative local preference of the gene's expression across brain regions. The Allen Brain Atlas contained six samples of the corpus callosum, four of the central glial substance and one of the cingulum bundle.

## 2.7 Anatomic uniformity of genetic signal across white matter regions

The question, whether the genetic influence is uniform across the brain's white matter locations, was approached in several ways. Firstly, explore the extent to which the effects of genome-wide significant SNPs were regionally specific, additional GWAs were performed on FA values in each of the 119,563 voxels within the TBSS skeleton, using the same model as above, for the genome-wide significant SNPs only.

Secondly, extended pedigrees enable the computation of genetic correlations (Almasy et al., 1997), which are a powerful way to quantify the extent to which two traits are influenced by the same additive genetic variation (also see 2.6 above). In contrast to genome-wide association studies, this method is most sensitive to rare genetic variants within pedigrees, rather than to common variation across unrelated individuals. Substantial heritability was regarded as a pre-requisite for the presence of genetic correlations, so only traits with  $h^2 > 0.2$  were considered for further analysis. Next, genetic correlations between global FA and FA within each of the regions were calculated by adding both as traits in a polygenic model in SOLAR, so that the covariance matrix between the two traits is decomposed into genetic and environmental correlations, in addition to covariance due to each of the covariates (age, sex, sex\*age, age<sup>2</sup> and sex\*age<sup>2</sup>).

Thirdly, to be more in line with GWAS's sensitivity for common genetic variation, we computed Spearman's rank correlation coefficients between the betas of global FA, and the betas of mean FA within each of the tracts. To minimize the influence of extreme outliers, beta values lower than  $-3$  and higher than  $3$  (6 for global FA, 1 for the left ALIC, 2 for the right posterior corona radiata, 3 for the right superior cerebellar peduncle, and 1 for the left UF) were excluded for this analysis.

Furthermore, because the majority of SNPs across the whole genome are not strongly associated with the phenotype, they can be assumed to randomly vary around 0. For this reason we tested whether SNPs significantly associated with global FA (according to five thresholds:  $p < 0.1$ ,  $p < 0.05$ ,  $p < 0.01$ ,  $p < 0.005$ ,  $p < 0.001$ ) were also more likely to be nominally associated with mean FA in the separate regions. To this end, chi-squared tests were performed between SNPs significant in the global FA GWAS and in each of the sub-region GWASs.

To consider the reverse scenario – to what extent is FA in each of the regions determined by independent genetic variation – polygenic models were run for each region including global FA as a covariate. This estimates the heritability of the variance unique to each region.

### 3. Results

#### 3.1 Heritability of FA

Consistent with previous reports including a subset of our data (Jahanshad et al., 2013a; Kochunov et al., 2010), the global FA measure was significantly heritable with  $h^2 = 52\%$  ( $p = 8.66 \times 10^{-11}$ ). In the same model, age co-varied significantly with global FA ( $X^2 = 141.64$ ;  $p = 1.17 \times 10^{-32}$ ) as did sex but to a lesser extent ( $X^2 = 4.05$ ;  $p = 0.04$ ), while neither age<sup>2</sup> nor any of the interactions were significant. Together the covariates explained 31% of the variance in FA. Re-running the model without the non-significant higher-order covariates resulted in an almost identical heritability estimate.

Similarly, average FA values for all tracts were significantly and substantially heritable (all  $h^2 > 0.26$ , all  $p < 0.0012$ ). Again, in line with previous reports in our sample (Jahanshad et al., 2013a; Kochunov et al., 2010), the highest heritability estimates was found for the right superior longitudinal fasciculus, with  $h^2 = 0.65$  ( $p = 8.49 \times 10^{-14}$ ) and the splenium of the corpus callosum ( $h^2 = 0.63$ ;  $p = 2.9 \times 10^{-14}$ ). Any slight differences in these estimates between the two analyses are due to the use of a sample-specific template in the present study, and a small reduction in sample size because of missing genotyping data. See supplementary Table S1 for the heritability estimates of FA in each tract separately.

Figure 1 shows the results of a voxel-wise analysis of the heritability of FA. In line with the tract-wise analysis, the highest heritability was found for voxels in the splenium and genu of the corpus callosum, the superior longitudinal fasciculi and in the internal and external capsules. In 58,202 (49%) of the voxels, the heritability estimate of FA was greater than 0.20, and in 7% (8,600 voxels) the heritability estimate was 0.50 or greater.



### 3.2 Genome-wide significant SNPs associated with global FA

GWA analysis of global FA was conducted in SOLAR, taking into account pedigree structure, age, sex, and four multi-dimensional scaling (MDS) components to account for population stratification. The genomic inflation factor was  $\lambda=1.03$ , indicating that any remaining population stratification effects after MDS regression, were negligible. The quantile-quantile (QQ) plot showed that p-values below  $p = 10^{-4}$  were more significant than predicted by chance (Fig. 2). Five SNPs were associated with global FA at the genome-wide significant level ( $p < 5 \times 10^{-8}$ ; Table 1, Fig. 3).

Four of the five genome-wide significant SNPs were intergenic (Table 1), while rs12249377 was located in an intron of serotonin receptor gene *HTR7*. SNP rs10853057 is located 7kb away from a g-protein coupled receptor gene, *GNA13*. The other SNPs were further than 50kb from any genes. Scatterplots showing the effects of these SNPs on FA per individual, and mean FA values grouped per genotype, are available in supplementary Figure S1. All five SNPs remained genome-wide significant after covarying for history of major depression, bipolar disorder, anxiety disorder, and psychosis (all  $p < 10^{-9}$ ); and after covarying for within-scanner motion (total amount translation and total rotation parameters; all  $p < 5 \times 10^{-8}$ ).

Voxel-wise associations revealed that, for all five top SNPs, the majority of voxels in the TBSS skeleton contributed to the genome-wide signal of global FA: the number of voxels with the same direction of effect (the same allele being associated with increased FA) ranged from 89% (rs1361277) to 91% (rs10853057) of voxels in the white matter skeleton. The voxel-wise results of SNPs rs10853057 and rs258415 are shown in Figure 4; other voxel-wise maps are available in supplementary Figure S2. For all SNPs, the most significantly influenced tracts with p-values  $< 0.001$ , were the genu and splenium of the corpus callosum, the internal capsules, parts of the superior longitudinal fasciculus, and parahippocampal white matter.

### 3.3 Gene expression

To infer which genes might be considered responsible for the effects of the genome-wide significant SNPs, we examined lymphocyte-derived gene expression profiles of genes within a 500kb interval on either side of each tag SNPs. Transcript data were available for eleven nearby genes, corresponding to 15 Illumina transcript probes (Table 2). Nine of these transcripts were significantly associated with their respective SNPs (Table 2) after correcting for the total number of transcripts tested (FDR=0.05).

Correlations with FA are reported in Table 3. While six transcripts were significantly phenotypically correlated with global FA, two transcripts, one of *GNA13* and one of *CCDC91*, showed significant genetic correlations, providing circumstantial evidence that these genes influence global FA. Transcripts of *HTR7*, *PCGF5* and *SMURF2* tended to be less heritable and showed significant phenotypic, but not genotypic, correlations with global FA, suggesting lower measurement reliability and/or shared environmental influences. Both the associations with SNPs and the genetic correlations with FA remained similar when

covarying for history of major psychiatric disorders; and after covarying for 6 parameters of within scanner movement (3 for translation and 3 for rotation).

For each of *CCDC91* and *GNAI3*, transcript data were available for 3 probes.<sup>1</sup> Expression of *GNAI3* was above average for all three probes in the corpus callosum (means across 6 samples:  $z_{\text{probe1}}=2.44$ ;  $z_{\text{probe2}}=1.74$ ,  $z_{\text{probe3}}=1.22$ ), in the cingulum bundle ( $z_{\text{probe1}}=2.93$ ;  $z_{\text{probe2}}=1.99$ ,  $z_{\text{probe3}}=1.78$ ), and in the central glial substance (means across 4 samples:  $z_{\text{probe1}}=2.68$ ;  $z_{\text{probe2}}=2.00$ ,  $z_{\text{probe3}}=2.32$ ). *CCDC91* is highly expressed in the cingulum bundle ( $z_{\text{probe1}}=2.17$ ;  $z_{\text{probe2}}=1.72$ ,  $z_{\text{probe3}}=2.49$ ), but did not demonstrate a clear preference for the corpus callosum or the cingulum bundles (mean z ranging from  $-0.04$  to  $0.70$  and from  $-0.44$  to  $1.18$ ; respectively).

### 3.4 Homogeneity of additive genetic effects across white matter regions

Supplementary Table S2 contains an overview of the statistics of the homogeneity of genetic effects across white matter regions derived from the Johns Hopkins University Labels atlas.

Phenotypic correlations with global FA were very high, between 0.48 and 0.82. Genetic correlations varied between 0.38 (for the left tapetum) and 0.94 (for the left anterior limb of the internal capsule and the corona radiata bilaterally). Of note, 29 of the 31 sub-regions had a genetic correlation of 0.50 or higher with the global FA measure. Similarly, when including global FA as a covariate in polygenic models for each of the tracts, global FA explained between 20% (left tapetum) and 61% (splenium of corpus callosum and left anterior limb of internal capsule) of the variance in each of the sub-regions.

Spearman's rank correlation coefficients between the betas resulting from the GWAs for global FA and for each of the sub-regions were all very high: between 0.42 (left tapetum) and 0.84 (splenium of corpus callosum). Hence the ranks of the betas of global FA explained between 18% and 70% of the rankings of the regional betas. All these correlations were highly significant at  $p < 2.2 \cdot 10^{-16}$ , but it should be noted that such genome-wide correlations are statistically over-powered because of the large number of measurement points across the genome.

In the five analyses of statistical enrichment for nominally significant SNPs of the global FA GWA within the sets of nominally significant SNPs of each of the regional GWAs ( $p < 0.1$ ,  $p < 0.05$ ,  $p < 0.01$ ,  $p < 0.005$ ,  $p < 0.001$ ), chi squared statistics ranged from 949 for the left tapetum (at  $p < 0.001$ ) to 250,452 for FA in the splenium (at  $p < 0.1$ ), and all were highly significant (all  $p < 2.2 \cdot 10^{-16}$ ).

Despite this high dependence of the regional measures on global FA, co-varying for global FA in the polygenic models revealed that the unique variance of most regions is also heritable. With  $h^2$  between 0.21 and 0.70 all the regions remained significantly heritable after statistically correcting for global FA. What is more, for some regions the heritability

<sup>1</sup>These data can be downloaded and visualized at the following 2 links, For GNAI3 and CCDC91 respectively: [http://human.brain-map.org/microarray/search/show?exact\\_match=false&search\\_term=GNAI3&search\\_type=gene](http://human.brain-map.org/microarray/search/show?exact_match=false&search_term=GNAI3&search_type=gene) [http://human.brain-map.org/microarray/search/show?exact\\_match=false&search\\_term=ccdc91&search\\_type=gene](http://human.brain-map.org/microarray/search/show?exact_match=false&search_term=ccdc91&search_type=gene)

estimate increased, indicating a significant amount of genetic variation is affecting those regions specifically.

#### 4. Discussion

We localized five genome-wide significant SNPs for global FA, indicating that variation in FA is sensitive to common genetic variation. Three of these SNPs were associated with nine lymphocyte-derived RNA expression levels from nearby genes, indicating that these variants (or variants in close LD) may regulate the expression of these transcripts. Transcript levels of *GNAI3* and *CCDC91* were significantly genetically correlated with global FA, suggesting that the protein products of these genes are involved in processes governing white matter microstructure.

Previous studies using *a priori* selected candidate genes and SNPs have associated FA with genetic variation in *DISCI* (Sprooten et al., 2011b), and *BDNF* (Chiang et al., 2011a), amongst others. However, white matter microstructure is a complex trait with a polygenic genetic architecture. To date, few GWA have been performed on DTI-derived traits (Chiang et al., 2012; Jahanshad et al., 2013b; Lopez et al., 2012; Sprooten et al., 2013b). Jahanshad et al., (2013b) found a genome-wide significant SNP significant (in *SPONI*) in association with the number of tractography streamlines between the left posterior cingulate gyrus and superior parietal lobe (Jahanshad et al., 2013b). In the same sample, Chiang et al. (2012) detected 24 SNPs in association with extracted FA from 18 regions of interest, including variants in genes involved in lipid metabolism (*LPIN2*, *HADH*) and cell adhesion (*OPCML*, *KAZN*).

Our top-associated SNP rs10853057 is located only 7kb downstream from the gene encoding the g-protein alpha 13 subunit (*GNAI3*). This SNP lies within in a DNase I hypersensitivity region and transcription factor binding site (<http://genome.ucsc.edu/cgi-bin/hgGateway>; (Consortium et al., 2012), indicating high regulatory potential. This supports our finding that *GNAI3* transcript levels were significantly associated with both our top SNP and with global FA suggesting that the g-protein alpha 13 subunit mediates an effect of rs10853057 on white matter development or maintenance. There are several conceivable mechanisms through which variation in *GNAI3* could affect white matter microstructure. For example, the g-proteins defined by the alpha 13 and 12 subunits interact directly with RhoGEFs, which in turn interact with RhoA (Kvachnina et al., 2005; Ponimaskin et al., 2002; Siehler, 2009), which is thought to regulate myelination (Feltri et al., 2008) and axon growth (Gross et al., 2007). From a broader perspective, Rho-proteins serve multiple other functions including focal adhesion, and cytoskeleton repair and maintenance (Dhanasekaran and Dermott, 1996), that could affect white matter organization.

The second strongest genome-wide association, rs12249377, was located in an intron of the serotonin receptor 7 gene (*HTR7*). We also found a significant association between rs12249377 and the expression of one of the *HTR7* transcripts suggesting that this marker may be detecting cis-acting regulatory effects upon the *HTR7* gene. However, there was only a phenotypic – but no genotypic – correlation between the transcript and FA, suggesting that *HTR7* expression influences white matter microstructure, but it is unclear

whether this effect is driven by secondary genetic or environmental factors (for example circadian rhythms). Our observation may be due to our lack of power to detect genetic correlation between *HTR7* expression levels and global FA due to the relatively low heritability of *HTR7* gene expression in lymphocytes. The serotonin system has been a major interest in behavioral neuroscience and neuropharmacology, and observations of knock-out animals and experiments with specific antagonists have demonstrated that *HTR7* is involved in mood regulation, circadian rhythms, and depressive behavior (Hedlund and Sutcliffe, 2004). Interestingly, *HTR7* couples to g-protein alpha s and alpha 12 subunits, the amino acid sequence of the latter being very similar to the alpha 13 subunit, with a similar potential to activate RhoA proteins (Kvachnina et al., 2005). Thus, *HTR7* and *GNA13* share their involvement in a molecular pathway that is critical for axon growth and myelination.

FA-associated SNP rs258415 was significantly associated with the expression of a gene 393kb away, namely *CCDC91*, which was in turn also genetically correlated with FA. Little is known about the function of *CCDC91*, although several studies indicate that it plays a role in brain structure and pathology. It is highly expressed in the central nervous system and it is involved in sub-cellular trafficking and localization of proteins to the Golgi complex. One study showed that there is an increased incidence of copy-number variants in *CCDC1* in patients with bipolar disorder (Chen et al., 2010). Also, its interaction partners, *GGA1* and *GGA2*, have been linked to the Alzheimer pathophysiology by interacting with beta-amyloid precursors (Herskowitz et al., 2012).

It may be surprising that our top SNPs do not implicate existing candidate genes for white matter structure specifically. None of the genes identified in the present paper would have been hypothesized as being associated with FA *a priori*. However, a review of other GWA studies of complex traits suggests that firstly, the absence of candidate genes in the results is the rule rather than the exception (Flint and Munafò, 2013; Seifuddin et al., 2012), and secondly the representation of intergenic regions in the results has been unexpectedly large (Hindorff et al., 2009). These two observations from GWA studies are likely due to several factors. Firstly, the polygenic nature of complex traits and the extremely high statistical penalty for multiple testing across the genome undoubtedly leads to a large number of unreported false negative SNPs, which likely include some previously hypothesized variants in candidate genes. Secondly, candidate genes are often based on animal experiments or linkage studies which are more suited to detecting rare variation of large effects. Third, the *a priori* information is incomplete, with many poorly characterized genes, regions transcribing non-coding RNAs that can regulate expression (Maurano et al., 2012), and epigenetic effects of chromatin structure (Mitchell et al., 2013).

A well-known, but important limitation of GWA is that it is not sensitive to rare genetic variation. In addition, replication of top GWA hits across studies has proven difficult, and even more so across populations (Ioannidis, 2007; Johnson et al., 2011). We were unable to acquire an independent replication sample from the same ancestry, since large imaging genetics samples of Mexican-American ethnicity are extremely rare to date. However, the observed correlations of transcript levels with top SNPs and global FA provided an internal replication, from independently acquired data, for the associated SNPs or their LD blocks in the present sample.

Another limitation of the present study is that it is impossible to directly measure expression from brain tissue *in vivo*. Here, gene expression data were derived from lymphocytes, and the extent to which these transcripts are similarly expressed in the brain is difficult to ascertain (Cai et al., 2010). Since the correspondence of gene expression in distal brain tissue varies depending on the gene, we confirmed the presence of the transcripts of interest in a tissue specific expression atlas (Hawrylycz et al., 2012).

## 5. Conclusions

Our estimate of global FA is a sensitive quantitative phenotype for genetic analysis of common genetic variation. In a large GWAS of 776 individuals of Mexican-American ancestry, we detected five genome-wide significant SNPs that may influence white matter microstructure in the general population. Analysis of gene transcript levels suggest a potential mechanism for the effects of two SNPs on white matter microstructure, via the expression of *GNAI3* and *CCDC91*, that encourage further investigation into the biological functions of these genes and their relation to brain structure and function in health and disease.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Financial support for this study was provided by the National Institute of Mental Health Grants MH0708143 (Principal Investigator [PI]: DCG), MH078111 (PI: JB), and MH083824 (PI: DCG & JB). SOLAR is supported by National Institute of Mental Health Grant MH59490 (PI: JB) ) and by NIBIB grant EB015611 (PI: PK).

## References

- Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet.* 1998; 62:1198–1211. [PubMed: 9545414]
- Almasy L, Dyer TD, Blangero J. Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. *Genet Epidemiol.* 1997; 14:953–958. [PubMed: 9433606]
- Beaulieu C. The basis of anisotropic water diffusion in the nervous system - a technical review. *NMR Biomed.* 2002; 15:435–455. [PubMed: 12489094]
- Behrens TE, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, Brady JM, Smith SM. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn Reson Med.* 2003; 50:1077–1088. [PubMed: 14587019]
- Blangero J, Diego VP, Dyer TD, Almeida M, Peralta J, Kent JW Jr, Williams JT, Almasy L, Goring HH. A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees. *Adv Genet.* 2013; 81:1–31. [PubMed: 23419715]
- Braskie MN, Jahanshad N, Stein JL, Barysheva M, Johnson K, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Ringman JM, Toga AW, Thompson PM. Relationship of a variant in the *NTRK1* gene to white matter microstructure in young adults. *J Neurosci.* 2012; 32:5964–5972. [PubMed: 22539856]
- Burdick JT, Chen WM, Abecasis GR, Cheung VG. In silico method for inferring genotypes in pedigrees. *Nat Genet.* 2006; 38:1002–1004. [PubMed: 16921375]
- Cai C, Langfelder P, Fuller TF, Oldham MC, Luo R, van den Berg LH, Ophoff RA, Horvath S. Is human blood a good surrogate for brain tissue in transcriptional studies? *BMC Genomics.* 2010; 11:589. [PubMed: 20961428]

- Chen X, Li X, Wang P, Liu Y, Zhang Z, Zhao G, Xu H, Zhu J, Qin X, Chen S, Hu L, Kong X. Novel association strategy with copy number variation for identifying new risk Loci of human diseases. *PLoS One*. 2010; 5:e12185. [PubMed: 20808825]
- Chiang MC, Barysheva M, McMahon KL, de Zubicaray GI, Johnson K, Montgomery GW, Martin NG, Toga AW, Wright MJ, Shapshak P, Thompson PM. Gene network effects on brain microstructure and intellectual performance identified in 472 twins. *J Neurosci*. 2012; 32:8732–8745. [PubMed: 22723713]
- Chiang MC, Barysheva M, Toga AW, Medland SE, Hansell NK, James MR, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM. BDNF gene effects on brain circuitry replicated in 455 twins. *Neuroimage*. 2011a; 55:448–454. [PubMed: 21195196]
- Chiang MC, McMahon KL, de Zubicaray GI, Martin NG, Hickie I, Toga AW, Wright MJ, Thompson PM. Genetics of white matter development: a DTI study of 705 twins and their siblings aged 12 to 29. *Neuroimage*. 2011b; 54:2308–2317. [PubMed: 20950689]
- Consortium EP, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. [PubMed: 22955616]
- Dhanasekaran N, Dermott JM. Signaling by the G12 class of G proteins. *Cell Signal*. 1996; 8:235–245. [PubMed: 8842523]
- Feltri ML, Suter U, Relvas JB. The function of RhoGTPases in axon ensheathment and myelination. *Glia*. 2008; 56:1508–1517. [PubMed: 18803320]
- Flint J, Munafo MR. Candidate and non-candidate genes in behavior genetics. *Curr Opin Neurobiol*. 2013; 23:57–61. [PubMed: 22878161]
- Glahn DC, Curran JE, Winkler AM, Carless MA, Kent JW Jr, Charlesworth JC, Johnson MP, Goring HH, Cole SA, Dyer TD, Moses EK, Olvera RL, Kochunov P, Duggirala R, Fox PT, Almasy L, Blangero J. High dimensional endophenotype ranking in the search for major depression risk genes. *Biol Psychiatry*. 2012; 71:6–14. [PubMed: 21982424]
- Gold BT, Johnson NF, Powell DK, Smith CD. White matter integrity and vulnerability to Alzheimer's disease: preliminary findings and future directions. *Biochim Biophys Acta*. 2012; 1822:416–422. [PubMed: 21803153]
- Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, Jowett JB, Abraham LJ, Rainwater DL, Comuzzie AG, Mahaney MC, Almasy L, MacCluer JW, Kissebah AH, Collier GR, Moses EK, Blangero J. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet*. 2007; 39:1208–1216. [PubMed: 17873875]
- Gross RE, Mei Q, Gutekunst CA, Torre E. The pivotal role of RhoA GTPase in the molecular signaling of axon growth inhibition after CNS injury and targeted therapeutic strategies. *Cell Transplant*. 2007; 16:245–262. [PubMed: 17503736]
- Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, van de Lagemaat LN, Smith KA, Ebbert A, Riley ZL, Abajian C, Beckmann CF, Bernard A, Bertagnolli D, Boe AF, Cartagena PM, Chakravarty MM, Chapin M, Chong J, Dalley RA, Daly BD, Dang C, Datta S, Dee N, Dolbeare TA, Faber V, Feng D, Fowler DR, Goldy J, Gregor BW, Haradon Z, Haynor DR, Hohmann JG, Horvath S, Howard RE, Jeromin A, Jochim JM, Kinnunen M, Lau C, Lazars ET, Lee C, Lemon TA, Li L, Li Y, Morris JA, Overly CC, Parker PD, Parry SE, Reding M, Royall JJ, Schulkin J, Sequeira PA, Slaughterbeck CR, Smith SC, Sodt AJ, Sunkin SM, Swanson BE, Vawter MP, Williams D, Wohnoutka P, Zielke HR, Geschwind DH, Hof PR, Smith SM, Koch C, Grant SG, Jones AR. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012; 489:391–399. [PubMed: 22996553]
- Hedlund PB, Sutcliffe JG. Functional, molecular and pharmacological advances in 5-HT7 receptor research. *Trends Pharmacol Sci*. 2004; 25:481–486. [PubMed: 15559250]
- Herskowitz JH, Offe K, Deshpande A, Kahn RA, Levey AI, Lah JJ. GGA1-mediated endocytic traffic of LR11/SorLA alters APP intracellular distribution and amyloid-beta production. *Mol Biol Cell*. 2012; 23:2645–2657. [PubMed: 22621900]
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009; 106:9362–9367. [PubMed: 19474294]

- Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005; 6:95–108. [PubMed: 15716906]
- Hoptman MJ, Nierenberg J, Bertisch HC, Catalano D, Ardekani BA, Branch CA, Delisi LE. A DTI study of white matter microstructure in individuals at high genetic risk for schizophrenia. *Schizophr Res.* 2008; 106:115–124. [PubMed: 18804959]
- Ioannidis JP. Non-replication and inconsistency in the genome-wide association setting. *Hum Hered.* 2007; 64:203–213. [PubMed: 17551261]
- Jahanshad N, Kochunov PV, Sprooten E, Mandl RC, Nichols TE, Almasy L, Blangero J, Brouwer RM, Curran JE, de Zubicaray GI, Duggirala R, Fox PT, Hong LE, Landman BA, Martin NG, McMahon KL, Medland SE, Mitchell BD, Olvera RL, Peterson CP, Starr JM, Sussmann JE, Toga AW, Wardlaw JM, Wright MJ, Hulshoff Pol HE, Bastin ME, McIntosh AM, Deary IJ, Thompson PM, Glahn DC. Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the ENIGMA-DTI working group. *Neuroimage.* 2013a; 81:455–469. [PubMed: 23629049]
- Jahanshad N, Rajagopalan P, Hua X, Hibar DP, Nir TM, Toga AW, Jack CR Jr, Saykin AJ, Green RC, Weiner MW, Medland SE, Montgomery GW, Hansell NK, McMahon KL, de Zubicaray GI, Martin NG, Wright MJ, Thompson PM, Alzheimer's Disease Neuroimaging I. Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity. *Proc Natl Acad Sci U S A.* 2013b; 110:4768–4773. [PubMed: 23471985]
- Jahanshad N, Valcour VG, Nir TM, Kohanim O, Busovaca E, Nicolas K, Thompson P. Disrupted brain networks in the aging HIV+ population. *Brain Connect.* 2012
- Johnson C, Drgon T, Walther D, Uhl GR. Genomic regions identified by overlapping clusters of nominally-positive SNPs from genome-wide studies of alcohol and illegal substance dependence. *PLoS One.* 2011; 6:e19210. [PubMed: 21818250]
- Kochunov P, Glahn DC, Lancaster JL, Winkler AM, Smith S, Thompson PM, Almasy L, Duggirala R, Fox PT, Blangero J. Genetics of microstructure of cerebral white matter using diffusion tensor imaging. *Neuroimage.* 2010; 53:1109–1116. [PubMed: 20117221]
- Kochunov P, Thompson PM, Lancaster JL, Bartzokis G, Smith S, Coyle T, Royall DR, Laird A, Fox PT. Relationship between white matter fractional anisotropy and other indices of cerebral health in normal aging: tract-based spatial statistics study of aging. *Neuroimage.* 2007; 35:478–487. [PubMed: 17292629]
- Konrad A, Vucurevic G, Musso F, Stoeter P, Dahmen N, Winterer G. ErbB4 genotype predicts left frontotemporal structural connectivity in human brain. *Neuropsychopharmacology.* 2009; 34:641–650. [PubMed: 18668031]
- Kubicki M, McCarley R, Westin CF, Park HJ, Maier S, Kikinis R, Jolesz FA, Shenton ME. A review of diffusion tensor imaging studies in schizophrenia. *J Psychiatr Res.* 2007; 41:15–30. [PubMed: 16023676]
- Kvachnina E, Liu G, Dityatev A, Renner U, Dumuis A, Richter DW, Dityateva G, Schachner M, Voyno-Yasenetskaya TA, Ponimaskin EG. 5-HT7 receptor is coupled to G alpha subunits of heterotrimeric G12-protein to regulate gene transcription and neuronal morphology. *J Neurosci.* 2005; 25:7821–7830. [PubMed: 16120784]
- Li J, Li XY, Feng DF, Gu L. Quantitative evaluation of microscopic injury with diffusion tensor imaging in a rat model of diffuse axonal injury. *Eur J Neurosci.* 2011; 33:933–945. [PubMed: 21385236]
- Lopez LM, Bastin ME, Maniega SM, Penke L, Davies G, Christoforou A, Valdes Hernandez MC, Royle NA, Tenesa A, Starr JM, Porteous DJ, Wardlaw JM, Deary IJ. A genome-wide search for genetic influences and biological pathways related to the brain's white matter integrity. *Neurobiol Aging.* 2012; 33(1847):e1841–e1814.
- Mahon K, Burdick KE, Szeszeko PR. A role for white matter abnormalities in the pathophysiology of bipolar disorder. *Neurosci Biobehav Rev.* 2010; 34:533–554. [PubMed: 19896972]
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, Shafer A, Neri F, Lee K, Kutayavin T, Stehling-Sun S, Johnson AK, Canfield TK, Giste E, Diegel M, Bates D, Hansen RS, Neph S, Sabo PJ, Heimfeld S, Raubitschek A, Ziegler S, Cotsapas C, Sotoodehnia N, Glass I, Sunyaev SR, Kaul R, Stamatoyannopoulos JA. Systematic

localization of common disease-associated variation in regulatory DNA. *Science*. 2012; 337:1190–1195. [PubMed: 22955828]

- McIntosh AM, Moorhead TW, Job D, Lymer GK, Munoz Maniega S, McKirdy J, Sussmann JE, Baig BJ, Bastin ME, Porteous D, Evans KL, Johnstone EC, Lawrie SM, Hall J. The effects of a neuregulin 1 variant on white matter density and integrity. *Mol Psychiatry*. 2008; 13:1054–1059. [PubMed: 17925794]
- Mitchell AC, Bharadwaj R, Whittle C, Krueger W, Mirnics K, Hurd Y, Rasmussen T, Akbarian S. The Genome in Three Dimensions: A New Frontier in Human Brain Research. *Biol Psychiatry*. 2013
- Mitchell BD, Kammerer CM, Blangero J, Mahaney MC, Rainwater DL, Dyke B, Hixson JE, Henkel RD, Sharp RM, Comuzzie AG, VandeBerg JL, Stern MP, MacCluer JW. Genetic and environmental contributions to cardiovascular risk factors in Mexican Americans. The San Antonio Family Heart Study. *Circulation*. 1996; 94:2159–2170. [PubMed: 8901667]
- Olvera RL, Bearden CE, Velligan DI, Almasy L, Carless MA, Curran JE, Williamson DE, Duggirala R, Blangero J, Glahn DC. Common genetic influences on depression, alcohol, and substance use disorders in Mexican-American families. *Am J Med Genet B Neuropsychiatr Genet*. 2011; 156B: 561–568. [PubMed: 21557468]
- Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol*. 2008; 32:381–385. [PubMed: 18348202]
- Ponimaskin EG, Profirovic J, Vaiskunaite R, Richter DW, Voyno-Yasenetskaya TA. 5-Hydroxytryptamine 4(a) receptor is coupled to the Galpha subunit of heterotrimeric G13 protein. *J Biol Chem*. 2002; 277:20812–20819. [PubMed: 11923294]
- Puppala S, Dodd GD, Fowler S, Arya R, Schneider J, Farook VS, Granato R, Dyer TD, Almasy L, Jenkinson CP, Diehl AK, Stern MP, Blangero J, Duggirala R. A genomewide search finds major susceptibility loci for gallbladder disease on chromosome 1 in Mexican Americans. *Am J Hum Genet*. 2006; 78:377–392. [PubMed: 16400619]
- Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kahler AK, Akterin S, Bergen SE, Collins AL, Crowley JJ, Fromer M, Kim Y, Lee SH, Magnusson PK, Sanchez N, Stahl EA, Williams S, Wray NR, Xia K, Bettella F, Borglum AD, Bulik-Sullivan BK, Cormican P, Craddock N, de Leeuw C, Durmishi N, Gill M, Golimbet V, Hamshere ML, Holmans P, Hougaard DM, Kendler KS, Lin K, Morris DW, Mors O, Mortensen PB, Neale BM, O'Neill FA, Owen MJ, Milovancevic MP, Posthuma D, Powell J, Richards AL, Riley BP, Ruderfer D, Rujescu D, Sigurdsson E, Silagadze T, Smit AB, Stefansson H, Steinberg S, Suvisaari J, Tosato S, Verhage M, Walters JT, Multicenter Genetic Studies of Schizophrenia C, Levinson DF, Gejman PV, Kendler KS, Laurent C, Mowry BJ, O'Donovan MC, Owen MJ, Pulver AE, Riley BP, Schwab SG, Wildenauer DB, Dudbridge F, Holmans P, Shi J, Albus M, Alexander M, Campion D, Cohen D, Dikeos D, Duan J, Eichhammer P, Godard S, Hansen M, Lerer FB, Liang KY, Maier W, Mallet J, Nertney DA, Nestadt G, Norton N, O'Neill FA, Papadimitriou GN, Ribble R, Sanders AR, Silverman JM, Walsh D, Williams NM, Wormley B, Psychosis Endophenotypes International C, Arranz MJ, Bakker S, Bender S, Bramon E, Collier D, Crespo-Facorro B, Hall J, Iyegbe C, Jablensky A, Kahn RS, Kalaydjieva L, Lawrie S, Lewis CM, Lin K, Linszen DH, Mata I, McIntosh A, Murray RM, Ophoff RA, Powell J, Rujescu D, Van Os J, Walshe M, Weisbrod M, Wiersma D, Wellcome Trust Case Control C, Management C, Donnelly P, Barroso I, Blackwell JM, Bramon E, Brown MA, Casas JP, Corvin AP, Deloukas P, Duncanson A, Jankowski J, Markus HS, Mathew CG, Palmer CN, Plomin R, Rautanen A, Sawcer SJ, Trembath RC, Viswanathan AC, Wood NW, Data Analysis G, Spencer CC, Band G, Bellenguez C, Freeman C, Hellenthal G, Giannoulatou E, Pirinen M, Pearson RD, Strange A, Su Z, Vukcevic D, Donnelly P, Dna GDQC, Informatics G, Langford C, Hunt SE, Edkins S, Gwilliam R, Blackburn H, Bumpstead SJ, Dronov S, Gillman M, Gray E, Hammond N, Jayakumar A, McCann OT, Liddle J, Potter SC, Ravindrarajah R, Ricketts M, Tashakkori-Ghanbaria A, Waller MJ, Weston P, Widaa S, Whittaker P, Barroso I, Deloukas P, Publications C, Mathew CG, Blackwell JM, Brown MA, Corvin AP, McCarthy MI, Spencer CC, Bramon E, Corvin AP, O'Donovan MC, Stefansson K, Scolnick E, Purcell S, McCarroll SA, Sklar P, Hultman CM, Sullivan PF. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet*. 2013



- Schmierer K, Wheeler-Kingshott CA, Boulby PA, Scaravilli F, Altmann DR, Barker GJ, Tofts PS, Miller DH. Diffusion tensor imaging of post mortem multiple sclerosis brain. *Neuroimage*. 2007; 35:467–477. [PubMed: 17258908]
- Seifuddin F, Mahon PB, Judy J, Pirooznia M, Jancic D, Taylor J, Goes FS, Potash JB, Zandi PP. Meta-analysis of genetic association studies on bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2012; 159B:508–518. [PubMed: 22573399]
- Siehler S. Regulation of RhoGEF proteins by G12/13-coupled receptors. *Br J Pharmacol*. 2009; 158:41–49. [PubMed: 19226283]
- Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE, Watkins KE, Ciccarelli O, Cader MZ, Matthews PM, Behrens TE. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage*. 2006; 31:1487–1505. [PubMed: 16624579]
- Sobel E, Lange K. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am J Hum Genet*. 1996; 58:1323–1337. [PubMed: 8651310]
- Sprooten E, Brumbaugh MS, Knowles EEM, McKay DR, Lewis J, Barrett J, Landau S, Cyr L, Kochunov P, Winkler AM, Pearlson GD, Glahn DC. Reduced white matter integrity in sibling pairs discordant for bipolar disorder. *Am Journal of Psychiatry*. 2013a
- Sprooten E, Fleming KM, Thomson PA, Bastin ME, Whalley HC, Hall J, Sussmann JE, McKirdy J, Blackwood D, Lawrie SM, McIntosh AM. White matter integrity as an intermediate phenotype: exploratory genome-wide association analysis in individuals at high risk of bipolar disorder. *Psychiatry Res*. 2013b; 206:223–231. [PubMed: 23218918]
- Sprooten E, Lymer GK, Munoz Maniega S, McKirdy J, Clayden JD, Bastin ME, Porteous D, Johnstone EC, Lawrie SM, Hall J, McIntosh AM. The relationship of anterior thalamic radiation integrity to psychosis risk associated neuregulin-1 variants. *Mol Psychiatry*. 2009; 14:237–238. 233. [PubMed: 19229203]
- Sprooten E, Sussmann JE, Clugston A, Peel A, McKirdy J, Moorhead TW, Anderson S, Shand AJ, Giles S, Bastin ME, Hall J, Johnstone EC, Lawrie SM, McIntosh AM. White matter integrity in individuals at high genetic risk of bipolar disorder. *Biol Psychiatry*. 2011a; 70:350–356. [PubMed: 21429475]
- Sprooten E, Sussmann JE, Moorhead TW, Whalley HC, Ffrench-Constant C, Blumberg HP, Bastin ME, Hall J, Lawrie SM, McIntosh AM. Association of white matter integrity with genetic variation in an exonic DISC1 SNP. *Mol Psychiatry*. 2011b; 16(685):688–689.
- Stebbins GT, Murphy CM. Diffusion tensor imaging in Alzheimer's disease and mild cognitive impairment. *Behav Neurol*. 2009; 21:39–49. [PubMed: 19847044]
- Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, Toro R, Appel K, Bartecek R, Bergmann O, Bernard M, Brown AA, Cannon DM, Chakravarty MM, Christoforou A, Domin M, Grimm O, Hollinshead M, Holmes AJ, Homuth G, Hottenga JJ, Langan C, Lopez LM, Hansell NK, Hwang KS, Kim S, Laje G, Lee PH, Liu X, Loth E, Lourdasamy A, Mattingsdal M, Mohnke S, Maniega SM, Nho K, Nugent AC, O'Brien C, Pappmeyer M, Putz B, Ramasamy A, Rasmussen J, Rijpkema M, Risacher SL, Roddey JC, Rose EJ, Ryten M, Shen L, Sprooten E, Strengman E, Teumer A, Trabzuni D, Turner J, van Eijk K, van Erp TG, van Tol MJ, Wittfeld K, Wolf C, Woudstra S, Aleman A, Alhusaini S, Almasy L, Binder EB, Brohawn DG, Cantor RM, Carless MA, Corvin A, Czisch M, Curran JE, Davies G, de Almeida MA, Delanty N, Depondt C, Duggirala R, Dyer TD, Erk S, Fagerness J, Fox PT, Freimer NB, Gill M, Goring HH, Hagler DJ, Hoehn D, Holsboer F, Hoogman M, Hosten N, Jahanshad N, Johnson MP, Kasperaviciute D, Kent JW Jr, Kochunov P, Lancaster JL, Lawrie SM, Liewald DC, Mandl R, Matarin M, Mattheisen M, Meisenzahl E, Melle I, Moses EK, Muhleisen TW, Nauck M, Nothen MM, Olvera RL, Pandolfo M, Pike GB, Puls R, Reinvang I, Renteria ME, Rietschel M, Roffman JL, Royle NA, Rujescu D, Savitz J, Schnack HG, Schnell K, Seiferth N, Smith C, Steen VM, Valdes Hernandez MC, Van den Heuvel M, van der Wee NJ, Van Haren NE, Veltman JA, Volzke H, Walker R, Westlye LT, Whelan CD, Agartz I, Boomsma DI, Cavalleri GL, Dale AM, Djurovic S, Drevets WC, Hagoort P, Hall J, Heinz A, Jack CR Jr, Foroud TM, Le Hellard S, Macciardi F, Montgomery GW, Poline JB, Porteous DJ, Sisodiya SM, Starr JM, Sussmann J, Toga AW, Veltman DJ, Walter H, Weiner MW, Alzheimer's Disease Neuroimaging I, Consortium E, Consortium I, Saguenay Youth Study G, Bis JC, Ikram MA, Smith AV, Gudnason V, Tzourio C, Vernooij MW, Launer LJ, DeCarli C, Seshadri S, Cohorts for H, Aging Research in Genomic Epidemiology C, Andreassen OA,

Apostolova LG, Bastin ME, Blangero J, Brunner HG, Buckner RL, Cichon S, Coppola G, de Zubicaray GI, Deary IJ, Donohoe G, de Geus EJ, Espeseth T, Fernandez G, Glahn DC, Grabe HJ, Hardy J, Hulshoff Pol HE, Jenkinson M, Kahn RS, McDonald C, McIntosh AM, McMahon FJ, McMahon KL, Meyer-Lindenberg A, Morris DW, Muller-Myhsok B, Nichols TE, Ophoff RA, Paus T, Pausova Z, Penninx BW, Potkin SG, Samann PG, Saykin AJ, Schumann G, Smoller JW, Wardlaw JM, Weale ME, Martin NG, Franke B, Wright MJ, Thompson PM, Enhancing Neuro Imaging Genetics through Meta-Analysis C. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet.* 2012; 44:552–561. [PubMed: 22504417]

Whalley HC, Sprooten E, Hackett S, Hall L, Blackwood DH, Glahn DC, Bastin M, Hall J, Lawrie SM, Sussmann JE, McIntosh AM. Polygenic risk and white matter integrity in individuals at high risk of mood disorder. *Biol Psychiatry.* 2013; 74:280–286. [PubMed: 23453289]

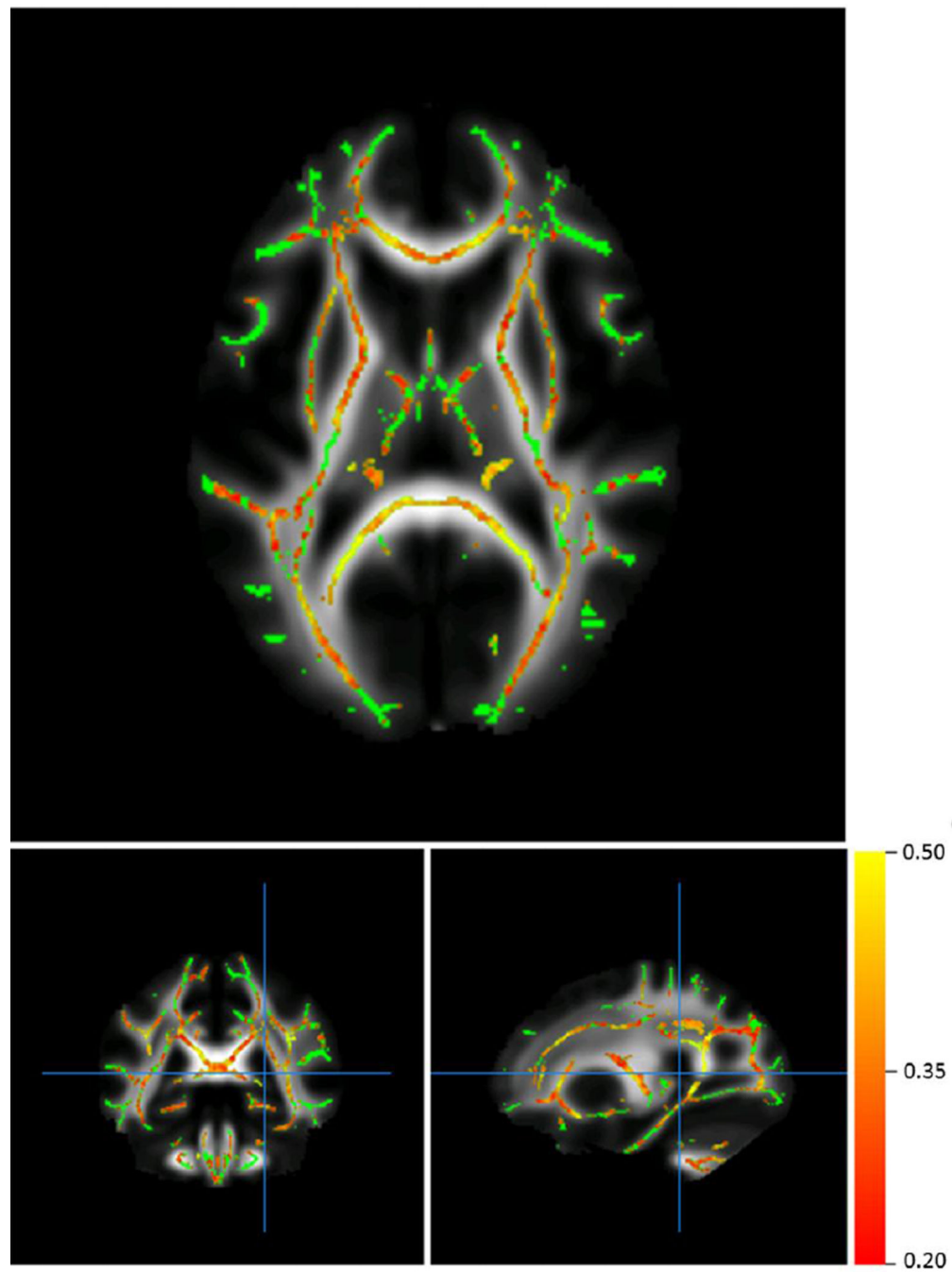
Williams JT, Van Eerdewegh P, Almasy L, Blangero J. Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. Likelihood formulation and simulation results. *Am J Hum Genet.* 1999; 65:1134–1147. [PubMed: 10486333]

Winterer G, Konrad A, Vucurevic G, Musso F, Stoeter P, Dahmen N. Association of 5' end neuregulin-1 (NRG1) gene variation with subcortical medial frontal microstructure in humans. *Neuroimage.* 2008; 40:712–718. [PubMed: 18255317]

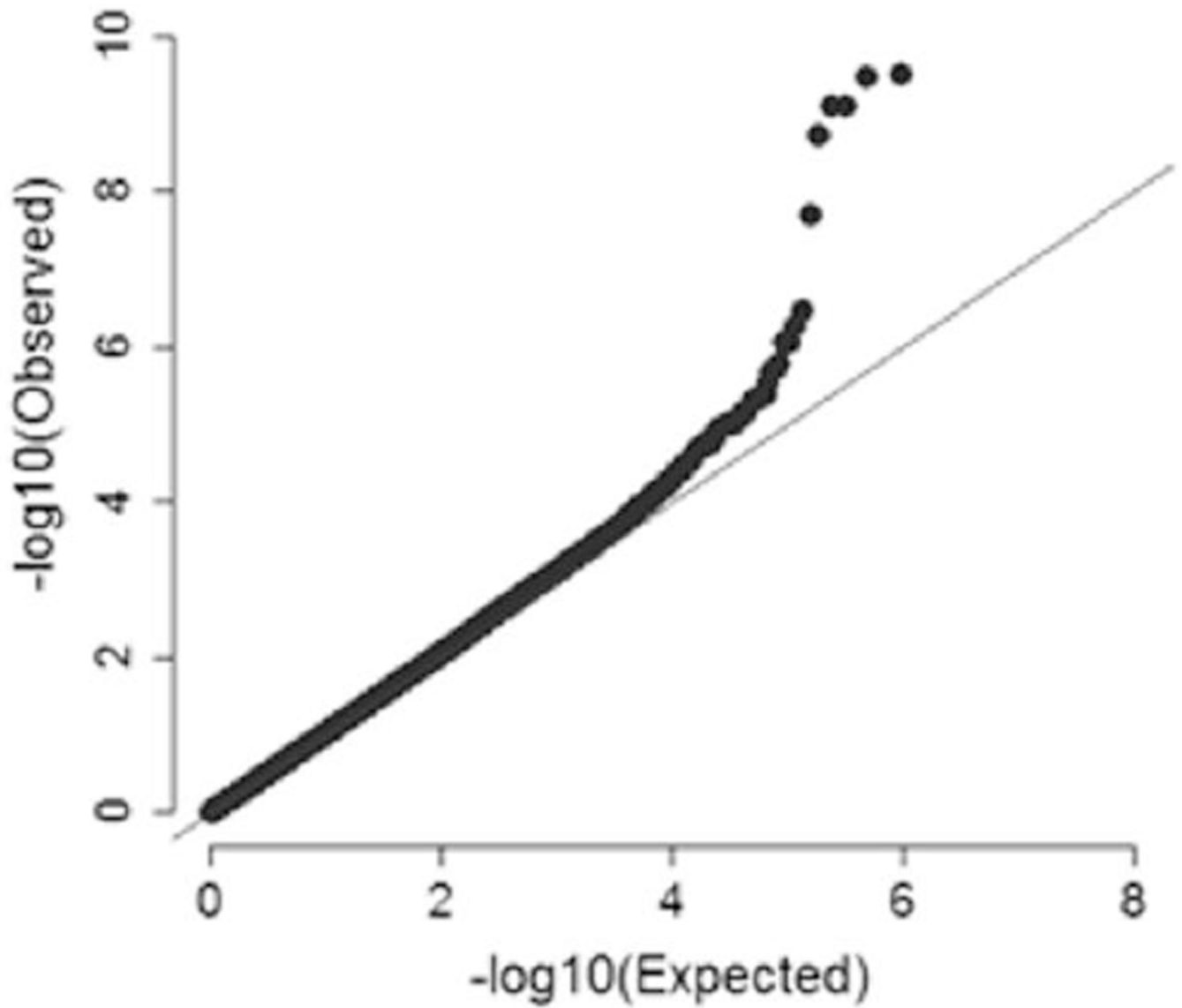
Zuliani R, Moorhead TW, Bastin ME, Johnstone EC, Lawrie SM, Brambilla P, O'Donovan MC, Owen MJ, Hall J, McIntosh AM. Genetic variants in the *ErbB4* gene are associated with white matter integrity. *Psychiatry Res.* 2011; 191:133–137. [PubMed: 21232925]

### Highlights

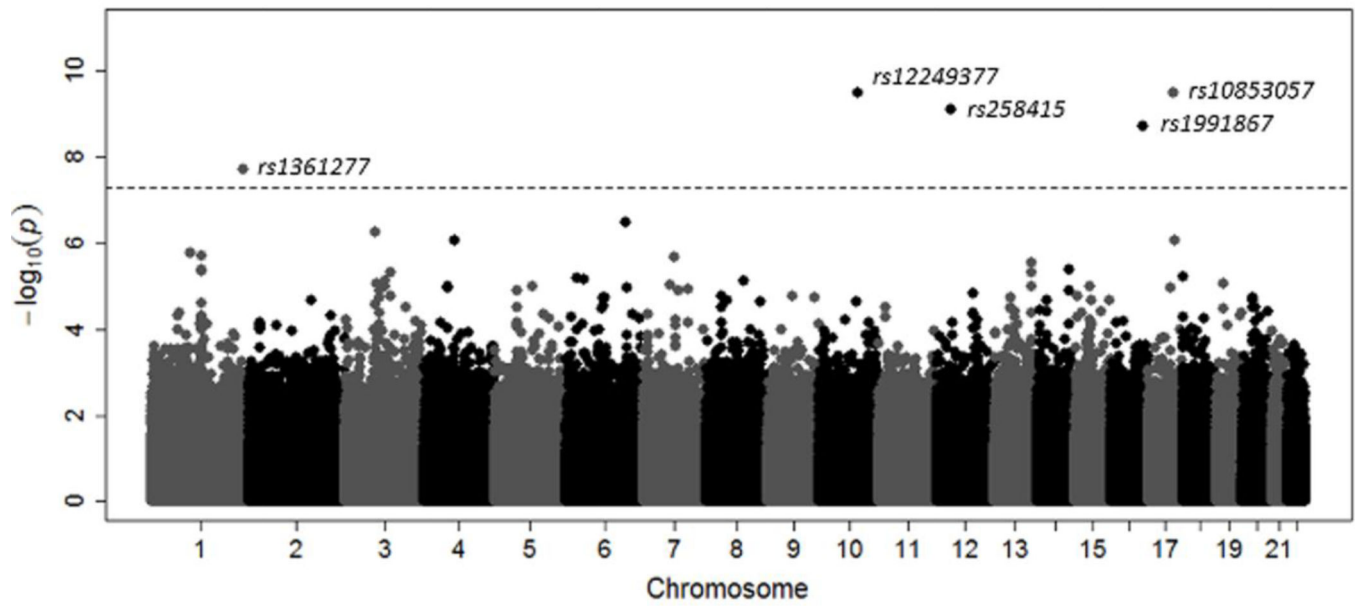
- Five genome-wide significant SNPs were detected in association with FA
- *HTR7* expression is associated with FA and with intronic genome-wide significant hit
- *GNA13* and *CCDC91* expression may mediate effects of genome-wide significant SNPs
- Additive genetic effects on FA are largely homogeneous across white matter tracts



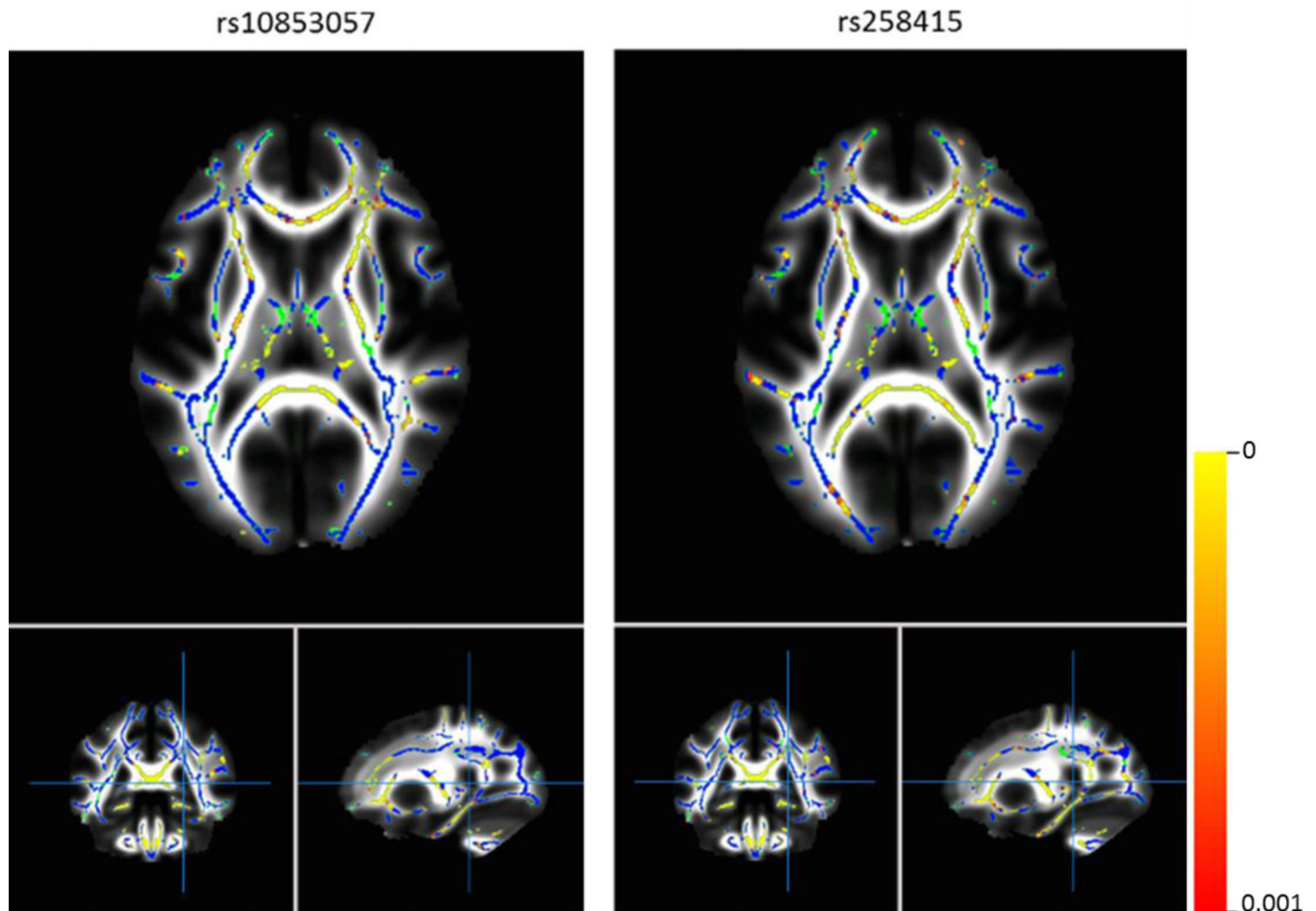
**Figure 1.** Voxel-wise heritability of FA across the TBSS skeleton (green). Voxels where the heritability of FA was 0.2 or greater are colored according to the red-yellow color bar below.



**Figure 2.** Quantile-quantile plot of GWA results on global FA, demonstrating population stratification effects were minimal, with genomic inflation factor  $\lambda=1.03$ .



**Figure 3.** Manhattan plot of GWA results on global FA, showing five SNPs significant at the genome-wide significance level of  $< 5 \times 10^{-8}$ .



**Figure 4.**

Voxel-wise association results for rs10853057 and rs258415 overlaid on top of the mean FA image. Voxels in red-yellow are significantly associated with the SNP at  $p < 0.001$  (uncorrected) in the same direction as the SNP's association with global FA. To further illustrate the degree of spatial homogeneity of the effect, the remaining voxels (with  $p > 0.001$ ) are colored blue if the allelic effect was in the same direction as for the global FA measure. Voxels are colored green if the SNP was associated with FA in direction opposite of its effect on global FA.

**Table 1**

Genome-wide significant SNPs on mean FA in the entire TBSS skeleton

SNP	X <sup>2</sup>	p	MAF	Chr	SNP function	Genes (<50kb)	HWE X <sup>2</sup>
rs10853057	39.56	3.18E-10	0.23	17	Intergenic	GNAI3	2.78
rs12249377	39.51	3.26E-10	0.31	10	Intron	HTR7	6.32
rs258415	37.77	7.97E-10	0.30	12	Intergenic	NA	0.98
rs1991867	36.03	1.94E-09	0.40	16	Intergenic	NA	0.40
rs1361277	31.72	2.00E-08	0.29	1	Intergenic	NA	1.16

MAF = Minor Allele Frequency; HWE = Hardy-Weinberg Equilibrium; Chr = Chromosome



Table 2

Effects of genotypes at top SNPs on expression level of nearby genes

Transcript	Gene	Nearby top SNP	h <sup>2</sup>	P (h <sup>2</sup> )	P (SNP)	P <sub>FDR</sub> (SNP)
ILMN_1676523	CCDC91	rs258415	0.38	4.00E-07	1.96E-06	7.35E-06*
ILMN_2190051	CCDC91	rs258415	0.21	0.009	0.06	0.09
ILMN_1758906	GNA13	rs10853057	0.42	1.00E-07	6.18E-11	9.27E-10*
ILMN_2176037	GNA13	rs10853057	0.18	0.006	7.52E-10	5.64E-09*
ILMN_1744217	HTR7	rs12249377	0.09	0.16	0.006	0.01*
ILMN_1652732	HTR7	rs12249377	0.18	0.01	0.73	0.78
ILMN_1727134	KLHDC5	rs258415	0.17	0.01	0.03	0.04*
ILMN_1781360	MPHOSPH6	rs1991867	0.32	1.24E-05	0.003	0.006*
ILMN_2189993	MRPFS35	rs258415	0.55	4.77E-11	0.003	0.006*
ILMN_1788059	PCGF5	rs12249377	0.19	0.007	2.45E-08	1.23E-07*
ILMN_1754468	RGS9	rs10853057	0.26	0.003	0.06	0.09
ILMN_2389984	RGS9	rs10853057	0.20	0.002	0.28	0.32
ILMN_1792283	RPP30	rs12249377	0.18	0.03	0.09	0.12
ILMN_1760798	RYR2	rs1361277	0.18	0.01	0.86	0.86
ILMN_1675429	SMURF2	rs10853057	0.12	0.05	0.0002	6.89E-04*

\* Significant at FDR &lt; 0.05

**Table 3**

Genetic correlations between global FA and nine transcripts

Transcript	Gene	$\rho_p$ with FA	P ( $\rho_p$ )	P <sub>FDR</sub> ( $\rho_p$ )	$\rho_g$ with FA	P ( $\rho_g$ )	P <sub>FDR</sub> ( $\rho_g$ ) <sup>I</sup>
ILMN_1676523	CCDC91	0.21	0.0002	0.0006	0.41	0.006	0.048*
ILMN_1758906	GNA13	0.15	0.01	0.02	0.13	0.33	0.42
ILMN_2176037	GNA13	0.23	0.0001	0.0005	0.42	0.01	0.048*
ILMN_1744217	HTR7	-0.17	0.004	0.007	-0.27	0.27	0.41
ILMN_1727134	KLHDC5	0.07	0.24	0.31	0.05	0.87	0.87
ILMN_1781360	MPHOSPH6	0.04	0.52	0.58	0.26	0.14	0.31
ILMN_2189993	MRPS35	0.00	0.95	0.95	0.13	0.39	0.44
ILMN_1788059	PCGF5	0.17	0.003	0.007	0.22	0.21	0.37
ILMN_1675429	SMURF2	-0.25	<0.0001	<0.0001	-0.40	0.14	0.31

<sup>I</sup> Only transcripts that significantly associated with top SNPs (Table 3) were considered;  $\rho_p$  = phenotypic correlation  $\rho_g$  = genotypic correlation;

\* Significant at FDR < 0.05