Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: Comparing meta and megaanalytical approaches for data pooling

CrossMark

Peter Kochunov [a,*,1], Neda Jahanshad [b,1], Emma Sprooten [c], Thomas E. Nichols [d,e], René C. Mandl [f], Laura Almasy [g], Tom Booth [h], Rachel M. Brouwer [f], Joanne E. Curran [g], Greig I. de Zubicaray [i], Rali Dimitrova [j], Ravi Duggirala [g], Peter T. Fox [k], L. Elliot Hong [a], Bennett A. Landman [l], Hervé Lemaitre [m], Lorna M. Lopez [h], Nicholas G. Martin [o], Katie L. McMahon [p], Braxton D. Mitchell [q], Rene L. Olvera [r], Charles P. Peterson [g], John M. Starr [h,s], Jessika E. Sussmann [j], Arthur W. Toga [b], Joanna M. Wardlaw [h,n], Margaret J. Wright [o], Susan N. Wright [a], Mark E. Bastin [h,n], Andrew M. McIntosh [j], Dorret I. Boomsma [t], René S. Kahn [f], Anouk den Braber [t], Eco J.C. de Geus [t], Ian J. Deary [h], Hilleke E. Hulshoff Pol [f], Douglas E. Williamson [r], John Blangero [g], Dennis van 't Ent [t], Paul M. Thompson [b,u], David C. Glahn [c]

[a] Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, USA
[b] Imaging Genetics Center, Institute of Neuroimaging and Informatics, Department of Neurology, Keck School of Medicine of USC, Los Angeles, CA, USA
[c] Olin Neuropsychiatry Research Center in the Institute of Living, Yale University School of Medicine, New Haven, CT, USA
[d] Department of Statistics & Warwick Manufacturing Group, The University of Warwick, Coventry, UK
[e] Oxford Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, Oxford University, UK
[f] Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands
[g] Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX, USA
[h] Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, The University of Edinburgh, Edinburgh, UK
[i] School of Psychology, University of Queensland, Brisbane, Australia
[j] Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK
[k] Research Imaging Institute, University of Texas Health Science Center San Antonio, San Antonio, TX, USA
[l] Department of Electrical Engineering, Vanderbilt University, Nashville, TN, USA
[m] U1000 Research Unit Neuroimaging and Psychiatry, INSERM-CEA-Faculté de Médecine Paris-Sud, Orsay, France
[n] Brain Research Imaging Centre, Centre for Clinical Brain Sciences, University of Edinburgh, UK
[o] QIMR Berghofer Medical Research Institute, Brisbane, Australia
[p] Centre for Advanced Imaging, University of Queensland, Brisbane, Australia
[q] Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA
[r] Department of Psychiatry, University of Texas Health Science Center San Antonio, San Antonio, TX, USA
[s] Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK
[t] Department of Biological Psychology, VU University, Amsterdam, The Netherlands
[u] Department of Neurology, Pediatrics, Engineering, Psychiatry, Radiology, & Ophthalmology, University of Southern California, Los Angeles, CA, USA

ARTICLE INFO

ABSTRACT

Combining datasets across independent studies can boost statistical power by increasing the numbers of observations and can achieve more accurate estimates of effect sizes. This is especially important for genetic studies where a large number of observations are required to obtain sufficient power to detect and replicate genetic effects. There is a need to develop and evaluate methods for joint-analytical analyses of rich datasets collected in imaging genetics studies. The ENIGMA-DTI consortium is developing and evaluating approaches for obtaining pooled estimates of heritability through meta-and mega-genetic analytical approaches, to estimate the general additive genetic contributions to the intersubject variance in fractional anisotropy (FA) measured from diffusion tensor imaging (DTI). We used the ENIGMA-DTI data harmonization protocol for uniform processing of DTI data from multiple sites. We evaluated this protocol in five family-based cohorts providing data from a total of 2248 children and adults (ages: 9–85) collected with various imaging protocols. We used the imaging genetics analysis tool, SOLAR-Eclipse, to combine twin and family data from Dutch, Australian and Mexican-American cohorts into

* Corresponding author.
  E-mail address: pkochunov@mprc.umaryland.edu (P. Kochunov).
[1] PK and NJ contributed equally to this work.

one large "mega-family". We showed that heritability estimates may vary from one cohort to another. We used two meta-analytical (the sample-size and standard-error weighted) approaches and a mega-genetic analysis to calculate heritability estimates across-population. We performed leave-one-out analysis of the joint estimates of heritability, removing a different cohort each time to understand the estimate variability. Overall, meta- and mega-genetic analyses of heritability produced robust estimates of heritability.

Introduction

Human brain mapping studies have shown substantial advantages of pooling data across multiple studies (Van Horn et al., 2004). Genetic analyses, particularly genome-wide association studies (GWAS), tend to be limited in statistical power as there is typically a small (<0.5%, Flint and Munafò, 2013) contribution to complex phenotypic variability from individual, common genetic variants. This limitation is especially problematic for imaging genetic studies of human brain. The structure and function of the human brain is greatly influenced by genetics, but the proportion of the variance due to individual differences in the human genome depends on the brain structure and measure assessed (Kochunov et al., 2009, 2010).

A large number of neuroimaging traits with ever-increasing spatial resolution are becoming increasingly available to describe the regional complexity of brain variability. This presents a daunting challenge where the number of degrees of freedom, in both the imaging and genetics, can be overwhelming for any single imaging-genetic study. Therefore, data pooling strategies are crucial whereby data from multiple large imaging genetics studies can be analyzed together.

Imaging and genetics have both greatly advanced neuroscience in recent years. The two fields have developed in parallel but in the last decade, there was a push to merge them to fully capitalize on their power leading to the development of the new field of imaging genetics. This field emerged as a variation of classical genetic analyses that related diagnostic, clinical and/or behavioral measures to locations and specific variants in the genome. This new field is thought to be able to provide new approaches to characterize, treat and potentially prevent some brain-related disorders. Insight into biological mechanisms that predispose individuals to these types of illnesses holds the promise of yielding potential new therapies and a significant reduction of this considerable burden. Advantages of imaging genetics include the presumed greater biological proximity to genetic variation and the quantitative nature of imaging phenotypes, which ideally suited for partitioning phenotypic variance into variance explained by genetic and environmental factors. Therefore, the statistical power of genetic analysis depends on both the closeness of a phenotype to the action of the gene and the precision of the measurements. Modern MRI offers phenotypic measurements that may provide a more detailed description of the disorder than clinical symptoms or neuropsychological assessments, and many of these measures have high precision and reproducibility. Our experience and that of others indicate that the inter-session, scan-rescan variability of many common imaging measurements can be low, in the range of 1–5% (Agartz et al., 2001; Julin et al., 1997; Kochunov and Duff Davis, 2009; Kochunov et al., 2012b; Lemieux et al., 1999; Lerch and Evans, 2005). Therefore, the imaging genetics approach may help ascertain effects of specific genetic variants on the human brain and may also discover genetic variants associated with neurological or psychiatric illnesses (Braskie et al., 2011; Chen et al., 2012; Glahn et al., 2007, 2010; Hasler and Northoff, 2011; Thompson et al., 2001; van den Heuvel et al., 2013).

In imaging genetic studies, up to a million voxel-based imaging traits may be analyzed. The required correction for multiple comparisons may limit the statistical power for gene discovery, even in the largest individual imaging studies of hundreds or even thousands of subjects. One solution is collaborative data sharing and pooling

through consortia such as Enhancing Neuroimaging Genetics through Meta-Analysis (ENIGMA) Consortium (http://enigma.ini.usc.edu). Recent examples highlight the potential of large, meta-analyses of genome-wide association studies (GWAS) to uncover genetic loci that are reliably and consistently associated with MRI-based phenotypes in worldwide datasets, including hippocampal volumes (Bis et al., 2012; Stein et al., 2012), intracranial volumes (Ikram et al., 2012; Stein et al., 2012), and head circumference (Taal et al., 2012). Recently, the ENIGMA-DTI Consortium working group was organized to develop methods to facilitate multi-site approaches to study genetic influences on white matter micro-architecture and integrity, assessed using diffusion tensor imaging (DTI). Here we specifically focus on the fractional anisotropy (FA) as it is the most commonly analyzed scalar parameter extracted from DTI (Basser and Pierpaoli, 1996; Basser et al., 1994). The absolute FA values are sensitive to fiber coherence, myelination levels, and axonal integrity, and have been widely used as an index of white matter integrity (Thomason and Thompson, 2011). FA has emerged as a sensitive index of diffuse abnormalities in many brain disorders including Alzheimer's disease (AD) (Clerx et al., 2012; Teipel et al., 2012); in many studies, it is related to cognitive performance (Penke et al., 2010a, 2010b) and is altered in numerous psychiatric disorders including major depressive disorder (Carballedo et al., 2012) and bipolar disorder (Barysheva et al., 2012; Sprooten et al., 2011). Patient-control differences in FA values are also among the most replicable and consistent neuroimaging findings in schizophrenia (Alba-Ferrara and de Erausquin, 2013; Friedman et al., 2008; Kochunov et al., 2012a; Mandl et al., 2012; Nazeri et al., 2013; Perez-Iglesias et al., 2011).

The goal of the ENIGMA-DTI Working Group is to develop generalizable analyses, methods, and techniques for extraction and combined genetic analysis of phenotypes from DTI data collected from imaging groups around the world, regardless of the imaging acquisition or specific population under study. Its overall aim is to discover genetic factors influencing or related to white matter architecture. The first step towards this goal was the development of homogenization protocols to reliably extract phenotypic measurements from data collected with different imaging equipment and parameters (Jahanshad et al., 2013). The next step is to evaluate different statistical approaches for data pooling and specifically compare meta and megaanalytical techniques to choose one approach that yields the greatest improvements in the power of genetic discovery while accommodating for potential for genetic diversity among samples. Two specific advantages of data pooling are the increased power for genetic discovery and the genetic diversity of the population sample. Data pooling makes it easier to identify genetic variants that exert only small individual effects (Zuk et al., 2012). However, pooling data may be confounded by variations in data acquisition across datasets, heterogeneities in study population, and other factors. Another limitation to data pooling arises from restrictions that can arise with sharing raw data (both ethical and regulatory), including either phenotypic or genetic information.

Here, we tested and compared the outcomes of three approaches to pool imaging genetic data from five separate cohorts worldwide, that had used various imaging acquisition parameters and population structures for analysis of heritability of the DTI–FA. Of course the ultimate goal is to detect specific variants on the genome that

relate to DTI measures; but in the past we have found it boosts power in GWAS to first rank phenotypes by heritability, discarding the least promising ones for data and dimensionality reduction (Jahanshad et al., 2013).

Two different approaches were used to perform aggregation of heritability estimates derived from each cohort individually, known as meta-analytical pooling, where no raw data were shared between sites. One meta-analytic approach tested here weighted heritability estimates by the standard errors derived from the heritability estimates (MA-SE) (Borenstein et al., 2009); the other approach, although not optimal in this scenario as explained in the discussion, is commonly used for meta-analysis of heritability estimates of twin-only studies (Bergen et al., 2007; Rhee and Waldman, 2002) and therefore was used here for comparison; it involves weighting individual results by the total *number* of subjects per sample (MA-N). The third approach, the "mega-analytic" approach, individually normalizes each cohort to the same imaging template, and then pools together all data. For a mega-analytic GWAS all individual subject genotype information, or some imputed version of it, would commonly be shared. But in the case presented here, as we were not performing a genome-wide scan, the shared data included only the population and family structure of the cohort in addition to the FA images and covariate information.

All data was processed using the harmonization protocol previously developed by our group ((Jahanshad et al., 2013) provided on-line at http://enigma.ini.usc.edu/ongoing/dti-working-group/). This included the use of the ENIGMA protocol for following the QA/QC steps for each site, registration to the ENIGMA-DTI target, extraction of white matter skeleton, followed by extraction of tract-average FA values. After processing all datasets with this protocol, we performed heritability analyses using the Sequential Oligogenic Linkage Analysis Routines (SOLAR) software package (http://www.nitrc.org/projects/se_linux) (Almasy and Blangero, 1998) for tract-average FA values as well as on a detailed voxel-wise level. The data pooling approaches we compared to assess power and stability in estimation.

As an extension of our prior paper, this paper presents much deeper view of DTI heritability in the largest sample ever analyzed. With an unprecedented sample size of over 2200 scans, which to the best of our knowledge is the largest for a DTI imaging genetics study – or any DTI study, for that matter – we show the ability to measure heritability consistently across populations in a multi-site manner and provide estimates with narrower confidence intervals, which is sorely needed in heritability analyses. We approximately double our previous sample size and extend our analyses to include populations of children. Our goal was to determine whether, regardless of population, we are able to prioritize targets for GWAS with respect to stability, reliability, and heritability estimates. We provide our combined heritability results online at http://enigma.ini.usc.edu/ongoing/dti-working-group/2014_nimg/.

Methods

Study subjects and imaging protocols

Five datasets (GOBS, QTIM, TAOS, NTR, and BrainSCALE; the acronyms are detailed below) contributed data from 2203 subjects, aged 9–85 years (Table 1). All datasets were family-based studies that allowed for measurements of heritability.

- GOBS — Genetics of Brain Structure and Function study
  o Subjects: The sample comprised 859 (351 M/508 F, mean age: 43.2 ± 15.0) Mexican-American individuals from 73 extended pedigrees (average size, 17.2 people, range = 1–247) from the San Antonio Family Study (Olvera et al., 2011). Only 814 subjects in the sample (350 M/464 F, mean age 43.1 ± 15.2, ranging from 19 to 85 years) had a significant genetic relationship with other subjects and provided contribution to the heritability measurement. However, the entire sample was used to estimate the covariate effects. Individuals in this cohort have actively participated in research for over 18 years and were randomly selected from the community with the constraints that they are of Mexican-American ancestry, part of a large family, and live in the San Antonio region. All participants provided written informed consent on forms approved by the Institutional Review Boards at the University of Texas Health Science Center San Antonio (UTHSCSA) and Yale University before participating in any aspect of the study. Imaging: All imaging was performed at the Research Imaging Institute, UTHSCSA, on a Siemens 3 T Trio scanner using a multi-channel phased array head coil. A single-shot single refocusing spin-echo, echo-planar imaging sequence was used to acquire diffusion-weighted data with a spatial resolution of $1.7 \times 1.7 \times 3.0$ mm. The sequence parameters were: TE/TR = 87/8000 ms, FOV = 200 mm, 55 isotropically distributed diffusion weighted directions, two diffusion weighting values, $b = 0$ and 700 s/mm^2 and three $b = 0$ (non-diffusion-weighted) images.
- QTIM — the Queensland Twin IMaging study
  o Subjects: The sample comprised of 625 young adult Australian twins and siblings (231 M/394 F, average age = 22.9 ± 2.1 year, range 21–29), all of European ancestry. The sample including 79 monozygotic (MZ) pairs, and 80 dizygotic (DZ) pairs and their siblings. Additional information on the cohort, inclusion and exclusion criteria, and the study goals may be found in (de Zubicaray et al., 2008).
  o Imaging: A high angular resolution diffusion imaging protocol was used, consisting of a 14.2-minute, 105-image acquisition, with a 4-Tesla Bruker Medspec MRI scanner. Imaging parameters were:

Table 1
Demographic and imaging information for the 5 cohorts used in this study are listed.

	GOBS	QTIM	TAOS	Adults NTR	BrainSCALE NTR
Relatedness	Pedigree	Twins and siblings	Twins and siblings	Twins	Twins
Race/ethnicity	Mexican-American	Caucasian	Caucasian	Caucasian	Caucasian
Number subjects contributed	859	625	319	246	199
Sex	351 M/508 F	231 M/394 F	158 M/161 F	93 M/153 F	100 M/99 F
Age range (years)	19–85	21–29	12–15	18–45	9–10
Healthy	Yes	Yes	Yes	Yes	Yes
Scanner	Siemens	Bruker/Siemens	Siemens	Intera	Achieva
N-acquisitions	1	1	1	1	2
Field strength	3 T	4 T	3 T	3 T	1.5 T
Voxel size (mm)	$1.7 \times 1.7 \times 3.0$	$1.8 \times 1.8 \times 2$	$1.7 \times 1.7 \times 3.0$	$2.5 \times 2.5 \times 2.5$	$2.5 \times 2.5 \times 2.5$
N-gradients	55	94	55	32	32
N-b_0	3	11	3	1	8
b-value (s/mm^2)	700	1159	700	1000	1000
Reference	Kochunov et al., 2011	de Zubicaray et al., 2008	Kochunov et al., 2011	den Braber et al., 2013	Brouwer et al., 2010

TR/TE = 6090/91.7 ms, FOV = 23 cm, with a 128 × 128 acquisition matrix. Each 3D volume consisted of 55 2-mm thick axial slices with no gap and 1.8 × 1.8 mm^2 in-plane resolution. 105 images were acquired per subject: 11 with no diffusion sensitization (i.e., T2-weighted b$_0$ images) and 94 diffusion-weighted (DW) images (b = 1159 s/mm^2) with gradient directions uniformly distributed on the hemisphere.

- TAOS — the Teen Alcohol Outcomes Study.
  - o Subjects: The sample comprised of 319 twin and sibling adolescent participants (158 M/161 F) residing in San Antonio, Texas. Subjects were aged 12 to 15 years, average age: 13.4 years ± 0.96 SD at the time of imaging. The study consisted of 12 pairs of twins (MZ = 4/DZ = 8) and siblings with the average family size of 2.1 ± 0.2. All subjects were healthy and free of drug abuse at the time of the imaging. All subjects and their parents signed informed assents and informed consents, respectively.
  - o Imaging: Diffusion imaging was performed at the Research Imaging Center, UTHSCSA, on a Siemens 3 T Trio scanner the same protocol as GOBS study.
- NTR — Netherlands Twin Register
  - o Subjects: 246 adults (93/153 M/F; average age 33.9 ± 10.1, range 19–57), monozygotic and dizygotic twin pairs (72/48/6 MZ/DZ/Sibs), were recruited from the Netherlands Twin Registry. Exclusion criteria consisted of having any metal material in the head, having a pacemaker, a known history of any major medical condition or psychiatric illness (den Braber et al., 2008, 2012).
  - o Imaging: Diffusion tensor data were collected on a 3.0 T Philips Intera MR scanner (32 diffusion-weighted volumes with different non-collinear diffusion directions with b-factor = 1000 s/mm^2 and one b-factor = 0 s/mm^2 image, flip angle = 90°; 38 axial slices of 3.0 mm; no slice gap; voxel size, 2.0 × 2.0 × 3.0 mm; FOV = 230 mm; TE = 94 ms; TR = 4863 ms; no cardiac gating; and total scan duration = 185 s) (den Braber et al., 2011).
- BrainSCALE — Brain Structure and Cognition: an Adolescent Longitudinal Twin Study into genetic Etiology
University Medical Center Utrecht and Netherlands Twin Register, Amsterdam
  - o Subjects: 199 children (100 M/99 F; average age 9.2 ± 0.1, range 9.0–9.6), mono and dizygotic twin pairs (42/57 MZ/DZ), were recruited from families participating in the BrainSCALE cohort (van Soelen et al., 2012) that was recruited from the Netherlands Twin Register (van Beijsterveldt et al., 2013). Imaging of this cohort was performed in the University Medical Center Utrecht (Peper et al., 2009). Exclusion criteria consisted of having any metal material in the head, having a pacemaker, a known history of any major medical condition or psychiatric illness. Zygosity was determined based on DNA polymorphisms, using 8–11 highly polymorphic di-, tri- and tetranucleotide genetic markers and confirmed by genome-wide single nucleotide polymorphism (SNP) data. All subjects and their parents signed informed assents and informed consents, respectively. The study was approved by the Central Committee on Research involving Human Subjects of the Netherlands (CCMO) and was in agreement with the Declaration of Helsinki (Edinburgh amendments).
  - o Imaging: MRI data were collected on a 1.5 Philips Achieva MR scanner (32 diffusion-weighted volumes with different non-collinear diffusion directions with b-factor = 1000 s/mm^2 and 8 diffusion-unweighted volumes with b-factor = 0 s/mm^2; parallel imaging SENSE factor = 2.5; flip angle = 90°; 60 slices of 2.5 mm; no slice gap; 96 × 96 acquisition matrix; reconstruction matrix 128 × 128; FOV = 240 mm; TE = 88 ms; TR = 9822 ms; two repetitions; no cardiac gating; and total scan duration = 296 s). More information may be found in Brouwer et al. (2010, 2012).

ENIGMA-DTI processing

We used the ENIGMA-DTI protocols for multi-site processing and extraction of tract-wise average FA values as described in our prior work (Jahanshad et al., 2013). Briefly, FA images from all subjects were non-linearly registered to ENIGMA target brain using FSL *fnirt* (Smith et al., 2006, 2007) as a default registration. This target was created from the images of the participating studies as the minimal deformation target as previously described (Jahanshad et al., 2013). The data were then processed using the tract-based spatial statistics (*TBSS*) analytic method (Smith et al., 2006, 2007) modified to project individual FA values on the ENIGMA-DTI skeleton. Following the extraction of the skeletonized white matter and projection of individual FA values, ENIGMA-tract-wise regions of interest (ROIs), derived from the Johns Hopkins University (JHU) white matter parcellation atlas (Mori et al., 2008), were transferred to extract the mean FA across the full skeleton and average FA values for twelve major white matter tracts, breaking down the corpus callosum into 3 regions, for a total of 15 ROIs. The regions analyzed are listed in Table 2. The full brain average FA values were calculated to include all voxels in the ENIGMA-DTI skeleton, not just those voxels in the segmented regions of interest, thereby also including peripheral white matter regions. The protocol, target brain, skeleton mask, source code and executables, are publicly available (http://enigma.ini.usc.edu/ongoing/dti-working-group/). Finally, we analyze the voxelwise FA value along the ENIGMA skeleton mask.

Single-site heritability estimation: Analysis of additive genetic variance

The variance components method, as implemented in the Sequential Oligogenic Linkage Analysis Routines (SOLAR) software package (http://www.nitrc.org/projects/se_linux) (Almasy and Blangero, 1998) was used for all individual cohort heritability estimations. Briefly, the algorithms in SOLAR employ maximum likelihood variance decomposition methods and are extensions of the strategy developed by Amos (1994). The covariance matrix Ω for a pedigree of individuals is given by:

$$\Omega = 2 \cdot \Phi \cdot \sigma_g^2 + I \cdot \sigma_e^2 \tag{1}$$

where σ_g^2 is the genetic variance due to the additive genetic factors, Φ is the kinship matrix representing the pair-wise kinship coefficients among all individuals, σ_e^2 is the variance due to individual − specific environmental effects, and I is an identity matrix (under the assumption that all environmental effects are uncorrelated among family

Table 2
Regions of interest (ROIs) examined along the ENIGMA-DTI skeleton as defined by the JHU white matter parcellation atlas (Mori et al., 2008). Analyses for other JHU ROIs are provided in the Supplement.

Average Skeleton FA	
GCC	Genu of the corpus callosum
BCC	Body of corpus callosum
SCC	Splenium of corpus callosum
FX	Fornix
CGC	Cingulum (cingulate gyrus) — L and R combined
CR	*Corona radiata* — L and R anterior and posterior sections combined
EC	External capsule — L and R combined
IC	Internal capsule — L and R anterior limb, posterior limb, and retrolenticular parts combined
IFO	Inferior fronto-occipital fasciculus — L and R combined
PTR	Posterior thalamic radiation — L and R combined
SFO	Superior fronto-occipital fasciculus
SLF	Superior longitudinal fasciculus
SS	Sagittal stratum (includes inferior longitudinal fasciculus and inferior fronto-occipital fasciculus) — L and R combined
CST	Corticospinal tract

members). Narrow sense heritability is defined as the fraction of phenotypic variance σ_P^2 attributable to additive genetic factors,

$$h^2 = \frac{\sigma_g^2}{\sigma_P^2}. \tag{2}$$

The variance parameters are estimated by comparing the observed phenotypic covariance matrix with the covariance matrix predicted by kinship (Almasy and Blangero, 1998). Significance of heritability is tested by comparing the likelihood of the model in which σ_g^2 is constrained to zero with that of a model in which σ_g^2 is estimated. Twice the difference between the log$_e$ likelihoods of these models yields a test statistic, which is asymptotically distributed as a 1/2:1/2 mixture of a χ^2 variable with 1 degree-of-freedom and a point mass at zero. Prior to testing for the significance of heritability, phenotype values for each individual within the cohort were adjusted for covariates including sex, age, age^2, age × sex interaction, and age^2 × sex interaction. Inverse Gaussian transformation was also applied to ensure normality of the measurements. Outputs from SOLAR include the heritability value, the significance value (p), and the standard error for each trait (ROI or voxel).

Mega-and-metagenetic analysis of additive genetic variance

A mega-analysis function '*polyclass*' was recently developed for the software package, SOLAR-Eclipse (Blangero et al., 2005). It was implemented in the LONI Pipeline infrastructure (Dinov et al., 2010) and used for both voxelwise and ROI-level analyses. This function 1) fitted a model for each of the sites separately, and 2) performed the mega-analysis where the data for all populations is fitted together using a single model. The processing steps for this function are shown in Fig. 1. The trait average and variability values can vary by site due to study design and measurement biases as shown by non-overlapping histograms of average FA values (Fig. 1A). Two data normalization steps are performed. In the first, the regression of covariates (as specified above) is performed, per site, and then followed by the per-site inverse Gaussian normalization of data (Fig. 1B). Individual cohort heritability estimates are obtained by fitting a polygenic model per population. The heritability value, its significance, and standard error are tabulated for the meta-genetic analysis.

A mega-genetic analysis refers to analysis multiple cohorts combined into a single pedigree (Jahanshad et al., 2013). In this step, a joint pedigree structure was created by merging the kinship matrices from all cohorts, and a polygenic model was then fit for the combined dataset. Meta-genetic analyses calculate weighted-mean heritability (h^2) and standard error estimates based on measurements from individual cohorts (Li et al., 2003). We chose two methods to calculate the weights: by standard error (Sutton, 2000) and, for comparison with most other meta-analysis studies of heritability, by study sample[2] (Verweij et al., 2010).

We weight the heritability from each study by the heritability standard error, as extracted from the variance component model of SOLAR. The heritability weighted by standard error (Sutton, 2000) is:

$$h_{\text{MA-SE}}^2(v) = \frac{\sum_j se_j^{-2} h_j^2(v)}{\sum_j se_j^{-2}}, \tag{3}$$

where $v = 1$ to N_v indexes voxels or ROIs, and $j = 1, ..., 5$ indexes site.

The mean heritability weighted by sample size is calculated as

$$h_{\text{MA-N}}^2(v) = \frac{\sum_j n_j h_j^2(v)}{\sum_j n_j}. \tag{4}$$

As the heritability estimates for all datasets were computed with SOLAR, the standard error for the heritability at each voxel was also available. Therefore, we computed the joint standard error for each MA as follows:

$$Var\left(h_{\text{MA-SE}}^2(v)\right) = \frac{1}{\sum_j se_j^{-2}(v)}, \tag{5}$$

$$Var\left(h_{\text{MA-N}}^2(v)\right) = \frac{\sum_j n_j^2 se_j^2(v)}{\left(\sum_j n_j\right)^2}. \tag{6}$$

Surprisingly, there is no standard method for calculating the significance, or p-value, for meta-analyses of such variance component estimates. The principal problem is the non-Normality of the heritability estimate, reflected in the fact that a variance component p-value cannot exceed 1/2. To demonstrate some possible approaches, we calculated the meta-analysis Z-score in two ways and compared the results.

The first approach is based on individual cohort p-values; first they are transformed to cohort Z values according to $z_j(v) = \Phi^{-1}(p_j(v))$, where Φ represents the Gaussian normal transformation; if all p-values are strictly less than 1/2, this transformation is valid. These Z values are then combined into meta-analysis Z-scores following (Willer et al., 2010):

$$Z_{\text{MA-SE}}(v) = \frac{\hat{\beta}(v)}{SE\left(\hat{\beta}(v)\right)} = \frac{\sum_j \hat{\beta}_j(v) \times se_j^{-2}(v)/\sum_j se_j^{-2}(v)}{\sqrt{1/\sum_j se_j^{-2}(v)}}, \tag{7}$$

$$Z_{\text{MA-N}}(v) = \frac{\sum_j z_j(v)\sqrt{n_j}}{\sqrt{\sum_j n_j}}, \tag{8}$$

where, for the SE-weighted meta-analysis, each site's effect is estimated as $\hat{\beta}_j(v) = z_j(v) \times se_j(v)$, which can be interpreted as z-scores in h^2 SE units.

In the second approach for calculating p-values, the meta-analysis p-values are calculated from the meta-analysis heritability estimates and their standard error, without consulting individual cohort p-values. Specifically, these were found by computing z-scores for both MA-SE and MA-N methods, found as ratios of the $h_{\text{MA-SE}}^2(v)$ and $h_{\text{MA-N}}^2(v)$ values (Eqs. (3) & (4)) with their standard errors (square root of Eqs. (5) & (6), respectively); from these z-scores, p-values were found based on the asymptotic normality of maximum likelihood estimates (Lehmann and Casella, 1998).

Multiple comparisons correction

Multiple statistical tests performed across voxels and/or ROIs can increase the chance of reporting a false positive finding at a given significance threshold, unless steps are taken to control for multiple comparisons. We use the false discovery rate (FDR) (Benjamini and Hochberg, 1995; Genovese et al., 2002) approach for controlling false positives.
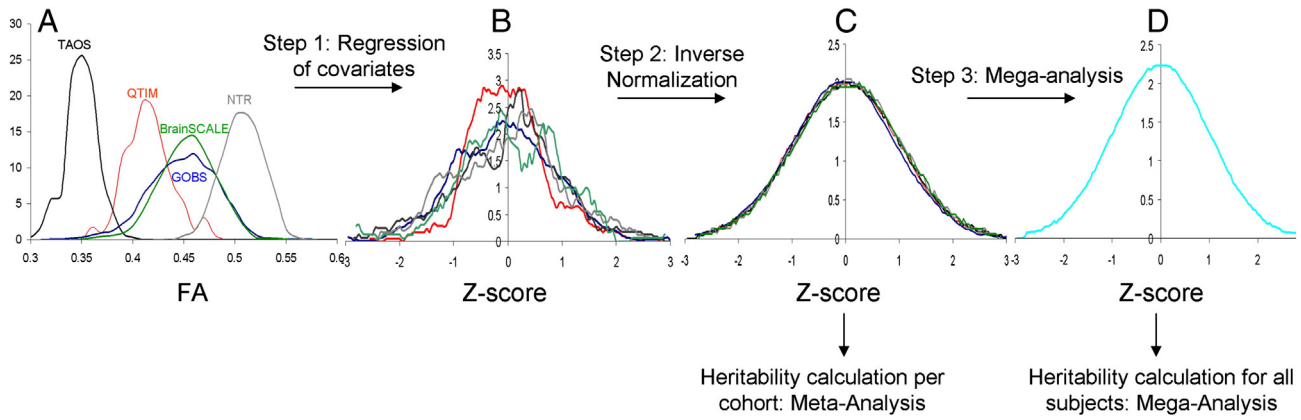
---

[2] The Verweij et al. (2010) study specifically focused on twin studies. Weighting a mix of family studies by sample size should be approached with caution as sample sizes in different family based studies do not reflect the same power. A certain sample size in a twin study may be more powerful in picking up heritability effects than that same sample size in a large extended pedigree due to lower degrees of shared genetic information between more distant relatives. Please see Discussion for a more detailed explanation.

Fig. 1. The mega-analysis workflow is shown. First, covariates are regressed from each site. Next, each site's data undergoes an inverse Gaussian normalization. Finally data are merged for a mega-analysis of heritability estimates using a joint kinship matrix.

Results

Heritability estimates for individual cohorts

Heritability analysis for the whole-brain average FA values demonstrated significant heritability for all five cohorts (Table 3, Fig. 2). The highest heritability estimate was observed for the NTR cohort ($h^2 = 0.84 \pm .03$; $p < 10^{-10}$), while the lowest heritability was observed for the youngest cohort, BrainSCALE ($h^2 = 0.27 \pm .13$; $p = .02$); there were no significant differences between samples in the heritability estimates for the average FA values (1-way ANOVA $p = 0.07$). The heritability for tract-wise average FA measurements were significant for all cohorts with exception of the TAOS cohort where heritability estimates for six tracts (EC, FX, IFO, SFO, SS, CST) had $p > 0.05$. When correcting for multiple comparisons across all regions, the BCC and SCC were also not significant in TAOS. The CST was also not significantly heritable in QTIM or BrainSCALE cohorts; the SFO was alson not significantly heritable in BrainSCALE. Also of note is that with the exception of the two youngest cohorts (TAOS and BrainSCALE), the standard errors for regional measurements were similar for the three remaining cohorts (e.g., 0.06, 0.07, 0.07, vs 0.12 and 0.26 for GOBS, QTIM, NTR, vs. BrainSCALE and TAOS, for average FA respectively).

The significance of voxelwise measurements of heritability varied substantially per cohort (Table 3). Following multiple comparisons correction across all voxels using FDR, we observed that the QTIM cohort had the highest fraction (79%) of significant voxels, while TAOS cohort had the lowest (1.6%) fraction (Table 4).

Meta- and Mega-analysis

The results for the three approaches for pooling data and obtaining overall estimates of heritability are shown in Table 3 and Fig. 3. The SE-weighted meta-analysis yielded the highest estimate ($h^2 = 0.71 \pm 0.06$; $p < 10^{-10}$) for the average FA across the whole skeleton, while the N-weighted meta-analysis produced the lowest ($h^2 = 0.51 \pm 0.06$; $p < 10^{-10}$) heritability estimate. The mega-genetic approach produced an intermediate heritability estimate but with the lowest standard error and the highest significance ($h^2 = 0.67 \pm 0.04$; $p < 10^{-10}$). The tract-wise heritability estimates for the SE-weighted meta- with mega-genetic analyses were highly correlated ($r = 0.76$, $p = 0.002$) (Table 5), however the correlations of the tract-wise estimates for the N-weighted approach with the SE-weighted approach and with the mega-analytical approach, were not significant ($r = 0.48$, $p = 0.11$; $r = 0.41$, $p = 0.15$).

The tract-wise heritability estimates of the N-weighted meta-analysis was highly correlated with the estimates from the TAOS sample ($r = 0.84$, $p = 0.0006$), while the results of the SE-weighted meta-analysis was strongly correlated with the NTR cohort ($r = 0.89$, $p = 0.0001$). The ROI results of the mega-analysis were more correlated with QTIM, NTR, and BrainSCALE ($r = 0.74, 0.54$ and 0.50, respectively).

Voxelwise, all three data pooling analytic approaches produced a similar spatial pattern of heritability (Fig. 4). The meta-analysis weighted by N (*top*) showed the lowest overall voxelwise heritability. The meta-analysis weighted by SE (*middle*) showed a more similar pattern to the mega-analysis (*bottom*), although the mega-analysis appeared to have a smoother spatial distribution of heritability. Notably, all three methods produced a significant number of voxel-wise estimates that would pass the most stringent, Bonferroni, correction for multiple testing (Table 6). Both methods of calculating p-values for meta-analysis produced a significant number of heritable voxels. The SE-MAs included many voxels whose p-values were exceedingly low (0). Despite this, the mega-analytic analysis was the only one to show a majority of voxels surviving Bonferroni correction.

We note a high voxel-wise correlation of heritability values 0.87–0.90 between the different joint approaches (Fig. 5). Meta-analyses show a correlation of 0.87. Additionally, when the mega-analysis is compared to the SE-weighted meta-analysis, a similar high correlation of 0.87 is obtained (Fig. 5b). A slightly higher correlation is observed when comparing the mega-analysis to the meta-analysis weighted by N (Fig. 5c).

Leave-one-cohort out analysis

Finally, the stability of the joint heritability estimates was ascertained in the regions of interest using a leave-one-out analysis. The aim was to study how the joint heritability estimate depended on the contribution from any one dataset. For the whole brain average FA values, removing one cohort resulted in different results for each type of joint analysis. On average, the SE-weighted meta-analysis approach produced the highest heritability estimate, yet with the widest range of results $h^2 = 0.69$ (for average FA for example; range 0.51 to 0.74) while the N-weighted meta-analysis approach produced the lowest heritability estimate $h^2 = 0.52$ with a narrower range of results (range 0.47 to 0.57). For both meta-analytic approaches, the highest h^2 estimate was observed when leaving out the GOBS cohort and the lowest was observed following removal of the NTR cohort. However, for the SE-weighted meta-analysis, the removal of the NTR cohort led to a substantially different heritability estimate than removal of each of the others, while for the N-weighted meta-analysis, removal of NTR was similar to removing QTIM, and differences in estimations were more subtle. Mega-analysis produced an intermediate estimate $h^2 = 0.67$ (cumulative h^2 estimates ranging from 0.58 to 0.77). The highest h^2 estimate was observed following removal of the BrainSCALE cohort and the lowest following removal of the QTIM cohort (Fig. 6).

Table 3

Individual cohort heritability results (and significance (p) values) for each ROI are listed along with the results for the joint analytic approaches. Meta-analysis p-values were computed in two ways as described in the Methods section, both of which are reported here. The first p-value presented in the MA columns uses individual cohort level data while the second is found directly from the MA heritability and SE values.

	GOBS	QTIM	TAOS	BrainSCALE	NTR	Metaanalysis (SE)	Metaanalysis (N)	Megaanalysis
Whole-brain	0.41 ± 0.08 (5.98E-09)	$0.61 \pm .05$ (p < 10^{-10})	0.49 ± 0.24 (0.04)	0.27 ± 0.13 (0.02)	0.84 ± 0.03 (p < 10^{-10})	0.71 ± 0.03 (0, p < 10^{-10})	0.51 ± 0.05 (p < 10^{-10}, p < 10^{-10})	0.67 ± 0.03 (p < 10^{-10})
BCC	0.49 ± 0.08 (p < 10^{-10})	0.69 ± 0.05 (p < 10^{-10})	0.69 ± 0.19 (0.004)	0.49 ± 0.11 (7.70E-05)	0.72 ± 0.05 (p < 10^{-10})	0.65 ± 0.03 (0, p < 10^{-10})	0.60 ± 0.04 (p < 10^{-10}, p < 10^{-10})	0.69 ± 0.03 (p < 10^{-10})
GCC	0.48 ± 0.09 (5.62E-10)	0.59 ± 0.06 (p < 10^{-10})	0.82 ± 0.13 (4.01E-04)	0.40 ± 0.12 (9.54E-04)	0.84 ± 0.03 (p < 10^{-10})	0.75 ± 0.02 (0, p < 10^{-10})	0.60 ± 0.04 (p < 10^{-10}, p < 10^{-10})	0.66 ± 0.03 (p < 10^{-10})
SCC	0.56 ± 0.09 (p < 10^{-10})	0.57 ± 0.06 (p < 10^{-10})	0.52 ± 0.17 (0.007)	0.37 ± 0.12 (0.002)	$.78 \pm .05$ (p < 10^{-10})	0.66 ± 0.03 (0, p < 10^{-10})	0.57 ± 0.05 (p < 10^{-10}, p < 10^{-10})	0.62 ± 0.03 (p < 10^{-10})
FX	0.21 ± 0.08 (8.23E-04)	0.62 ± 0.06 (p < 10^{-10})	0.50 ± 0.29 (0.077)	0.45 ± 0.13 (0.001)	0.53 ± 0.08 (2.29E07)	0.49 ± 0.04 (0, p < 10^{-10})	0.43 ± 0.05 (p < 10^{-10}, p < 10^{-10})	0.56 ± 0.04 (p < 10^{-10})
CGC	0.41 ± 0.08 (2.16E-08)	0.65 ± 0.05 (p < 10^{-10})	0.78 ± 0.20 (0.002)	0.47 ± 0.11 (1.14E-04)	0.76 ± 0.05 (p < 10^{-10})	0.66 ± 0.03 (0, p < 10^{-10})	0.58 ± 0.05 (p < 10^{-10}, p < 10^{-10})	0.63 ± 0.03 (p < 10^{-10})
CR	0.50 ± 0.08 (p < 10^{-10})	0.77 ± 0.04 (p < 10^{-10})	0.76 ± 0.16 (.006)	0.49 ± 0.12 (3.03E-04)	0.77 ± 0.05 (p < 10^{-10})	0.73 ± 0.03 (0, p < 10^{-10})	0.64 ± 0.04 (p < 10^{-10}, p < 10^{-10})	0.75 ± 0.03 (p < 10^{-10})
EC	0.49 ± 0.08 (4.03E-10)	0.78 ± 0.03 (p < 10^{-10})	0.02 ± 0.37 (0.40)	0.46 ± 0.13 (7.68E-04)	0.83 ± 0.03 (p < 10^{-10})	0.77 ± 0.02 (0, p < 10^{-10})	0.54 ± 0.06 (p < 10^{-10}, p < 10^{-10})	0.75 ± 0.03 (p < 10^{-10})
IC	0.46 ± 0.08 (5.55E-10)	0.67 ± 0.05 (p < 10^{-10})	0.62 ± 0.21 (0.02)	0.48 ± 0.11 (1.26E-04)	0.81 ± 0.04 (p < 10^{-10})	0.71 ± 0.03 (0, p < 10^{-10})	0.58 ± 0.05 (p < 10^{-10}, p < 10^{-10})	0.71 ± 0.03 (p < 10^{-10})
IFO	0.37 ± 0.09 (1.54E-06)	0.68 ± 0.05 (p < 10^{-10})	0.26 ± 0.44 (0.22)	0.65 ± 0.08 (4.26E-08)	0.81 ± 0.04 (p < 10^{-10})	0.71 ± 0.03 (0, p < 10^{-10})	0.51 ± 0.07 (p < 10^{-10}, p < 10^{-10})	0.73 ± 0.03 (p < 10^{-10})
PTR	0.38 ± 0.08 (2.48E-09)	0.72 ± 0.04 (p < 10^{-10})	0.48 ± 0.19 (0.01)	0.39 ± 0.12 (0.002)	0.86 ± 0.03 (p < 10^{-10})	0.76 ± 0.02 (0, p < 10^{-10})	0.54 ± 0.04 (p < 10^{-10}, p < 10^{-10})	0.69 ± 0.03 (p < 10^{-10})
SFO	0.36 ± 0.08 (1.16E-07)	0.67 ± 0.05 (p < 10^{-10})	0.29 ± 0.23 (0.10)	0.39 ± 0.13 (0.02)	0.73 ± 0.05 (p < 10^{-10})	0.62 ± 0.03 (0, p < 10^{-10})	0.48 ± 0.05 (p < 10^{-10}, p < 10^{-10})	0.64 ± 0.03 (p < 10^{-10})
SLF	0.57 ± 0.09 (p < 10^{-10})	0.78 ± 0.03 (p < 10^{-10})	0.51 ± 0.29 (0.05)	0.41 ± 0.13 (0.002)	0.88 ± 0.03 (p < 10^{-10})	0.82 ± 0.02 (0,0)	0.64 ± 0.06 (p < 10^{-10}, p < 10^{-10})	0.77 ± 0.02 (p < 10^{-10})
SS	0.44 ± 0.08 (1.04E-09)	0.74 ± 0.04 (p < 10^{-10})	0.05 ± 0.31 (0.50)	0.25 ± 0.14 (0.03)	0.70 ± 0.06 (p < 10^{-10})	0.66 ± 0.03 (0, p < 10^{-10})	0.48 ± 0.06 (p < 10^{-10}, p < 10^{-10})	0.65 ± 0.03 (p < 10^{-10})
CST	0.33 ± 0.08 (1.72E-06)	0.25 ± 0.10 (0.008)	0.29 ± 0.39 (0.23)	0.37 ± 0.13 (0.004)	0.61 ± 0.06 (p < 10^{-10})	0.44 ± 0.04 (0, p < 10^{-10})	0.33 ± 0.07 (p < 10^{-10},2.54E-06)	0.42 ± 0.04 (p < 10^{-10})

The tract-wise, leave-one-cohort-out analysis produced mainly stable results consistent with the full group analysis, with one notable exception. The N-weighted and mega-analyses produced consistent estimates that followed the trend of the full group analysis ($r = 0.91$–0.99) (Fig. 7). The SE-weighted meta-analysis however, appeared to be most influenced by exclusion of the NTR dataset, which had the highest individual heritability and lowest SE. Removal of that dataset led to substantially lower heritability estimates across the ROIs and lowered the correlation with the results from the complete dataset ($r = 0.79$). Another notable observation is that the mega-analytic approach (Fig. 7) produced generally higher heritability estimates with a narrower error margin compared to both the meta analytical methods. Genetic analyses for other regions of the Johns Hopkins DTI atlas are provided in supplement (see Supplement).
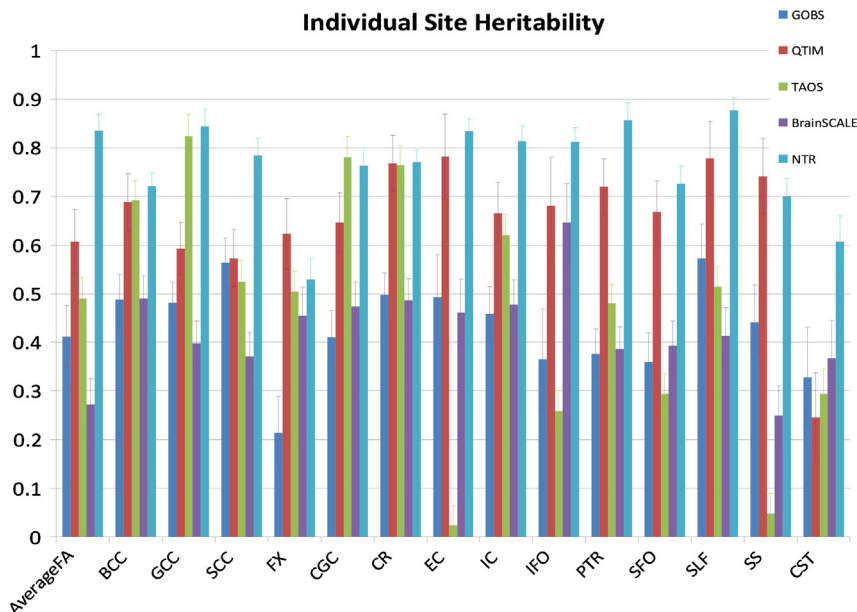

Fig. 2. Individual site heritability estimates and standard error are plotted for each of 15 regions of interest.

Table 4
Voxelwise multiple comparisons correction using FDR was performed for each heritability map based on the resulting p-values. The percent of the skeleton that was significantly heritable, along with the FDR critical p-value and minimum p-value along the image are listed.

	% Significantly heritable (FDR)[a]	FDR critical p-value	Minimum p-value
GOBS	59.09	0.0295	$p < 10^{-10}$
QTIM	79.10	0.0395	$p < 10^{-10}$
TAOS	.61	0.0008	3.0×10^{-7}
BrainSCALE	16.36	0.0082	2.1×10^{-10}
NTR	40.22	0.0201	$p < 10^{-10}$

[a] Total voxels in mask = 116524.

Discussion

We evaluated three data-pooling strategies for performing a multi-site analysis of additive genetic contributions to variability in the fractional anisotropy (FA) of cerebral white matter (WM). The ENIGMA-DTI data harmonization protocol, presented in our prior work (Jahanshad et al., 2013) served as the foundation for this analysis.

This protocol was used to assemble and homogenize data collected at five sites, using diverse imaging protocols and cohorts that varied in age (9–85 years) and the family structure (twins, twins and siblings, and extended family). Our efforts aim to identify FA measurements that are significantly heritable regardless of age, ethnicity, family structure, and methodological variations. To summarize our findings, we observed that (as expected) the heritability estimates do vary across cohorts. Datasets with smaller numbers of subjects, or less familial relationships, were more likely to produce heritability estimates not representative of the pooled trend. Nonetheless, including these datasets improves the overall outcome of the joint analysis. We also observe that three analytical approaches for data pooling vary in both the estimates and the significance of pooled heritability and these estimates may be sensitive to inclusion/exclusion of specific datasets. Overall, our study concluded that despite methodological differences, combining genetically informative datasets using meta-and/or-mega-genetic approaches led to a significant improvement of the statistical significance. Overall, the pooled heritability estimates across the brain showed high heritability, with additive genetic factors explaining over 50% of intersubject variance in FA values across most regions.
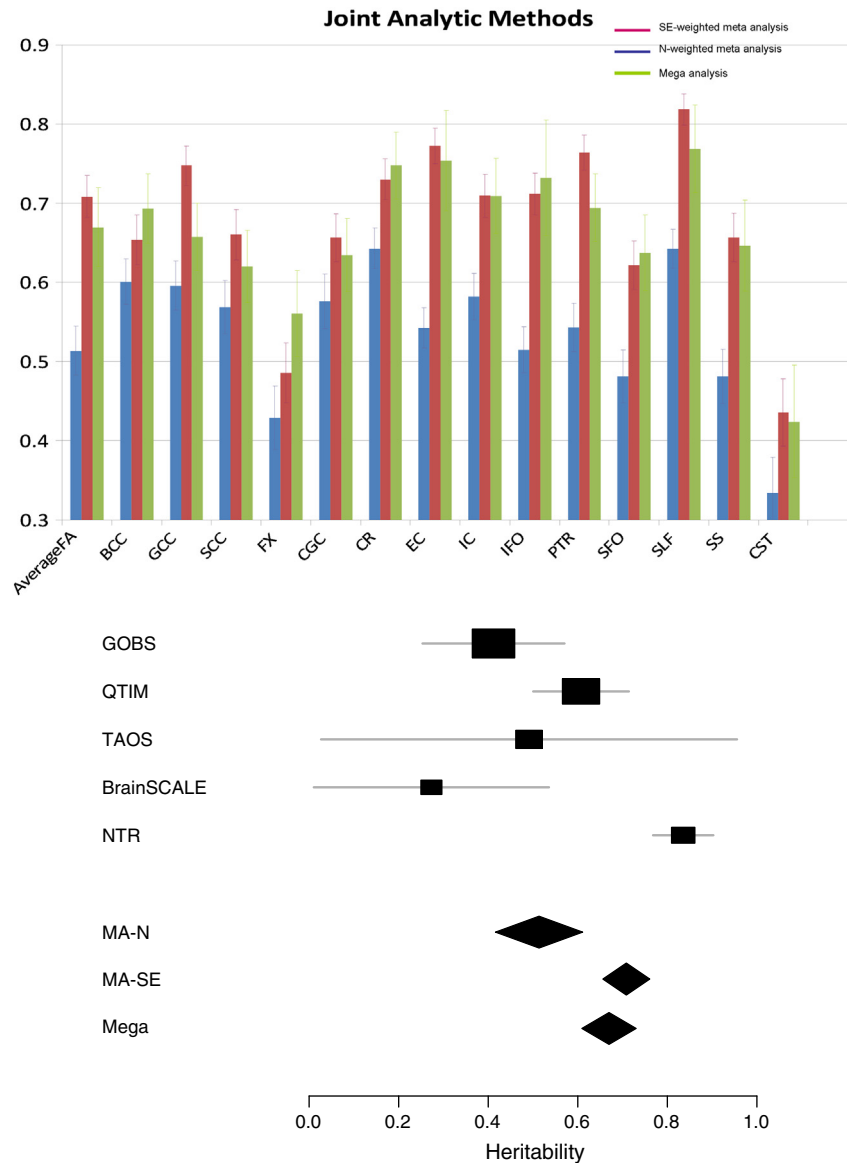

Fig. 3. Top: Heritability estimates for meta and mega analytical approaches combining the 5 cohorts. Bottom: A forest plot of the average FA value (h2 +/− the 95% CI) is shown highlighting each cohort and all joint models.

Table 5
Result of the linear correlation analysis (r-coefficients) between tract-wise heritability measurements for three data pooling methods and five cohorts. The tract-wise heritability estimates of the N-weighted meta-analysis was highly correlated with the estimates from the TAOS sample ($r = 0.84$, p = 0.0006), while the results of the SE-weighted meta-analysis was strongly correlated with the NTR cohort ($r = 0.89$, p = 0.0001). The ROI results of the mega-analysis were more correlated with QTIM, NTR, and BrainSCALE ($r = 0.74$, 0.54 and 0.50, respectively).

	GOBS	QTIM	TAOS	Brain-SCALE	NTR	Metaanalysis (SE)	Metaanalysis (N)	Megaanalysis
GOBS		0.11	0.23	−0.21	0.24	0.36	0.49	0.24
QTIM	0.11		−0.44	−0.01	0.11	0.50	0.00	0.73
TAOS	0.23	−0.44		0.20	0.13	0.00	0.84	0.11
BrainSCALE	−0.21	−0.01	0.20		0.23	0.13	0.34	0.50
NTR	0.24	0.11	0.13	0.23		0.89	0.46	0.54
Metaanalysis (SE)	0.36	0.50	0.00	0.13	0.89		0.48	0.76
Metaanalysis (N)	0.49	0.00	0.84	0.34	0.46	0.48		0.41
Megaanalysis	0.24	0.73	0.11	0.50	0.54	0.76	0.41	

This manuscript extends and validates the ENIGMA-DTI harmonization protocol (Jahanshad et al., 2013) to a wider range of cohorts. This protocol ensures that imaging data from multiple sites can be harmonized for a pooled analysis by transforming them in the common space (a custom spatial normalization template derived from four diverse high-resolution datasets) and extracting regional measures; it includes the processing steps necessary to prepare DTI data for quantitative genetic analysis, including extraction of tract-wise average and voxel-wise FA data, as well as data quality control and assurance procedures. Ideally, harmonization of data collection should occur at the onset of the multicenter imaging study as demonstrated by the efforts of the Alzheimer's Disease Neuroimaging Initiative (Weiner et al., 2012) and others (Teipel et al., 2011; Zhan et al., 2013). Clearly, this is only applicable to data that has yet to be collected. The ENIGMA-DTI harmonization effort was developed as a *post hoc* solution to help prepare existing data for pooled genetic analyses to facilitate the search for genetic associations with DTI-extracted traits. This manuscript evaluated the ENIGMA-DTI protocol by testing its ability to reproduce and replicate findings of significant additive genetic contribution to FA values in data collected at five sites.

The contribution of additive genetic variation (i.e., heritability) to individual variability was calculated for the whole brain skeleton and regional FA values in five cohorts. The five cohorts were diverse in both the age of participants (9–85 years) and family structure (twin, twins + siblings, extended families). Heritability is a property of a population sample and may vary from sample to sample. It should not serve as the single parameter to estimate and compare across various populations. However, for certain planned collaborative efforts to search the genome and discover individual variants contributing to basic white matter properties, it is desirable to establish patterns of heritability that are common regardless of the population under study. Despite differences in age, structure and ethnicity among our cohorts, there was no
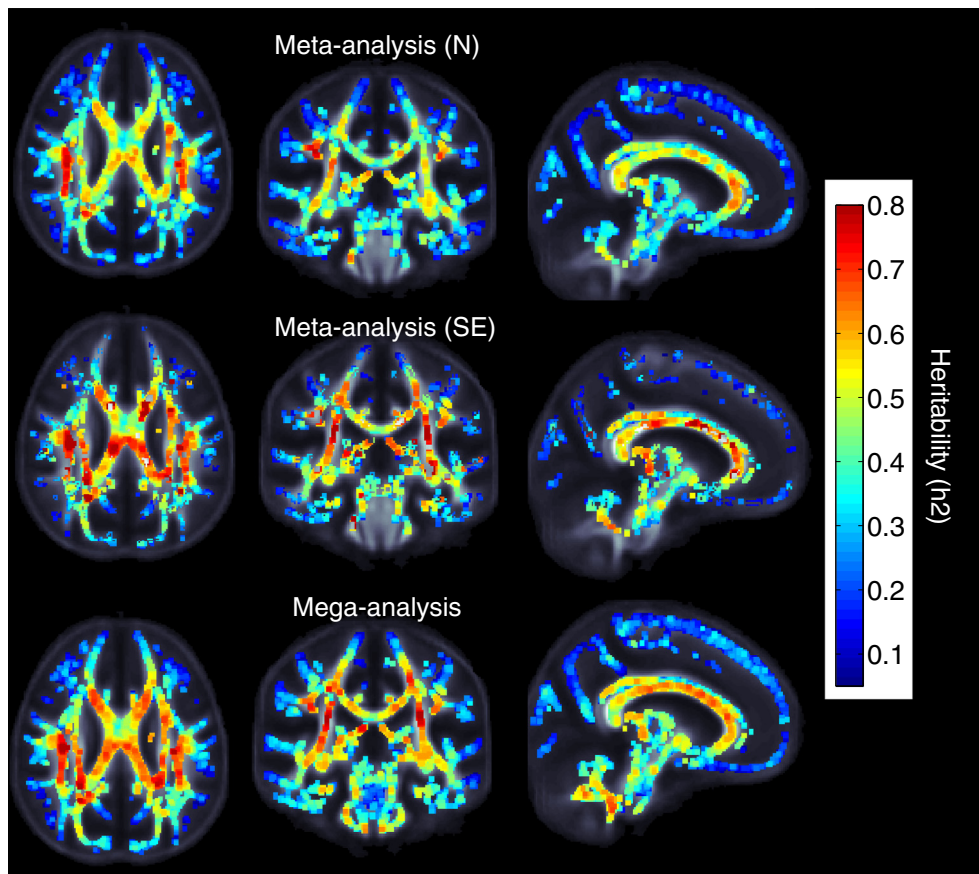

Fig. 4. Voxelwise maps of meta and mega–analysis of heritability. The meta-analysis weighted by N (top) showed the lowest overall voxelwise heritability. The meta-analysis weighted by SE (middle) showed a more similar pattern to the mega-analysis (bottom), although the mega-analysis appeared to have a smoother spatial distribution of heritability.

Table 6

Multiple comparisons corrections for joint analyses using the most stringent Bonferroni correction. All three methods have produced a significant number of voxel-wise estimates that would pass the most stringent, Bonferroni, correction for multiple testing. Meta-analysis p-values were computed in two ways as described in the Methods section, and values for both corrections are listed below.

	% Significantly heritable[a]	Bonferroni critical p-value[a]	Minimum p-value
Meta-SE	34.87, 49.86	$4.3 * 10^{-7}$	$p < 10^{-10}$
Meta-N	10.15, 35.75	$4.3 * 10^{-7}$	$p < 10^{-10}$
Mega	53.1	$4.3 * 10^{-7}$	$p < 10^{-10}$

[a] Total voxels in mask = 116,524.

significant difference in the average heritability estimates. While this does not signify that the same genes contribute to the level of heritability in the different populations, we can be confident that each of the regions has the potential to serve as a promising endophenotype to discover genes that exert controls over white matter structure in general. Our experiments demonstrate that heritability estimates from smaller cohorts with non-twin sibling family structures (TAOS) may be underpowered, particularly for smaller regions, and possibly not be representative of the group at large compared to those of extended pedigrees or primarily twin based cohorts. Despite these variations, we show joint

analytic procedures have limited sensitivity to such differences and can still benefit from the addition of individually underpowered cohorts.

The cohorts that yielded the highest and the lowest heritability estimates for the full skeleton (NTR and BrainSCALE, respectively) were twin samples recruited from the same twin register in The Netherlands. Moreover, the two populations were imaged using similar DTI-protocols implemented on Philips MRI scanners. The low heritability estimate in the BrainSCALE cohort may be due to a combination of factors. Its original study suggested that the low heritability estimate might have been due to the young age of the sample (average age = 9.2 ± 0.1 years) (Brouwer et al., 2010). It hypothesized that the ongoing cerebral myelination may lead to higher intersubject variance, thereby reducing heritability estimates. Unsurprisingly, the subject's age was not a significant covariate in this cohort (p > 0.5) and this may be responsible for the inflation of the overall variance in FA values (Brouwer et al., 2010). While potentially masked by differences is pedigree structure, the trends in heritability across cohorts do not appear to be due to the young age in the cohort. The next youngest cohort, TAOS, was also collected in adolescent subjects with a narrow age window (average age = 13.4 years $\pm$ 0.96) where the subject's age was also a non-significant covariate (p = 0.41). However, the heritability estimates from this cohort were nearly twice that observed in the
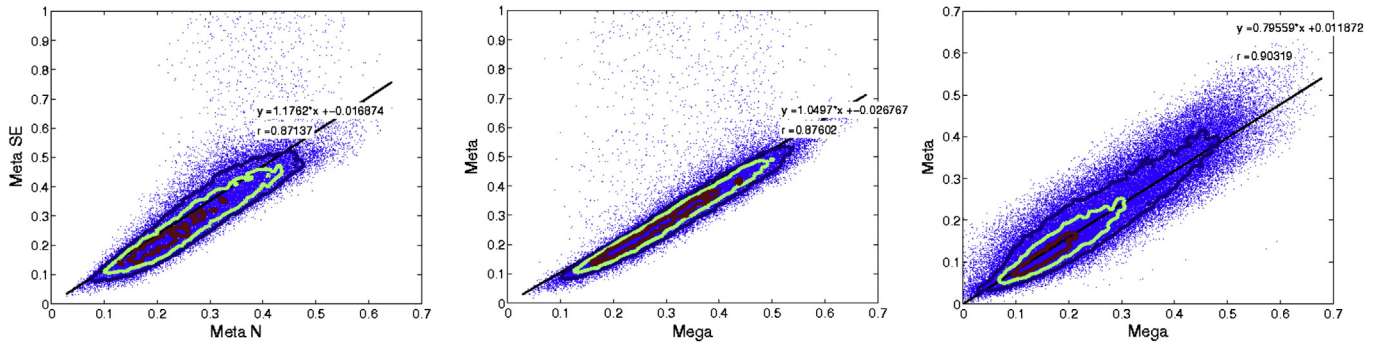

Fig. 5. Scatter plots of voxel-wise heritability estimates for three data pooling methods. The highest correlation was observed between SE-weighted meta and mega genetic analyses (r = 0.90). The lowest was observed for the SE-and N-weighted approaches (r = 0.87).
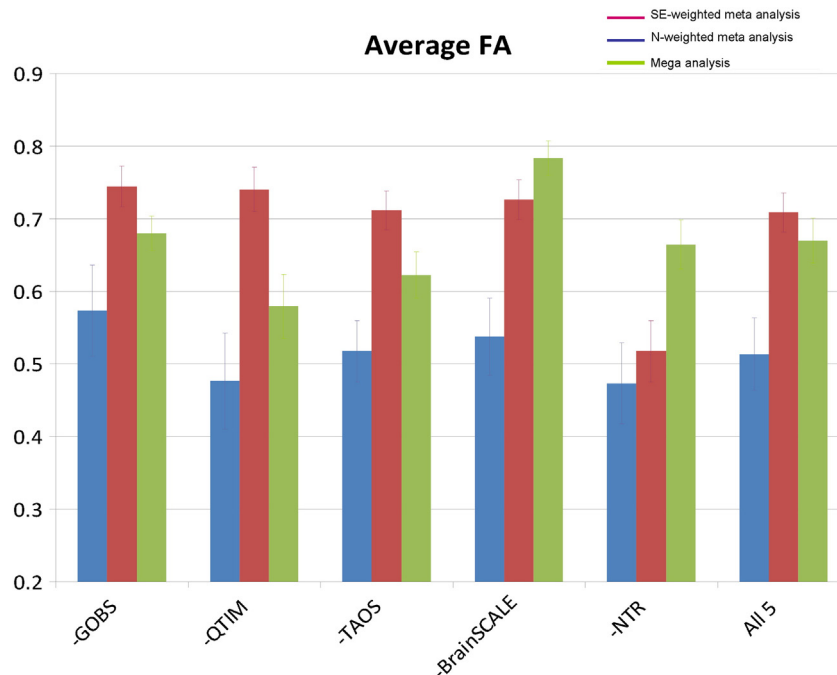

Fig. 6. Heritability and standard error results of the leave-one-out analysis for three data pooling approaches. The x-axis shows the cohort that was left out of the analysis.

BrainSCALE sample ($h^2 = 0.49$ vs. 0.27). The heritability estimates in the TAOS cohort had a low significance. This is because it was composed of siblings and therefore had a lower statistical power compared to twin cohorts. Clearly, this topic of age-by-heritability change during cerebral development deserves further study. These said, in this study, the variation in heritability estimates for cohorts with small subject number is likely to be caused by subject selection rather than methodological issues and/or biases.

The three data pooling methods succeeded in producing similar estimates of heritability using data from all five cohorts. However, there were a few notable differences. The two-meta-analytical approaches produced the lowest and the highest heritability estimate for the average FA values ($h^2 = 0.51$, $p < 10^{-10}$ and 0.71, $p < 10^{-10}$ for N- and SE-weighted approaches, respectively), while mega-genetic analysis produced intermediate heritability estimates with the highest significance ($h^2 = 0.67$, $p = <10^{-10}$). A similar pattern was observed for the tract-wise and voxel-wise measurements: over 50% of the voxel-wise estimates obtained by the mega-genetic approach satisfied even the strictest Bonferroni threshold for multiple comparisons. Importantly, the voxel-wise correlation of heritability estimates showed high levels of shared regional variance ($r = 0.87$–0.90), suggesting a high concordance among the three data-pooling approaches.
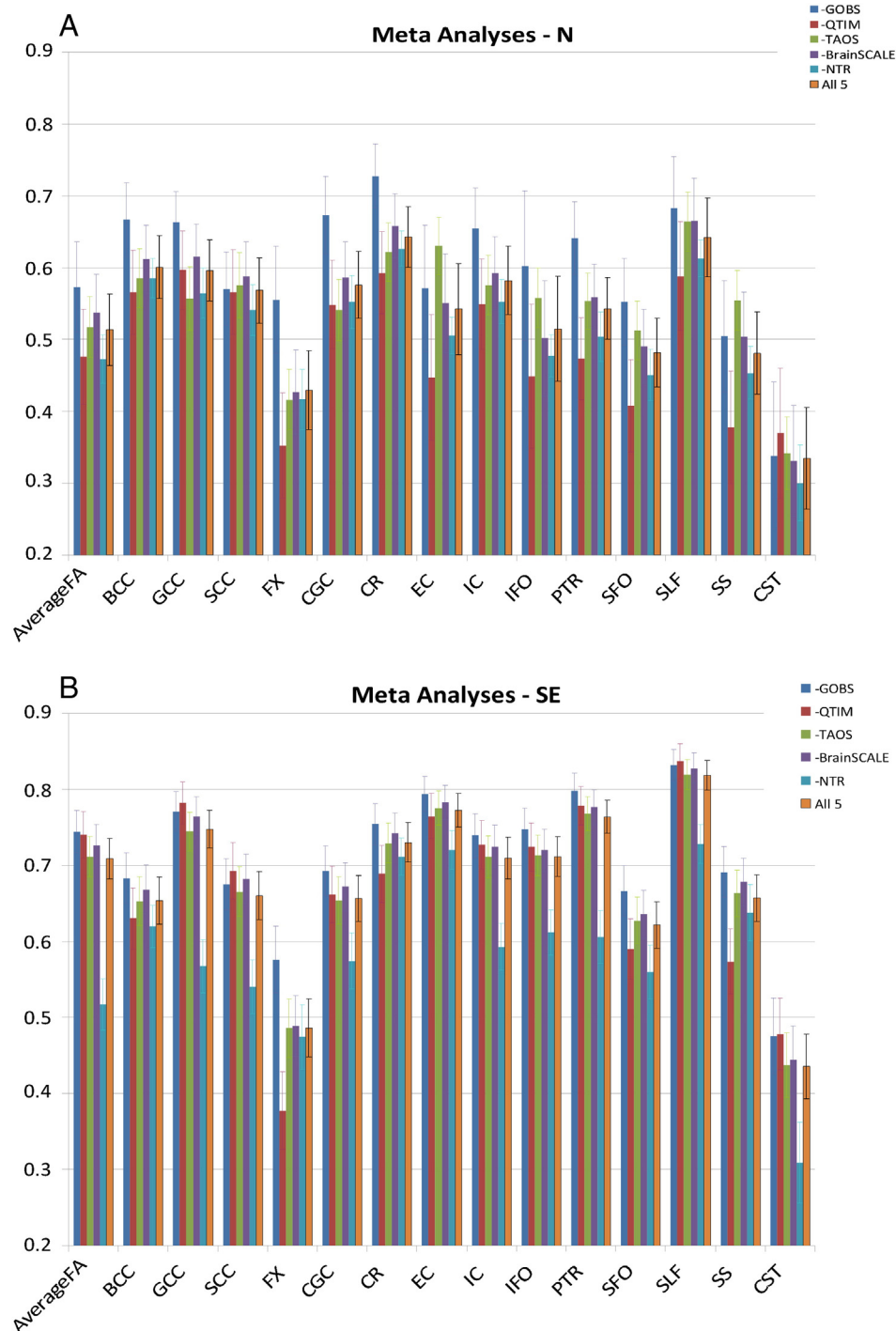
Fig. 7. Results of the leave-one-cohort out analysis for three data pooling methods. The SE-weighted meta analysis is most severely affected by the removal of the NTR cohort, which individually, had the highest heritability estimate and lowest SE. The N-weighted meta-analysis has the widest error-bars of the three methods.
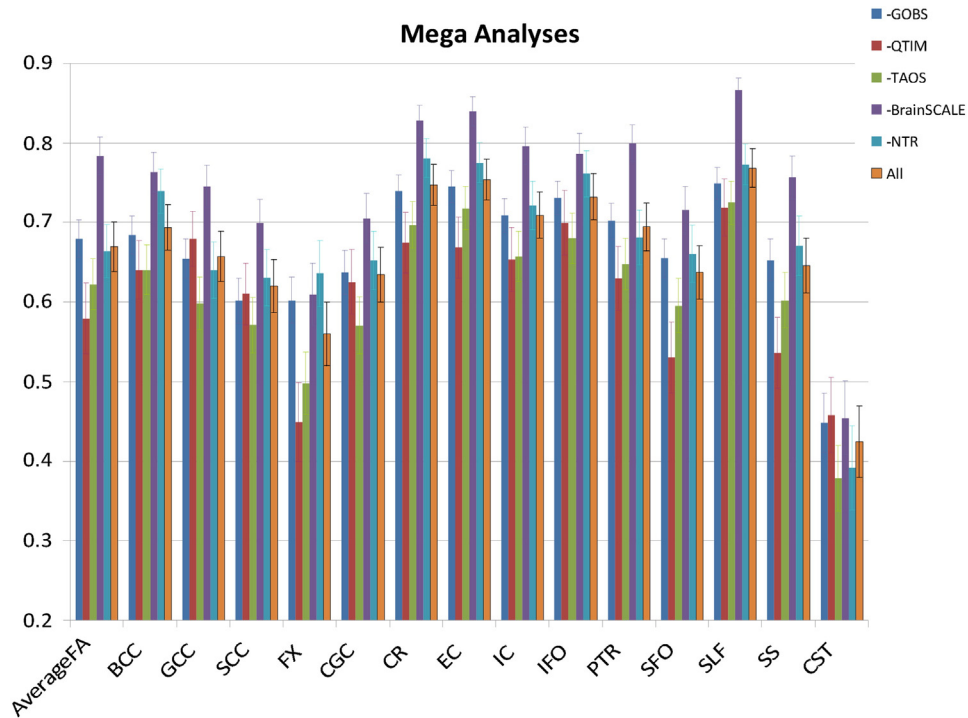
Fig. 7 (*continued*).

Importantly, aggregated heritability estimates presented here showed a high level of concordance with previous estimates calculated in the GOBS and only the twin pairs from the QTIM cohorts (N = 1151) (Jahanshad et al., 2013). For instance, the average FA heritability estimate by the mega-genetic approach was nearly identical to a mega-genetic estimate obtained in the prior study $h^2 = 0.67$, p = 7.6 $\cdot 10^{-49}$ vs. $h^2 = 0.68$, p = $1.9 \cdot 10^{-11}$, for present and past mega-genetic estimates. While the present meta-genetic heritability estimates here were higher than these obtained in the previous sample ($h^2 = 0.51$ and 0.71 vs. 0.42 and 0.43 for the N- and SE-weighted estimates for present and previous studies) (Jahanshad et al., 2013), the by-tract correlations of current (N =2203) vs past (N = 1151) analysis showed high consistency ($r = 0.70$–.86; p ≤ 0.01) for all three data pooling methods (Table 7).

Meta-analysis methods approximate global effect-size by computing a weighted estimate of the effect-sizes from individual studies. The weighs, or scaling factors, for effect sizes are assigned based on the precision of the effect size estimates per study. Studies with more precise estimate of the effect size may therefore be given higher weights. In general, the precision of the effect-size is directly related to the study's sample size (N). The sample size, or N-weighted, meta-analysis is, therefore, a commonly used approach. On the other hand, the design of genetic studies also plays an important role in affecting the precision of the estimate (Borenstein et al., 2009). Particularly in the case of variance-component derived heritability estimates, different family based study designs may have different correlation structures such that the sample size may not directly reflect the estimate precision. For instance, matched twin studies provide more precise estimates of heritability than populations with the same number of more distantly related relatives. On the other hand, the latter type provides a better estimate of covariate effects and thereby may provide more precise estimates of genetic effects (Blangero et al., 2013).

The N-weighted meta-analytic approach used in this study was developed for meta-analysis of twin data (Li et al., 2003; Rhee and Waldman, 2002; Verweij et al., 2010). Here it was used for a meta-analysis of both twin and family population. This may be a limitation as scaling of statistical power per subject may be different for family-based versus twin-based studies. However, we believe that in our case it is a minor limitation. To test this, we used power calculation methods implemented in SOLAR-Eclipse (Blangero et al., 2013) to model the statistical power for detecting heritability in each of the five samples. We observed that for the range of expected heritability estimates ($h^2 = 0.3$–0.7), the statistical power per cohort was directly proportional to the sample size (r = ~0.9), supporting the basic assumption of N-weighted meta-analysis. Further, we compared the estimates from N-weighted approach to these derived using the standard-error weighted meta and mega-genetic analyses. The estimates derived from the N-weighted meta-analysis overlapped with the confidence intervals of the other approaches, suggesting that it homed in on appropriate heritability estimates. Furthermore, a single weighting scheme of a meta-analysis may not be available for every condition and it is also possible that the difference weighting may have minimal effects (Brannick et al., 2011). Nonetheless, this limitation signifies the urgent need for development of more universal meta-genetic analytical approaches for an effective sample-size estimate based on within and between family variance. For instance, simulation modeling of statistical power based on the study designs (Blangero et al., 2013) may become a promising novel approach to derive the scaling factors.

The leave-one-out analysis was performed to test the stability of the pooled heritability estimates on inclusion and exclusion of specific cohorts. The average FA heritability estimates following the removal of each of the five cohorts were similar to the overall pooled heritability estimate for the both the N-weighted meta- and mega-genetic analyses ($h^2 = 0.51$ and 0.67, respectively). The N-weighted meta-analysis showed the smallest variability in heritability estimates because the

Table 7
Correlation between our past (Jahanshad et al., 2013) and present by-tract heritability estimates.

Past estimates	Present N-weighted	Present SE-weighted	Present mega
N-weighted	0.77	0.86	0.80
SE-weighted	0.69	0.86	0.86
Mega	0.70	0.84	0.86

cohorts with the highest and lowest heritability estimates (NTR and BrainSCALE) had the two smallest numbers of subjects and this limited their contribution to this pooled heritability estimate. Nonetheless, the N-weighted meta-analysis was shown to be most sensitive to removal of the GOBS cohort (Fig. 7A). Removal of this cohort led to a change in the heritability estimate from $h^2 = 0.51$ to 0.57. This was expected since the GOBS cohort had the largest number of subjects and thereby exerted the highest weight in the N-weighted combined heritability estimate. On the other hand, the SE-weighted approach was sensitive to removal of the NTR cohort. The removal of this cohort reduced the pooled MA-SE heritability estimate from $h^2 = 0.71$ to 0.51. This led to the highest standard deviation in pooled heritability estimates among the three methods (std dev = 0.10, 0.07 and 0.05 for SE-weighted, mega and N-weighted approaches, respectively). This was also expected, as the NTR cohort had the highest heritability estimates with the lowest standard error. The highest change in mega-genetic analysis $h^2 = 0.67$ to 0.78 (for average FA) was observed following the exclusion of the BrainSCALE cohort due to its lowest heritability estimates and its twin-design that produced population structure with high statistical power. While overall, the pooled estimates are exceedingly similar across regions of interest and voxelwise FA measurements, when data-sharing is not a viable option for mega-analysis, careful consideration of cohort structure and inherent power should be taken into account for meta-analytic approaches.

Overall, we observed that regional heritability estimates varied among individual cohorts and to a lesser degree between the joint analysis methods. As the heritability of a phenotype is dependent on the study population, we would not expect to see the same heritability estimates across geographically and ethnically diverse cohorts. Pooling diverse datasets and populations through joint analytical approach of variance-component analysis provides global means of heritability across a variety of populations. As no standards exist for jointly analyzing such data, we provided results from three different types of analyses and observe the variability between these joint estimators. In addition, we used leave-one out analyses to evaluate stability of each method and its sensitivity to individual cohorts. These stability tests suggest that adequately powered independent studies will likely produce heritability estimates that will fall within the range derived from the set of joint analyses. However, given that only 5 cohorts were used in this study and 4 out of 5 were of children or young adults, and common environmental effects were not assessed, it is also possible that heritability estimates of an independent cohort will be significantly different. The ENIGMA-DTI protocols available online (http://enigma.ini.usc.edu/ongoing/dti-working-group/), and the joint analytical methods described here, may now be used by others to expand upon the joint heritability analyses. Ideally, with enough cohorts, joint estimation of heritability will be analyzed using random-effects modeling (Normand, 1999) to further account for underlying differences and similarities between cohorts.

Conclusion

Here we test the reliability of the FA endophenotypes measured from DTI data and extensively evaluated three data pooling approaches to study the additive genetic contribution to variability in FA. This step of optimizing methods and protocols brings us closer to fulfilling the overall aim of the ENIGMA Consortium — using imaging phenotypes studies to help discover genes that associate with brain structure and may in turn be associated with, behavior, cognition, and neuropsychiatric disorders. We used five datasets of various ages, ethnicities and family structure to demonstrate that data pooling approaches provide robust estimates of FA heritability in almost all white matter regions, suggesting that despite differences in acquisition and cohort structure, data pooling is feasible. We showed that both global and regional heritability estimates from pooled approaches were much more similar than heritability estimates derived from individual cohorts were to

each other. We determined stability in the pooling approaches by showing the differences in pooled estimates, using subsets of the data. While we generally showed that mega-analytic approaches have greater stability and highest power for jointly analyzing data, we note that meta-analytic approaches are comparable when full individual-level data sharing is not feasible, although caution and careful evaluation of cohort trends should be taken when making the choice of a meta-analytic approach. We provide our joint heritability results online at http://enigma.ini.usc.edu/ongoing/dti-working-group/2014_nimg/ to serve as a voxelwise heritability atlas for future studies.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2014.03.033.

References

Agartz, I., Okuguwa, G., Nordstrom, M., Greitz, D., Magnotta, V., Sedvall, G., 2001. Reliability and reproducibility of brain tissue volumetry from segmented MR scans. Eur. Arch. Psychiatry Clin. Neurosci. 251, 255–261.

Alba-Ferrara, L.M., de Erausquin, G.A., 2013. What does anisotropy measure? Insights from increased and decreased anisotropy in selective fiber tracts in schizophrenia. Front. Integr. Neurosci. 7, 9.

Almasy, L., Blangero, J., 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. Am. J. Hum. Genet. 62, 1198–1211.

Amos, C.I., 1994. Robust variance-components approach for assessing genetic linkage in pedigrees. Am. J. Hum. Genet. 54, 535–543.

Barysheva, M., Jahanshad, N., Foland-Ross, L., Altshuler, L.L., Thompson, P.M., 2012. White matter microstructural abnormalities in bipolar disorder: a whole brain diffusion tensor imaging study.

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

Basser, P.J., Mattiello, J., LeBihan, D., 1994. MR diffusion tensor spectroscopy and imaging. Biophys. J. 66, 259–267.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate — a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Methodol. 57, 289–300.

Bergen, S.E., Gardner, C.O., Kendler, K.S., 2007. Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: a meta-analysis. Twin Res. Hum. Genet. 10, 423–433.

Bis, J.C., Decarli, C., Smith, A.V., van der Lijn, F., Crivello, F., Fornage, M., Debette, S., Shulman, J.M., Schmidt, H., Srikanth, V., Schuur, M., Yu, L., Choi, S.H., Sigurdsson, S., Verhaaren, B.F., Destefano, A.L., Lambert, J.C., Jack Jr., C.R., Struchalin, M., Stankovich, J., Ibrahim-Verbaas, C.A., Fleischman, D., Zijdenbos, A., den Heijer, T., Mazoyer, B., Coker, L.H., Enzinger, C., Danoy, P., Amin, N., Arfanakis, K., van Buchem, M.A., de Bruijn, R.F., Beiser, A., Dufouil, C., Huang, J., Cavalieri, M., Thomson, R., Niessen, W.J., Chibnik, L.B., Gislason, G.K., Hofman, A., Pikula, A., Amouyel, P., Freeman, K.B., Phan, T.G., Oostra, B.A., Stein, J.L., Medland, S.E., Vasquez, A.A., Hibar, D.P., Wright, M.J., Franke, B., Martin, N.G., Thompson, P.M., Nalls, M.A., Uitterlinden, A.G., Au, R., Elbaz, A., Beare, R.J., van Swieten, J.C., Lopez, O.L., Harris, T.B., Chouraki, V., Breteler, M.M., De Jager, P.L., Becker, J.T., Vernooij, M.W., Knopman, D., Fazekas, F., Wolf, P.A., van der Lugt, A., Gudnason, V., Longstreth Jr., W.T., Brown, M.A., Bennett, D.A., van Duijn, C.M., Mosley, T.H., Schmidt, R., Tzourio, C., Launer, L.J., Ikram, M.A., Seshadri, S., 2012. Common variants at 12q14 and 12q24 are associated with hippocampal volume. Nat. Genet. 44, 545–551.
Blangero, J., Goring, H.H., Kent Jr., J.W., Williams, J.T., Peterson, C.P., Almasy, L., Dyer, T.D., 2005. Quantitative trait nucleotide analysis using Bayesian model selection. Hum. Biol. 77, 541–559.
Blangero, J., Diego, V.P., Dyer, T.D., Almeida, M., Peralta, J., Kent, J.W., Williams, J.T., Almasy, L., Goring, H.H.H., 2013. A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees. Adv. Genet. 81 (81), 1–31.
Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. Introduction to meta-analysis. John Wiley & Sons, Chichester, U.K.
Brannick, M.T., Yang, L.Q., Cafri, G., 2011. Comparison of weights for meta-analysis of r and d under realistic conditions. Organ. Res. Methods 14, 587–607.
Braskie, M.N., Ringman, J.M., Thompson, P.M., 2011. Neuroimaging measures as endophenotypes in Alzheimer's disease. Int. J. Alzheimers Dis. 2011, 490140.
Brouwer, R.M., Mandl, R.C., Peper, J.S., van Baal, G.C., Kahn, R.S., Boomsma, D.I., Hulshoff Pol, H.E., 2010. Heritability of DTI and MTR in nine-year-old children. Neuroimage 53, 1085–1092.
Brouwer, R.M., Mandl, R.C., Schnack, H.G., van Soelen, I.L., van Baal, G.C., Peper, J.S., Kahn, R.S., Boomsma, D.I., Hulshoff Pol, H.E., 2012. White matter development in early puberty: a longitudinal volumetric and diffusion tensor imaging twin study. PLoS One 7, e32316.
Carballedo, A., Amico, F., Ugwu, I., Fagan, A.J., Fahey, C., Morris, D., Meaney, J.F., Leemans, A., Frodl, T., 2012. Reduced fractional anisotropy in the uncinate fasciculus in patients with major depression carrying the met-allele of the Val66Met brain-derived neurotrophic factor genotype. Am. J. Med. Genet. B Neuropsychiatr. Genet. 159B, 537–548.
Chen, C.H., Gutierrez, E.D., Thompson, W., Panizzon, M.S., Jernigan, T.L., Eyler, L.T., Fennema-Notestine, C., Jak, A.J., Neale, M.C., Franz, C.E., Lyons, M.J., Grant, M.D., Fischl, B., Seidman, L.J., Tsuang, M.T., Kremen, W.S., Dale, A.M., 2012. Hierarchical genetic organization of human cortical surface area. Science 335, 1634–1636.
Clerx, L., Visser, P.J., Verhey, F., Aalten, P., 2012. New MRI markers for Alzheimer's disease: a meta-analysis of diffusion tensor imaging and a comparison with medial temporal lobe measurements. J. Alzheimers Dis. 29, 405–429.
de Zubicaray, G.I., Chiang, M.C., McMahon, K.L., Shattuck, D.W., Toga, A.W., Martin, N.G., Wright, M.J., Thompson, P.M., 2008. Meeting the challenges of neuroimaging genetics. Brain Imaging Behav. 2, 258–263.
den Braber, A., Ent, D., Blokland, G.A., van Grootheest, D.S., Cath, D.C., Veltman, D.J., de Ruiter, M., Boomsma, D.I., 2008. An fMRI study in monozygotic twins discordant for obsessive-compulsive symptoms. Biol. Psychol. 79, 91–102.
den Braber, A., van 't Ent, D., Boomsma, D.I., Cath, D.C., Veltman, D.J., Thompson, P.M., de Geus, E.J., 2011. White matter differences in monozygotic twins discordant or concordant for obsessive-compulsive symptoms: a combined diffusion tensor imaging/voxel-based morphometry study. Biol. Psychiatry 70, 969–977.
den Braber, A., van 't Ent, D., Cath, D.C., Veltman, D.J., Boomsma, D.I., de Geus, E.J., 2012. Brain activation during response interference in twins discordant or concordant for obsessive compulsive symptoms. Twin Res. Hum. Genet. 15, 372–383.
den Braber, A., de Geus, E.J., Boomsma, D.I., van 't Ent, D., 2013. Obsessive-compulsive symptoms and related sex differences in brain structure: an MRI study in Dutch twins. Twin Res. Hum. Genet. 16, 516–524.
Dinov, I., Lozev, K., Petrosyan, P., Liu, Z., Eggert, P., Pierce, J., Zamanyan, A., Chakrapani, S., Van Horn, J., Parker, D.S., Magsipoc, R., Leung, K., Gutman, B., Woods, R., Toga, A., 2010. Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. PLoS One 5.
Flint, J., Munafò, M.R., 2013. Candidate and non-candidate genes in behavior genetics. Curr. Opin. Neurobiol. 23 (1), 57–61 (Feb).
Friedman, J.I., Tang, C., Carpenter, D., Buchsbaum, M., Schmeidler, J., Flanagan, L., Golembo, S., Kanellopoulou, I., Ng, J., Hof, P.R., Harvey, P.D., Tsopelas, N.D., Stewart, D., Davis, K.L., 2008. Diffusion tensor imaging findings in first-episode and chronic schizophrenia patients. Am. J. Psychiatry 165, 1024–1032.
Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15, 870–878.
Glahn, D.C., Thompson, P.M., Blangero, J., 2007. Neuroimaging endophenotypes: strategies for finding genes influencing brain structure and function. Hum. Brain Mapp. 28, 488–501.
Glahn, D.C., Winkler, A.M., Kochunov, P., Almasy, L., Duggirala, R., Carless, M.A., Curran, J.C., Olvera, R.L., Laird, A.R., Smith, S.M., Beckmann, C.F., Fox, P.T., Blangero, J., 2010. Genetic control over the resting brain. Proc. Natl. Acad. Sci. U. S. A. 107, 1223–1228.
Hasler, G., Northoff, G., 2011. Discovering imaging endophenotypes for major depression. Mol. Psychiatry 16, 604–619.
Ikram, M.A., Fornage, M., Smith, A.V., Seshadri, S., Schmidt, R., Debette, S., Vrooman, H.A., Sigurdsson, S., Ropele, S., Taal, H.R., Mook-Kanamori, D.O., Coker, L.H., Longstreth Jr., W.T., Niessen, W.J., DeStefano, A.L., Beiser, A., Zijdenbos, A.P., Struchalin, M., Jack Jr., C.R., Rivadeneira, F., Uitterlinden, A.G., Knopman, D.S., Hartikainen, A.L., Pennell, C.E., Thiering, E., Steegers, E.A., Hakonarson, H., Heinrich, J., Palmer, L.J., Jarvelin, M.R., McCarthy, M.I., Grant, S.F., St Pourcain, B., Timpson, N.J., Smith, G.D., Sovio, U., Nalls, M.A., Au, R., Hofman, A., Gudnason, H., van der Lugt, A., Harris, T.B., Meeks, W.M.,
Vernooij, M.W., van Buchem, M.A., Catellier, D., Jaddoe, V.W., Gudnason, V., Windham, B.G., Wolf, P.A., van Duijn, C.M., Mosley Jr., T.H., Schmidt, H., Launer, L.J., Breteler, M.M., DeCarli, C., 2012. Common variants at 6q22 and 17q21 are associated with intracranial volume. Nat. Genet. 44, 539–544.
Jahanshad, N., Kochunov, P.V., Sprooten, E., Mandl, R.C., Nichols, T.E., Almasy, L., Blangero, J., Brouwer, R.M., Curran, J.E., de Zubicaray, G.I., Duggirala, R., Fox, P.T., Hong, L.E., Landman, B.A., Martin, N.G., McMahon, K.L., Medland, S.E., Mitchell, B.D., Olvera, R. L., Peterson, C.P., Starr, J.M., Sussmann, J.E., Toga, A.W., Wardlaw, J.M., Wright, M.J., Hulshoff Pol, H.E., Bastin, M.E., McIntosh, A.M., Deary, I.J., Thompson, P.M., Glahn, D. C., 2013. Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMA-DTI working group. Neuroimage 81, 455–469.
Julin, P., Melin, T., Andersen, C., Isberg, B., Svensson, L., Wahlund, L.O., 1997. Reliability of interactive three-dimensional brain volumetry using MP-RAGE magnetic resonance imaging. Psychiatry Res. 76 (1), 41–49 (Nov 28).
Kochunov, P., Duff Davis, M., 2009. Development of structural MR brain imaging protocols to study genetics and maturation. Methods 50, 136–146.
Kochunov, P., Glahn, D., Fox, P.T., Lancaster, J., Saleem, K., Shelledy, W., Zilles, K., Thompson, P., Coulon, O., Blangero, J., Fox, P., Rogers, J., 2009. Genetics of primary cerebral gyrification: heritability of length, depth and area of primary sulci in an extended pedigree of Papio baboons. Neuroimage 15, 1126–1132.
Kochunov, P., Glahn, D., Lancaster, J., Wincker, P., Smith, S., Thompson, P., Almasy, L., Duggirala, R., Fox, P., Blangero, J., 2010. Genetics of microstructure of cerebral white matter using diffusion tensor imaging. Neuroimage 15, 1109–1116.
Kochunov, P., Glahn, D.C., Nichols, T.E., Winkler, A.M., Hong, E.L., Holcomb, H.H., Stein, J.L., Thompson, P.M., Curran, J.E., Carless, M.A., Olvera, R.L., Johnson, M.P., Cole, S.A., Kochunov, V., Kent, J., Blangero, J., 2011. Genetic analysis of cortical thickness and fractional anisotropy of water diffusion in the brain. Front. Neurosci. 5, 120 (Oct 19).
Kochunov, P., Glahn, D.C., Rowland, L.M., Olvera, R., Wincker, P., Yang, D., Sampath, H., Carpenter, W., Duggirala, R., Curran, J., Blangero, J., Hong, L.E., 2012a. Testing the hypothesis of accelerated cerebral white matter aging in schizophrenia and major depression. Biol. Psychiatry. http://dx.doi.org/10.1016/j.biopsych.2012.10.002.
Kochunov, P., Rogers, W., Mangin, J.F., Lancaster, J., 2012b. A library of cortical morphology analysis tools to study development, aging and genetics of cerebral cortex. Neuroinformatics 10, 81–96.
Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation, 2nd ed. Springer, New York, London (xxvi, 589 pp.).
Lemieux, L., Hagemann, G., Krakow, K., Woermann, F.G., 1999. Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. Magn. Reson. Med. 42, 127–135.
Lerch, J.P., Evans, A.C., 2005. Cortical thickness analysis examined through power analysis and a population simulation. Neuroimage 24, 163–173.
Li, M.D., Cheng, R., Ma, J.Z., Swan, G.E., 2003. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. Addiction 98, 23–31.
Mandl, R.C., Rais, M., van Baal, G.C., van Haren, N.E., Cahn, W., Kahn, R.S., Hulshoff Pol, H.E., 2012. Altered white matter connectivity in never-medicated patients with schizophrenia. Hum. Brain Mapp. 34 (9), 2353–2365 (Sep).
Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A.V., Mahmood, A., Woods, R., Toga, A.W., Pike, G.B., Neto, P.R., Evans, A., Zhang, J., Huang, H., Miller, M. I., van Zijl, P., Mazziotta, J., 2008. Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. Neuroimage 40 (2), 570–582 (Apr 1).
Nazeri, A., Mallar Chakravarty, M., Felsky, D., Lobaugh, N.J., Rajji, T.K., Mulsant, B.H., Voineskos, A.N., 2013. Alterations of superficial white matter in schizophrenia and relationship to cognitive performance. Neuropsychopharmacology 38 (10), 1954–1962 (Sep).
Normand, S.L.T., 1999. Meta-analysis: formulating, evaluating, combining, and reporting. Stat. Med. 18 (3), 321–359. http://dx.doi.org/10.1002/(Sici)1097-0258(19990215) 18:3<321:Aid-Sim28>3.0.Co;2-P (PubMed PMID: WOS:000078439200008).
Olvera, R.L., Bearden, C.E., Velligan, D.I., Almasy, L., Carless, M.A., Curran, J.E., Williamson, D.E., Duggirala, R., Blangero, J., Glahn, D.C., 2011. Common genetic influences on depression, alcohol, and substance use disorders in Mexican-American families. Am. J. Med. Genet. B Neuropsychiatr. Genet. 156B, 561–568.
Penke, L., Munoz Maniega, S., Houlihan, L.M., Murray, C., Gow, A.J., Clayden, J.D., Bastin, M. E., Wardlaw, J.M., Deary, I.J., 2010a. White matter integrity in the splenium of the corpus callosum is related to successful cognitive aging and partly mediates the protective effect of an ancestral polymorphism in ADRB2. Behav. Genet. 40, 146–156.
Penke, L., Munoz Maniega, S., Murray, C., Gow, A.J., Hernandez, M.C., Clayden, J.D., Starr, J. M., Wardlaw, J.M., Bastin, M.E., Deary, I.J., 2010b. A general factor of brain white matter integrity predicts information processing speed in healthy older people. J. Neurosci. 30, 7569–7574.
Peper, J.S., Schnack, H.G., Brouwer, R.M., Van Baal, G.C., Pjetri, E., Székely, E., van Leeuwen, M., van den Berg, S.M., Collins, D.L., Evans, A.C., Boomsma, D.I., Kahn, R.S., Hulshoff Pol, H.E., 2009. Heritability of regional and global brain structure at the onset of puberty: a magnetic resonance imaging study in 9-year-old twin pairs. Hum. Brain Mapp. 30 (7), 2184–2196 (Jul).
Perez-Iglesias, R., Tordesillas-Gutierrez, D., McGuire, P.K., Barker, G.J., Roiz-Santianez, R., Mata, I., de Lucas, E.M., Rodriguez-Sanchez, J.M., Ayesa-Arriola, R., Vazquez-Barquero, J.L., Crespo-Facorro, B., 2011. White matter integrity and cognitive impairment in first-episode psychosis. Am. J. Psychiatry 167, 451–458.
Rhee, S.H., Waldman, I.D., 2002. Genetic and environmental influences on antisocial behavior: a meta-analysis of twin and adoption studies. Psychol. Bull. 128, 490–529.
Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31, 1487–1505.

Smith, S.M., Johansen-Berg, H., Jenkinson, M., Rueckert, D., Nichols, T.E., Miller, K.L., Robson, M.D., Jones, D.K., Klein, J.C., Bartsch, A.J., Behrens, T.E., 2007. Acquisition and voxelwise analysis of multi-subject diffusion data with tract-based spatial statistics. Nat. Protoc. 2, 499–503.

Sprooten, E., Sussmann, J.E., Clugston, A., Peel, A., McKirdy, J., Moorhead, T.W., Anderson, S., Shand, A.J., Giles, S., Bastin, M.E., Hall, J., Johnstone, E.C., Lawrie, S.M., McIntosh, A.M., 2011. White matter integrity in individuals at high genetic risk of bipolar disorder. Biol. Psychiatry 70, 350–356.

Stein, J.L., Medland, S.E., Vasquez, A.A., Hibar, D.P., Senstad, R.E., Winkler, A.M., Toro, R., Appel, K., Bartecek, R., Bergmann, O., Bernard, M., Brown, A.A., Cannon, D.M., Chakravarty, M.M., Christoforou, A., Domin, M., Grimm, O., Hollinshead, M., Holmes, A.J., Homuth, G., Hottenga, J.J., Langan, C., Lopez, L.M., Hansell, N.K., Hwang, K.S., Kim, S., Laje, G., Lee, P.H., Liu, X., Loth, E., Lourdusamy, A., Mattingsdal, M., Mohnke, S., Maniega, S.M., Nho, K., Nugent, A.C., O'Brien, C., Papmeyer, M., Putz, B., Ramasamy, A., Rasmussen, J., Rijpkema, M., Risacher, S.L., Roddey, J.C., Rose, E.J., Ryten, M., Shen, L., Sprooten, E., Strengman, E., Teumer, A., Trabzuni, D., Turner, J., van Eijk, K., van Erp, T.G., van Tol, M.J., Wittfeld, K., Wolf, C., Woudstra, S., Aleman, A., Alhusaini, S., Almasy, L., Binder, E.B., Brohawn, D.G., Cantor, R.M., Carless, M.A., Corvin, A., Czisch, M., Curran, J.E., Davies, G., de Almeida, M.A., Delanty, N., Depondt, C., Duggirala, R., Dyer, T.D., Erk, S., Fagerness, J., Fox, P.T., Freimer, N.B., Gill, M., Goring, H.H., Hagler, D.J., Hoehn, D., Holsboer, F., Hoogman, M., Hosten, N., Jahanshad, N., Johnson, M.P., Kasperaviciute, D., Kent Jr., J.W., Kochunov, P., Lancaster, J.L., Lawrie, S.M., Liewald, D.C., Mandl, R., Matarin, M., Mattheisen, M., Meisenzahl, E., Melle, I., Moses, E.K., Muhleisen, T.W., Nauck, M., Nothen, M.M., Olvera, R.L., Pandolfo, M., Pike, G.B., Puls, R., Reinvang, I., Renteria, M.E., Rietschel, M., Roffman, J.L., Royle, N.A., Rujescu, D., Savitz, J., Schnack, H.G., Schnell, K., Seiferth, N., Smith, C., Steen, V.M., Valdes Hernandez, M.C., Van den Heuvel, M., van der Wee, N.J., Van Haren, N.E., Veltman, J.A., Volzke, H., Walker, R., Westlye, L.T., Whelan, C.D., Agartz, I., Boomsma, D.I., Cavalleri, G.L., Dale, A.M., Djurovic, S., Drevets, W.C., Hagoort, P., Hall, J., Heinz, A., Jack Jr., C.R., Foroud, T.M., Le Hellard, S., Macciardi, F., Montgomery, G.W., Poline, J.B., Porteous, D.J., Sisodiya, S.M., Starr, J.M., Sussmann, J., Toga, A.W., Veltman, D.J., Walter, H., Weiner, M.W., Bis, J.C., Ikram, M.A., Smith, A.V., Gudnason, V., Tzourio, C., Vernooij, M.W., Launer, L.J., DeCarli, C., Seshadri, S., Andreassen, O.A., Apostolova, L.G., Bastin, M.E., Blangero, J., Brunner, H.G., Buckner, R.L., Cichon, S., Coppola, G., de Zubicaray, G.I., Deary, I.J., Donohoe, G., de Geus, E.J., Espeseth, T., Fernandez, G., Glahn, D.C., Grabe, H.J., Hardy, J., Hulshoff Pol, H.E., Jenkinson, M., Kahn, R.S., McDonald, C., McIntosh, A.M., McMahon, F.J., McMahon, K.L., Meyer-Lindenberg, A., Morris, D.W., Muller-Myhsok, B., Nichols, T.E., Ophoff, R.A., Paus, T., Pausova, Z., Penninx, B.W., Potkin, S.G., Samann, P.G., Saykin, A.J., Schumann, G., Smoller, J.W., Wardlaw, J.M., Weale, M.E., Martin, N.G., Franke, B., Wright, M.J., Thompson, P.M., 2012. Identification of common variants associated with human hippocampal and intracranial volumes. Nat. Genet. 44, 552–561.

Sutton, A.J., 2000. Methods for Meta-analysis in Medical Research. J. Wiley, Chichester, New York (xvii, 317 pp.).

Taal, H.R., St Pourcain, B., Thiering, E., Das, S., Mook-Kanamori, D.O., Warrington, N.M., Kaakinen, M., Kreiner-Moller, E., Bradfield, J.P., Freathy, R.M., Geller, F., Guxens, M., Cousminer, D.L., Kerkhof, M., Timpson, N.J., Ikram, M.A., Beilin, L.J., Bonnelykke, K., Buxton, J.L., Charoen, P., Chawes, B.L., Eriksson, J., Evans, D.M., Hofman, A., Kemp, J.P., Kim, C.E., Klopp, N., Lahti, J., Lye, S.J., McMahon, G., Mentch, F.D., Muller-Nurasyid, M., O'Reilly, P.F., Prokopenko, I., Rivadeneira, F., Steegers, E.A., Sunyer, J., Tiesler, C., Yaghootkar, H., Breteler, M.M., Decarli, C., Debette, S., Fornage, M., Gudnason, V., Launer, L.J., van der Lugt, A., Mosley Jr., T.H., Seshadri, S., Smith, A.V., Vernooij, M.W., Blakemore, A.I., Chiavacci, R.M., Feenstra, B., Fernandez-Banet, J., Grant, S.F., Hartikainen, A.L., van der Heijden, A.J., Iniguez, C., Lathrop, M., McArdle, W.L., Molgaard, A., Newnham, J.P., Palmer, L.J., Palotie, A., Pouta, A., Ring, S.M., Sovio, U., Standl, M., Uitterlinden, A.G., Wichmann, H.E., Vissing, N.H., van Duijn, C.M., McCarthy, M.I., Koppelman, G.H., Estivill, X., Hattersley, A.T., Melbye, M., Bisgaard, H., Pennell, C.E., Widen, E., Hakonarson, H., Smith, G.D., Heinrich, J., Jarvelin, M.R., Jaddoe, V.W., 2012. Common variants at 12q15 and 12q24 are associated with infant head circumference. Nat. Genet. 44, 532–538.

Teipel, S.J., Reuter, S., Stieltjes, B., Acosta-Cabronero, J., Ernemann, U., Fellgiebel, A., Filippi, M., Frisoni, G., Hentschel, F., Jessen, F., Kloppel, S., Meindl, T., Pouwels, P.J., Hauenstein, K.H., Hampel, H., 2011. Multicenter stability of diffusion tensor imaging measures: a European clinical and physical phantom study. Psychiatry Res. 194, 363–371.

Teipel, S.J., Wegrzyn, M., Meindl, T., Frisoni, G., Bokde, A.L., Fellgiebel, A., Filippi, M., Hampel, H., Kloppel, S., Hauenstein, K., Ewers, M., 2012. Anatomical MRI and DTI in the diagnosis of Alzheimer's disease: a European multicenter study. J. Alzheimers Dis. 31 (Suppl 3), S33–S47.

Thomason, M.E., Thompson, P.M., 2011. Diffusion imaging, white matter, and psychopathology. Annu. Rev. Clin. Psychol. 7, 63–85.

Thompson, P.M., Cannon, T.D., Narr, K.L., van Erp, T., Poutanen, V.P., Huttunen, M., Lonnqvist, J., Standertskjold-Nordenstam, C.G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C.I., Toga, A.W., 2001. Genetic influences on brain structure. Nat. Neurosci. 4, 1253–1258.

van Beijsterveldt, C.E., Groen-Blokhuis, M., Hottenga, J.J., Franic, S., Hudziak, J.J., Lamb, D., Huppertz, C., de Zeeuw, E., Nivard, M., Schutte, N., Swagerman, S., Glasner, T., van Fulpen, M., Brouwer, C., Stroet, T., Nowotny, D., Ehli, E.A., Davies, G.E., Scheet, P., Orlebeke, J.F., Kan, K.J., Smit, D., Dolan, C.V., Middeldorp, C.M., de Geus, E.J., Bartels, M., Boomsma, D.I., 2013. The Young Netherlands Twin Register (YNTR): longitudinal twin and family studies in over 70,000 children. Twin Res. Hum. Genet. 16, 252–267.

van den Heuvel, M.P., van Soelen, I.L., Stam, C.J., Kahn, R.S., Boomsma, D.I., Hulshoff Pol, H.E., 2013. Genetic control of functional brain network efficiency in children. Eur. Neuropsychopharmacol. 23 (1), 19–23 (Jan).

Van Horn, J.D., Grafton, S.T., Rockmore, D., Gazzaniga, M.S., 2004. Sharing neuroimaging studies of human cognition. Nat. Neurosci. 7, 473–481.

van Soelen, I.L., Brouwer, R.M., Peper, J.S., van Leeuwen, M., Koenis, M.M., van Beijsterveldt, T.C., Swagerman, S.C., Kahn, R.S., Hulshoff Pol, H.E., Boomsma, D.I., 2012. Brain SCALE: brain structure and cognition: an adolescent longitudinal twin study into the genetic etiology of individual differences. Twin Res. Hum. Genet. 15, 453–467.

Verweij, K.J., Zietsch, B.P., Lynskey, M.T., Medland, S.E., Neale, M.C., Martin, N.G., Boomsma, D.I., Vink, J.M., 2010. Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. Addiction 105, 417–430.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., Morris, J.C., Petersen, R.C., Saykin, A.J., Schmidt, M.E., Shaw, L., Siuciak, J.A., Soares, H., Toga, A.W., Trojanowski, J.Q., 2012. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimers Dement. 8, S1–S68.

Willer, C.J., Li, Y., Abecasis, G.R., 2010. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26 (17), 2190–2191. http://dx.doi.org/10.1093/bioinformatics/btq340 (Epub 2010/07/10; PubMed PMID: 20616382; PubMed Central PMCID: PMC2922887.

Zhan, L., Jahanshad, N., Ennis, D.B., Jin, Y., Bernstein, M.A., Borowski, B.J., Jack Jr., C.R., Toga, A.W., Leow, A.D., Thompson, P.M., 2013. Angular versus spatial resolution trade-offs for diffusion imaging under time constraints. Hum. Brain Mapp. 34 (10), 2688–2706 (Oct).

Zuk, O., Hechter, E., Sunyaev, S.R., Lander, E.S., 2012. The mystery of missing heritability: genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. U. S. A. 109, 1193–1198.