Behavioral/Systems/Cognitive

Heritability of Working Memory Brain Activation

Gabriëlla A. M. Blokland,[1,2,3] **Katie L. McMahon,**[2] **Paul M. Thompson,**[4] **Nicholas G. Martin,**[1] **Greig I. de Zubicaray,**[3] **and Margaret J. Wright**[1,3]

[1]Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland 4006, Australia, [2]Centre for Advanced Imaging, and [3]School of Psychology, University of Queensland, Brisbane, Queensland 4072, Australia, and [4]Laboratory of Neuro Imaging, Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, California 90095

Although key to understanding individual variation in task-related brain activation, the genetic contribution to these individual differences remains largely unknown. Here we report voxel-by-voxel genetic model fitting in a large sample of 319 healthy, young adult, human identical and fraternal twins (mean ± SD age, 23.6 ± 1.8 years) who performed an *n*-back working memory task during functional magnetic resonance imaging (fMRI) at a high magnetic field (4 tesla). Patterns of task-related brain response (BOLD signal difference of 2-back minus 0-back) were significantly heritable, with the highest estimates (40–65%) in the inferior, middle, and superior frontal gyri, left supplementary motor area, precentral and postcentral gyri, middle cingulate cortex, superior medial gyrus, angular gyrus, superior parietal lobule, including precuneus, and superior occipital gyri. Furthermore, high test-retest reliability for a subsample of 40 twins indicates that nongenetic variance in the fMRI brain response is largely due to unique environmental influences rather than measurement error. Individual variations in activation of the working memory network are therefore significantly influenced by genetic factors. By establishing the heritability of cognitive brain function in a large sample that affords good statistical power, and using voxel-by-voxel analyses, this study provides the necessary evidence for task-related brain activation to be considered as an endophenotype for psychiatric or neurological disorders, and represents a substantial new contribution to the field of neuroimaging genetics. These genetic brain maps should facilitate discovery of gene variants influencing cognitive brain function through genome-wide association studies, potentially opening up new avenues in the treatment of brain disorders.

Introduction

Functional magnetic resonance imaging (fMRI) is a powerful tool for interrogating the mechanisms of the brain's response to different environmental stimuli. However, even with a rigidly standardized stimulus or task, we know that the brain's response is highly variable between people. It is of considerable interest to know how much of this variability is due to genetic differences between people, how much to their (unique) environmental experience, and how much to measurement error. To tease apart these contributions to individual variability, we measured brain response to an *n*-back working memory (WM) task. Activation during this task is affected in persons with genetically influenced

neurodegenerative (Wishart et al., 2004) and neuropsychiatric disorders (Callicott et al., 1998; Matsuo et al., 2007), and in healthy family members at increased genetic risk for some of these disorders (Callicott et al., 2003; Drapier et al., 2008). In addition, prior studies have shown that working memory task performance measures are heritable phenotypes (Ando et al., 2001; Luciano et al., 2001; Polderman et al., 2006). Furthermore, the *n*-back task robustly activates areas (Owen et al., 2005) shown to be highly heritable in twin studies of brain structure (Schmitt et al., 2007). Although substantial heritability for certain features of brain structure has now been firmly established by twin studies, only a handful of neuroimaging studies has examined the heritability of task-related neural activity, as captured by fMRI. These studies, mostly involving a region-of-interest (ROI) approach, have had mixed findings, ranging from no genetic effect to strong genetic influences (Côté et al., 2007; Matthews et al., 2007; Polk et al., 2007; Blokland et al., 2008; Koten et al., 2009). This may be attributable to differences in task paradigms and analysis approaches and/or lack of statistical power due to small sample sizes.

Here we investigated the relative contributions of genetic (heritability) and environmental influences to individual variation in task-related brain activation across the brain using fMRI in a large genetically informative sample of identical [monozygotic (MZ)] and fraternal [dizygotic (DZ)] twins. This study is a substantive extension on our first analysis (Blokland et al., 2008), both in terms of the complexity of the voxel-by-voxel analyses and the sample size. A total of 319 healthy, young adult MZ and

Table 1. Demographic characteristics of the twin sample[a]

(mean ± SD)	Females (n = 199 individuals)	Males (n = 120 individuals)	Total (n = 319 individuals)
Age (years)	23.5 ± 1.8	23.7 ± 1.9	23.6 ± 1.8
FIQ[b]	113.4 ± 10.9	117.5 ± 12.9**	114.9 ± 11.9
Gestational age (weeks)[c]	37.9 ± 2.8	38.1 ± 2.7	38.0 ± 2.7
Birth weight (g)	2575.4 ± 477.4	2715.9 ± 525.4*	2628.5 ± 500.0
Socioeconomic index	52.6 ± 24.2	54.7 ± 24.6	53.4 ± 24.3

[a]The sample included 75 MZ pairs, 66 DZ pairs, and 37 unpaired twins. Although unpaired MZ and DZ twins did not contribute to the estimation of the genetic and environmental parameters, they did contribute to the estimation of mean and variance effects (i.e. they allowed a more accurate estimation of phenotypic correlations and phenotypic effects).

[b]FIQ was measured using five subtests from the Multidimensional Aptitude Battery (Jackson, 1984), as close as possible to the twins' 16th birthday. The observed higher mean is likely due to the fact that the MAB test was created and normalized for Canadian samples, and therefore results on this test may differ when used in a different country. In addition, the presence of an ascertainment bias cannot be excluded, as more intelligent and often more highly educated people tend to volunteer for these studies more frequently. However, the higher FIQ mean does not affect the representativeness of this sample because FIQ follows a normal distribution, with scores ranging from 85 to 146, thus showing good variability.

[c]Gestational age, birth weight, and parental socioeconomic status (McMillan et al., 2009) were obtained from parental reports, either when the twins were 12 or 16 years. There were no significant mean or variance differences between co-twins or by zygosity. Males had slightly higher FIQ ($d = 0.34$, $p < 0.01$) and birth weight ($d = 0.28$, $p < 0.05$) than females. *$p < 0.05$; **$p < 0.01$.

DZ twins performed the 0- and 2-back versions of an established spatial, numerical *n*-back task (Callicott et al., 1998, 2003; Blokland et al., 2008). This task requires online monitoring, updating, and manipulation of remembered information, and is therefore assumed to place great demands on a number of key processes essential to working memory (Owen et al., 2005).

Materials and Methods

Participants. Twins between 20 and 30 years of age who had participated in the Brisbane Twin Cognition study at 16 years of age (Wright and Martin, 2004) were invited to participate in the present study. Before inclusion, twins were assessed for handedness; and were screened (by self-report) for their suitability for imaging; for significant medical, psychiatric, or neurological conditions, including head injuries; for a current or past diagnosis of substance abuse; and for current use of cognition-affecting medication. Of the 366 participants thus far, 7 were excluded due to excessive head motion or incomplete datasets, and 40 because of inadequate task performance. The final sample of 319 twins included 75 MZ pairs (46 female, 29 male), 66 DZ pairs (30 female, 11 male, 25 opposite sex), and 37 unpaired twins (12 MZ, 25 DZ; 22 female, 15 male), aged 23.6 ± 1.8 years (age range, 20–28 years), and all were right handed. Table 1 describes the sample demographics. Seventy-one percent of twin pairs were scanned on the same day, with the remainder, on average, within 12 d of each other. The scanning session lasted 60 min, and each participant received a $100 gift voucher in appreciation of their time. The study was approved by the Human Research Ethics Committees of the Queensland Institute of Medical Research, University of Queensland, and Uniting Health Care. Written informed consent was obtained from each participant.

Genotyping. Details regarding initial determination of zygosity have been described previously (Wright and Martin, 2004; Blokland et al., 2008). Zygosity typing was later confirmed by genome-wide single nucleotide polymorphism genotyping (Illumina 610K chip).

Psychometric intelligence testing, birth information, and socioeconomic status. General intellectual ability was assessed at age 16 years as part of the Brisbane Twin Cognition study (Wright and Martin, 2004) using three verbal and two performance subtests from the Multidimensional Aptitude Battery (Jackson, 1984), a measure that shows good temporal stability (Jackson, 1984; Harrell et al., 1987; Luciano et al., 2003) and is highly correlated with the Wechsler Adult Intelligence Scale (Luciano et al., 2003). A full-scale intelligence quotient (FIQ) standardized score was derived from these subtests (Wright and Martin, 2004). FIQ for one DZ female was unavailable due to computer hardware failure. The mean interval between cognitive testing and MRI scanning was 7 years and 9 months (range, 3.8–11.6 years). Birth information (gestational age and

birth weight) was obtained from a parent, usually the mother, as part of the cognition study. Socioeconomic status was obtained from parental report and rated according to the Australian Socioeconomic Index 2006 (McMillan et al., 2009).

Experimental procedure. Imaging was conducted on a 4 tesla Bruker Medspec whole-body scanner (Bruker) in Brisbane, Australia. Functional images were acquired using a T2*-weighted gradient echo-planar imaging (EPI) sequence, sensitive to blood oxygen level-dependent (BOLD) contrast [interleaved; repetition time (TR) = 2100 ms; echo time (TE) = 30 ms; flip angle = 90°; field of view (FOV) = 230 × 230 mm], and using a radio-frequency receive–transmit transverse electromagnetic head coil (M&R Devices) (Vaughan, 1999). Geometric distortions in the EPI images caused by magnetic field inhomogeneities at high field were corrected using a point-spread mapping approach (Zeng and Constable, 2002; Zaitsev et al., 2003). Over a continuous imaging run, we acquired 127 axial brain volumes, one volume every 2.1 s, with 36 coronal slices of 3 mm thickness (64 × 64 matrix; voxel size, 3.6 × 3.6 × 3.0 mm), and with a 0.6 mm slice gap. In addition to the functional scans, 3D T1-weighted images were acquired (MPRAGE; TR = 1500 ms; TE = 3.35 ms; TI = 700 ms; pulse angle = 8°; coronal orientation; FOV = 230 mm³; 256 × 256 × 256 matrix; slice thickness = 0.9 mm).

During functional imaging, the participants performed the 0-back and 2-back versions of the spatial *n*-back working memory task based on Callicott et al. (1998, 2003). See Figure 1 in Blokland et al. (2008). In this task, a number (1–4, randomized) was presented in a fixed position in one of four white circles, positioned at the corners of a diamond-shaped square. Stimuli were projected using a digital light video projector and presented on a screen at the foot of the scanner bed, viewed through a mirror mounted on the head coil. A fiber-optic response box, with four buttons arranged in the same configuration as the numbers presented on the screen, was used for responses. Participants pressed one of the four buttons to match the target stimulus. For $n = 0$ (i.e., 0-back), a simple button press in response to the number displayed was required. For $n = 2$ (i.e., 2-back), participants pressed the button corresponding to the number presented two trials before the current one. Thus, the 2-back condition required both the maintenance of the last two numbers in memory and the updating of these encoded stimuli as each new stimulus was presented (Fletcher and Henson, 2001). While difficulty increased from 0-back to 2-back, the stimulus information and demands on response selection and execution were the same within levels. Task conditions were run in blocks with the level of the task shown on the screen, and the background color of the diamond-shaped square changing from blue (0-back) to yellow (2-back). Participants were scanned through 16 alternating blocks of the 0-back and 2-back conditions (i.e., 8 blocks/condition). Each block consisted of 16 trials, with a stimulus presentation time of 200 ms and an interstimulus interval of 800 ms, resulting in a total experimental length of 4 min and 16 s (256 s). Participants were fully trained on the task before being positioned in the scanner, with each performing a minimum of four training blocks (two per condition). The importance of effort and commitment to the task was emphasized. To familiarize participants with the scanner and the response box, an additional set of practice trials were given once they were positioned in the magnet, with the pulse sequence running in the background. Task performance was measured as the percentage of correct responses (accuracy) and average response time (across correct trials) for each of the task conditions separately. The mean percentages correct in the 0- and 2-back conditions were 88.3 and 71.5, respectively (Table 1). On the 0-back condition, 98.1% of the sample had >50% accuracy, and 99.7% had >40% accuracy. On the 2-back condition, 84.6% of the sample had >50% accuracy, and 93.4% had >40% accuracy.

Image preprocessing. Images were processed and analyzed using Statistical Parametric Mapping software (SPM5, Wellcome Department of Cognitive Neurology) implemented in MATLAB (MathWorks). The first five EPI volumes were discarded to ensure that steady-state tissue magnetization was reached. Time-series volumes were realigned and unwarped using a robust rigid-body transformation procedure (Freire et al., 2002). A mean image generated during realignment was then coregistered with the participant's 3D T1 image, and the latter spatially normalized via nonlinear basis functions to the standard T1 template image in

MNI atlas space using the unified segmentation approach (Ashburner and Friston, 1999). The nonlinear transformations were next applied to the time-series volumes from which the mean was generated. Normalized volumes were then resampled to 3 mm^3 voxels and smoothed with an $8 \times 8 \times 8$ mm full width at half-maximum isotropic Gaussian kernel. Global signal effects were estimated and removed using a voxel-level general linear model (Macey et al., 2004). High-pass (cutoff: 128 s) and low-pass (AR1 model) filtering were applied to discard signals of no interest.

Image analysis. Image analysis was conducted in two stages. First, block design fixed-effects models were fitted at the single-subject level. Separate regressors were constructed for the 0- and 2-back conditions comprising a boxcar reference waveform convolved with a canonical hemodynamic response function. Second, the resulting single-subject 2-back>0-back *t*-contrast images were entered into a second-level group random-effects model (one-sample *t* test), regardless of zygosity. Using the Volumes Toolbox (http://sourceforge.net/projects/spmtools; authored by V. Glauche, Universität Freiburg, Freiburg, Germany), *t* scores were extracted from single-subject 2>0-back contrast images in each of 15,804 voxels comprising a brain mask created from the random-effects analysis ($p < 0.05$, family wise error (FWE) corrected; extent threshold, 25 voxels). Voxel activation values from the 2>0-back *t*-contrast images approximated normal distributions and did not require transformation. For each case, in each voxel, the square of the Mahalanobis distance was calculated to screen for and remove outliers. In addition, total gray matter volume was calculated by segmenting the 3D T1-weighted images in SPM5.

Genetic modeling. Using the statistical package Mx (Neale et al., 2002; Neale and Maes, 2004), maximum likelihood twin correlations were estimated, and univariate structural equation models examined the means and genetic (A) and environmental (E) sources of variance for task-related brain activation extracted from single-subject 2>0-back *t*-contrast images in each of the voxels specified by the group analysis. Age, sex, 2-back performance accuracy, and FIQ (estimated at age 16 years) were included as covariates, leaving their effects free to vary in each voxel. Correlations between genetic factors are fixed at 1 for MZ twin pairs, as they share 100% of their genes, and 0.5 for DZ pairs as they share, on average, 50% of their genes (see supplemental Fig. 1*c*, available at www.jneurosci.org as supplemental material). Thus, if the patterns and intensities of neural activity associated with the recruitment of working memory are significantly more similar in MZ twins than in DZ twins, this is strong evidence of heritability. By definition, nonshared environmental factors (e.g., illness, prenatal or postnatal traumas, peer groups), which also includes measurement error, are left uncorrelated. The significance of A was determined by testing whether dropping this parameter resulted in a significant ($p < 0.05$) decrease in the goodness-of-fit χ^2 statistic. Probability maps were height thresholded per voxel at an uncorrected $p < 0.05$, then cluster thresholded to correct for multiple statistical tests. At 10,000 iterations, a voxel-level threshold of $p < 0.05$, a mask image of 15,804 voxels, and an 8 mm smoothing kernel, Monte Carlo simulation (AlphaSim; http://afni.nimh.nih.gov/pub/dist/doc/program_help/AlphaSim.html) determined a minimum cluster size of 147 voxels for cluster significance at $p < 0.05$. Cluster thresholding of p value maps was performed using FSL and AFNI command line tools (http://www.fmrib.ox.ac.uk/fsl/ and http://afni.nimh.nih.gov, respectively).

We assessed whether the AE model was appropriate in several steps. We compared the maximum likelihood twin correlation maps for MZ and DZ pairs. From the MZ and DZ twin correlations (r_{MZ} and r_{DZ}, respectively), we calculated initial estimates of heritability [$h^2 = 2^*(r_{MZ} - r_{DZ})$] and common environmental influence [$c^2 = (2^*r_{DZ}) - r_{MZ}$] (Falconer and Mackay, 1996), and plotted the observed sampling distributions of h^2 and c^2 against the expected sampling distributions under the null hypothesis of no effect. Figure 1 shows that the differences between the observed and expected sampling distributions are consistent with the presence of genetic influences (heritability) and the absence of common environmental influences (C). The midline of the observed h^2 sampling distribution is well above zero, whereas the midline of the observed c^2 distribution is slightly below zero. The low DZ correlations, being less than half the MZ correlation and the fact that the midline of the
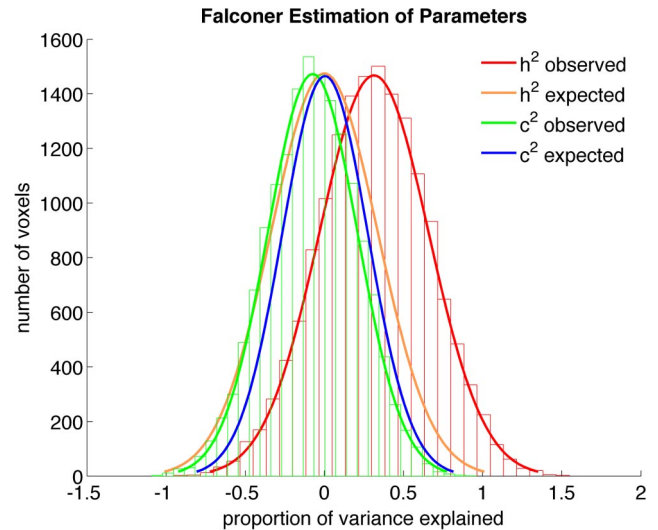

Figure 1. Observed and expected sampling distributions for genetic and environmental parameters. h^2 observed $= [2^*(r_{MZ} - r_{DZ})]$ and c^2 observed $= [(2^*r_{DZ}) - r_{MZ}]$ (Falconer and Mackay, 1996). h^2 and c^2 expected are normal distributions with a mean of zero and an expected sampling variance estimated as $[4^*((1 - r_{MZ})^2)^2/m + (1 - r_{DZ})^2)^2/n)]$ for h^2 expected, and $[(4^*(1 - r_{DZ})^2)^2/n + ((1 - r_{MZ})^2)^2/m)]$ for c^2 expected, where n and m refer to the numbers of DZ and MZ twin pairs, respectively, and r_{MZ} and r_{DZ} are set to zero under the null hypothesis of no heritability and no common environmental influence (Visscher, 2004).

observed c^2 distribution is slightly below zero suggested genetic dominance (D) in some regions. However, we did not have enough statistical power to estimate this separately in a structural equation model that includes additive genetic factors (A), dominance genetic factors (D), and unique environmental factors (E) (see supplemental Fig. 1*b*, available at www.jneurosci.org as supplemental material). In <2% of group-activated voxels, the variance component D was significant at $p < 0.05$. When applying cluster thresholding to the D map, no clusters survived. Based on this result, it was decided to fit an AE model to all group-activated voxels, with A including both additive and dominant genetic effects.

Reproducibility. Twenty twin pairs (5 MZF, 5 MZM, 5 DZF, 5 DZM) were rescanned 117 ± 56 d (range, 74–291 d) after their initial scan. Reproducibility of task-related brain activation across sessions was assessed by calculating voxelwise intraclass correlations (Shrout and Fleiss, 1979), and by carrying out a paired-samples *t* test ($p < 0.05$, FWE corrected) between contrast images from the two time points.

Results

Genetic modeling
We first established a brain mask of commonly activated regions (2-back>0-back *t* contrast) identified in a group-level random-effects analysis regardless of zygosity, and restricted our analyses to these regions (Fig. 2*a*). The group random-effects analysis showed the most significant increase in BOLD signal during the 2-back compared with the 0-back condition ($p < 0.05$, FWE corrected) (with the MNI coordinates for left [L] and right [R] hemispheres provided in brackets [*x, y, z*]), as follows: middle frontal gyri (including dorsolateral prefrontal cortex) (L: [−27, 6, 54]; right (R): [27, 12, 51], [42, 39, 21]; lateral orbitofrontal gyrus (R: [33, 24, −6]; superior frontal gyrus (R: [6, 24, 42]); inferior frontal gyri (L: [−45, 6, 33], [−45, 9, 24]; R: [45, 12, 27], [51, 15, 12], [51, 18, 3]); supramarginal gyri (L: [−42, −42, 42]; R: [33, −57, 45], [45, −39, 45]); superior parietal gyri (L: [−24, −63, 42]; R: [30, −63, 48]); insular cortex (L: [−33, 21, −3]); precuneus (L: [−9, −60, 54]; R: [9, −57, 54]); angular gyrus (R: [36, −54, 48], [57, −39, 18]); cerebellum (L: [−36, −54, −30], [−36, −60, −27], [−30, −57, −30], [−24, −63, −27], [0, −51, −18];
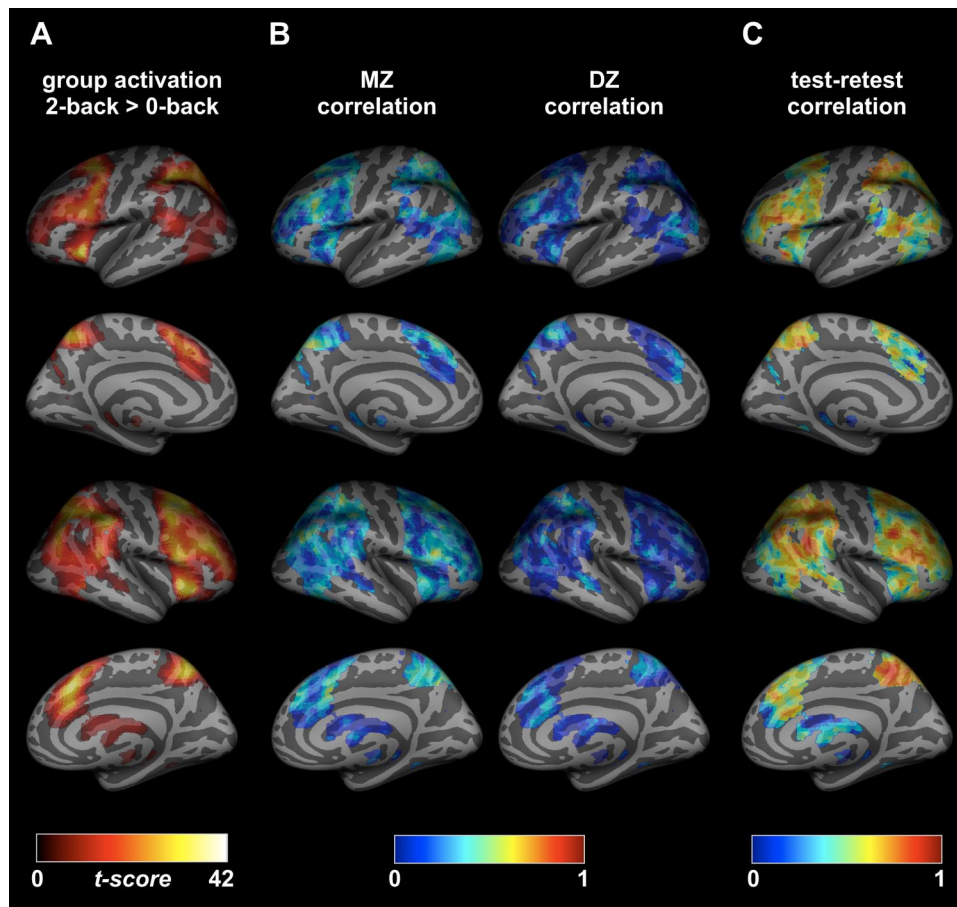
Figure 2. Group activation, twin correlations, and test-retest reliability. *A–C*, Group random-effects analysis for the 2-back >0-back *t* contrast (*p* < 0.05, FWE corrected) (*A*), maximum likelihood MZ and DZ twin correlations (*B*), and test-retest correlations within the group activation mask (*C*). Confidence intervals for twin and test-retest correlations are available in the supplemental material (available at www.jneurosci.org as supplemental material). Statistical maps are rendered on the Freesurfer inflated brain (CorTechs Labs) (Fischl et al., 1999) using the SPM SurfRend Toolbox (http://spmsurfrend.sourceforge.net; authored by Itamar Kahn, Universität Freiburg, Freiburg, Germany) and NeuroLens (Neurovascular Imaging Laboratory, L'Unité de Neuro-imagerie Fonctionnelle, Montréal, QC, Canada), separately for lateral and medial views in the left and right hemispheres.

R: [33, −60, −30]); middle occipital gyri (L: [−27, −69, 30]; R: [36, −69, 36]); caudate nuclei (L: [−18, 3, 18]; R: [18, 6, 15]); middle temporal gyrus (R: [54, −51, −6]); and inferior temporal gyrus (R: [51, −54, −9]). Activation peaks were labeled using an image-based probabilistic brain atlas based on cytoarchitectonic cortical parcellation (Eickhoff et al., 2007). The activation pattern is consistent with the literature, including our previous study in a subsample of the current sample (Callicott et al., 1998, 2003; Blokland et al., 2008).

Maximum likelihood twin correlations for task-related brain activation are shown in Figure 2*b*. For the regions activated by this task, overall MZ twin correlations were more than twice the size of the DZ correlations, suggesting that individual variation in WM activation is genetically influenced. DZ twin correlations in many voxels did not differ significantly from zero (for confidence intervals, see supplemental Fig. 2, available at www.jneurosci.org as supplemental material).

Based on our preliminary investigations (see Materials and Methods), we opted to fit a model that includes genetic (A), and unique environmental factors (E) to explain the variance in task-related brain activation (for confidence intervals, see supplemental Fig. 1c, available at www.jneurosci.org as supplemental material). The resulting genetic component *p* value map was height thresholded at an uncorrected *p* < 0.05, and cluster thresholded based on Monte Carlo simulation. The A and E esti-

mates and the A clusters surviving this strict significance criterion are shown in Figure 3. At a cluster threshold of 147 voxels, three large clusters were significant, totaling 8224 voxels, with moderately high heritability estimates (~40–65%) in inferior frontal gyri (pars triangularis) (L: [−48, 24, 21]; R: [45, 30, 12], [48, 24, 24]); inferior frontal gyrus (pars opercularis) (R: [57, 15, 18], [60, 12, 18]); middle frontal gyri (L: [−39, 30, 33]; R: [30, 36, 42], [33, 39, 21], [36, 24, 33], [36, 54, 15]); superior frontal gyrus (R: [21, 21, 45], [24, 27, 33]); supplementary motor area (SMA; Brodmann area 6) (L: [−3, 21, 51]); precentral gyrus (L: [−42, 0, 51], [−39, −3, 54]); postcentral gyrus (Brodmann area 2) (R: [33, −42, 54]); middle cingulate cortex (R: [9, 15, 45]); superior medial gyrus (L: [3, 36, 33]); angular gyrus (R: [42, −72, 30], [45, −66, 33]); primary somatosensory cortex (Brodmann area 3a) (R: [9, − 45, 60]); superior parietal lobule, including precuneus (L: [−9, −45, 57], [−6, −51, 48], [0, −63, 45]; R: [3, −54, 48], [12, −48, 66], [12, −72, 48], [18, −69, 51], [21, −42, 48]), and superior occipital gyri (L: [−21, −66, 33]; R: [21, −69, 45]).

Somewhat lower estimates (~20–39%) were found in inferior frontal gyri (pars triangularis) (L: [−36, 27, 0], [−30, 30, 3]); inferior frontal gyri (pars orbitalis) (L: [−27, 30, −9]; R: [30, 27, −15], [39, 21, −18]); middle frontal gyrus (R: [36, 54, 3]); precentral gyrus (L: [−42, 0, 27]); anterior cingulate cortex (L: [−6, 33, 21]); insula (L: [−30, 27, −3]; R: [36, 27, −6], [48, 12, −6]); superior medial gyrus (L: [−9, 24, 36], [0, 18, 42]); fusiform

gyrus (L: [−45, −60, −18]); anterior ventral bank of intraparietal sulcus (L: [−48, −36, 39]), superior parietal lobule (L: [−27, −75, 45]); supramarginal gyrus (R: [60, −36, 27], [54, −39, 21]); inferior occipital gyrus (L: [−42, −72, −9]); middle occipital gyri (L: [−36, −78, 6]; R: [33, −63, 36], [39, −75, 6]); superior occipital gyrus (R: [30, −66, 39]); inferior temporal gyrus (R: [45, −72, −3]); middle temporal gyri (L: [−48, −72, 12]; R: [45, −72, 12], [54, −51, 0], [60, −45, 3]); superior temporal gyrus (R: [42, −36, 6], [60, −45, 12]); and right pallidum (R: [18, 3, 0]).

Excluding task performance (i.e., 2-back performance accuracy) as a covariate had a negligible effect on the heritability estimates, with the minimum (0%), maximum (65%), and mean heritability (23%) across tested voxels remaining the same, and the number of voxels reaching significant heritability (i.e., surviving cluster thresholding) increasing by only 173 voxels, to 8397 voxels.

To assess whether task performance in this study was due to genetic factors, we fitted univariate ACE models (for confidence intervals, see supplemental Fig. 1a, available at www.jneurosci.org as supplemental material) to the performance measures with normal distributions: 0-back response time, 2-back performance, and FIQ. The distribution for 0-back accuracy showed a strong positive skew, reflecting the low level of difficulty for this condition, and this variable was therefore not modeled. Table 2 shows the model-fitting results and percentage of variance estimates for the performance phenotypes. Performance on the 2-back condition was found to be strongly influenced by genes; accuracy in particular (57%), but response time as well (34%).

Figure 3. Variance component estimates for *n*-back task-related brain activation. *A*, Percentages of variance explained by genetic (a^2) and unique environmental factors (e^2). *B*, Probability map for a^2, indicating which genetic estimates were significant after height ($p < 0.05$) and cluster (>147 voxels) thresholding.

Test-retest reliability

We found high reproducibility of task-related BOLD signal for a subsample of 40 twins rescanned ∼3 months after their initial scan. Voxelwise intraclass correlations between *t* scores at time points 1 and 2 ranged between 0.7 and 0.9 for most activated areas, as shown in Figure 2c (for confidence intervals and *p* values, see supplemental Fig. 3, available at www.jneurosci.org as supplemental material). The paired-samples *t* test provided additional support for the reliability of our measure. Across the group of 40 twins, merely 28 voxels were significantly different between time points 1 and 2 ($p < 0.05$, FWE corrected, two-tailed), clustered around the following peak locations: [30, −63, 57], $k = 12$ voxels; [−3, 6, 60], $k = 12$ voxels; [−33, 27, 18], $k = 2$ voxels; and [−33, −60, −21], $k = 2$ voxels.

Performance measures were significantly reproducible as well, with intraclass correlations between time points 1 and 2 in the moderate to high range ($p < 0.01$, two-tailed) at 0.73 for 0-back accuracy, 0.79 for 0-back reaction time (RT), 0.77 for 2-back accuracy, and 0.56 for 2-back RT, with a significant improvement in response accuracy in the 2-back condition ($t = −3.376$, df =
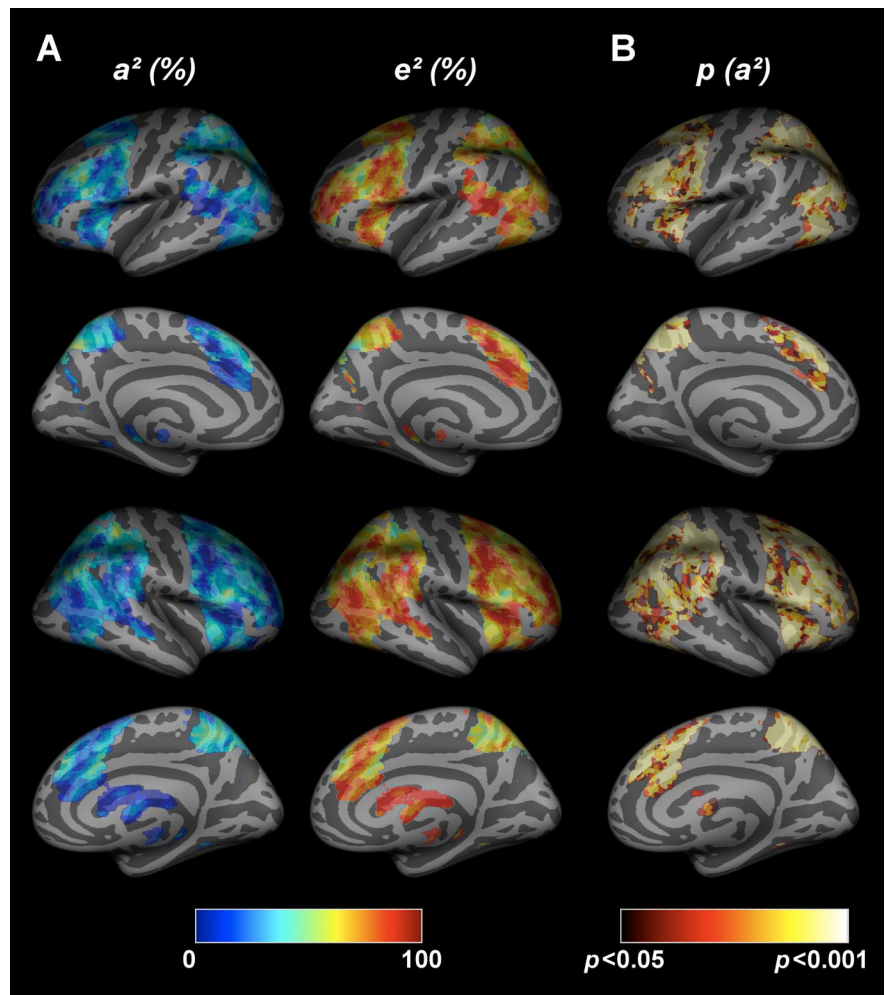
39, $p < 0.05$). This practice effect was stronger for females ($t = −2.859$, df = 19, $p < 0.05$) than for males ($t = −1.845$, df = 19, $p = 0.08$), consistent with the finding that performance at the first time point was slightly lower in females than in males, and suggesting a "catch-up" effect.

Influences of sex, age, and other covariates on task-related brain activation

We also quantified and analyzed the effects of sex, age, 2-back performance accuracy, and FIQ on the BOLD response (Fig. 4a; supplemental Table 4, available at www.jneurosci.org as supplemental material), which may mediate individual variation. As shown in Figure 4b, all covariates survived cluster thresholding, but sex had the strongest effect on brain activation (supplemental Table 1, available at www.jneurosci.org as supplemental material, shows that sex influenced task performance and FIQ as well). At a threshold of 147 voxels, one large cluster and five smaller clusters were significant, totaling 3613 voxels. Males had significantly stronger activation than females in several areas, but differences were most pronounced in inferior frontal gyrus (pars opercularis) (R: [51, 12, 27], [54, 15, 30]); middle frontal gyri (L: [−27, 6, 60]; R: [24, 33, 33], [33, 39, 24], [36, 3, 57], [36, 45, 24], [48, 12, 42]); superior frontal gyri (L: [−24, −3, 54], [−12, 6, 54]; R: [24, 3, 63], [24, 12, 54], [27, 3, 63]); precentral gyri, including

Table 2. Twin correlations and variance component estimates for task performance

Phenotypes	Twin correlations (95% CI)		Model fit: −2LL (Δχ2 (Δdf), p value)				ACE estimates (%) full ACE model (95% CI)			ACE estimates (%) best fitting model (95% CI)	
	MZ	DZ	ACE	AE	CE	E	A	C	E	A	E
0-Back											
Accuracy (%)	0.33* (0.13, 0.50)	0.08 (0.00, 0.31)									
RT (ms)	0.54* (0.38, 0.66)	0.19 (0.00, 0.42)	3464.94	**3464.94 (0.00 (1), ns)**	3471.29 (6.34 (1), <0.05)	3497.69 (32.75 (2), <0.001)	53 (14, 66)	0 (0, 33)	47 (34, 63)	53 (37, 66)	47 (34, 63)
2-Back											
Accuracy (%)	0.55* (0.38, 0.68)	0.34* (0.12, 0.53)	2732.06	**2732.42 (0.36 (1), ns)**	2734.96 (2.90 (1), ns)	2765.43 (33.37 (2), <0.001)	42 (0, 68)	13 (0, 52)	45 (32, 62)	57 (41, 69)	44 (31, 59)
RT (ms)	0.29* (0.11, 0.44)	0.20 (0.00, 0.35)	3895.92	**3895.92 (0.00 (1), ns)**	3896.52 (0.60 (1), ns)	3907.90 (11.98 (2), <0.01)	34 (0, 50)	0 (0, 34)	66 (50, 87)	34 (15, 50)	66 (50, 85)
FIQ	0.77* (0.67, 0.84)	0.52* (0.32, 0.66)	2371.02	**2373.00 (1.98 (1), ns)**	2380.60 (9.58 (1), <0.01)	2455.34 (84.32 (2), <0.001)	50 (18, 82)	27 (0, 56)	23 (16, 33)	78 (68, 84)	22 (16, 32)

Maximum-likelihood twin correlations, univariate genetic model fitting results, and variance component estimates from the full ACE and best fitting models for the performance measures and FIQ. The best fitting model is shown in bold. Nested submodels are compared to the full ACE model by testing whether dropping a parameter resulted in a significant increase in the goodness-of-fit χ^2 (the difference in −2 times the log likelihood (−2LL) of a model and a nested submodel follows a χ^2 distribution with degrees of freedom equal to the difference in the number of parameters). Estimates are corrected for age and sex; assumption testing supported homogeneity of means and variances across birth order and zygosity. −2LL = minus 2*log-likelihood; $\Delta\chi^2$ = change in χ^2; Δdf = change in degrees of freedom; CI, confidence interval; ns, not significant. *p < 0.05.

Brodmann area 6/SMA (L: [−48, 6, 42], [−45, 0, 48], [−36, −3, 42] [−12, 0, 66], [−3, 9, 54]; R: [27, −3, 48], [30, 0, 51], [36, −3, 48], [42, 6, 51]); middle cingulate cortex (L: [−3, 21, 33]; R: [3, 21, 36]); superior parietal lobule, including precuneus (L: [−24, −66, 54], [−18, −66, 45], [−15, −72, 51], [−9, −60, 60], [−3, −69, 48]; R: [3, −54, 57], [6, −45, 66], [24, −72, 48]); middle temporal gyrus (R: [45, −57, 9]); middle occipital gyrus (L: [−24, −75, 27]); superior occipital gyrus (L: [−21, −66, 39], [−18, −75, 42]), with a t-contrast score difference of 0.13–0.70. This is quite a large effect given a mean ± SD voxel t-contrast score of 0.32 ± 0.45 across the working memory network (range, −2.31 to +3.77).

Interestingly, even within the narrow age range of the sample, higher age predicted lower activation in inferior frontal gyri (pars opercularis) (L: [−54, 6, 12], [−54, 12, 27], [−48, 12, 21]; R: [42, 6, 24], [42, 15, 36], [54, 9, 21]); middle frontal gyrus (R: [33, 6, 36], [36, 3, 57], [39, 9, 39]); superior frontal gyrus (R: [18, 12, 60], [27, 6, 60]), precentral gyri (L: [−54, 6, 27], [−51, 9, 33], [−48, 3, 33], [−45, −3, 30], [−36, 3, 30]; R: [36, −3, 51], [39, 3, 45], [45, 6, 48], [48, 0, 33], [51, 6, 30], [51, 9, 36]); insula (L: [−39, 12, 3], [−33, 18, 6], [−30, 27, 3], R: [33, 21, 6], [39, 15, 0]); anterior ventral bank of intraparietal sulcus (L: [−51, −39, 42], [−45, −48, 48], [−45, −45, 33], [−42, −54, 48], [−39, −54, 39], [−36, −48, 42]; R: [33, −48, 30], [39, −42, 36], [42, −42, 42]); supramarginal gyrus (L: [−54, −39, 33], [−51, −45, 42], [−48, −42, 27], [−45, −51, 45]); angular gyrus (R: [33, −54, 42], [36, −69, 42], [42, −45, 36], [48, −60, 21]); superior parietal lobule, including precuneus (L: [−9, −69, 51]; R: [6, −69, 51], [18, −72, 48]); rolandic operculum (L: [−54, 9, 3], [−51, 3, 9], [−45, 3, 15]); middle temporal gyrus (R: [42, −72, 21], [45, −63, 12], [45, −54, 15], [48, −66, 15]), temporal pole (L: [−54, 12, −6], [−54, 12, −3]); middle occipital gyri (L: [−33, −72, 36], [−30, −69, 39], [−33, −66, 30], [−30, −63, 36]; R: [33, −72, 39], [36, −69, 33], [42, −75, 15]); superior occipital gyrus (R: [21, −66, 39], [27, −66, 27]); SMA (R: [12, 12, 48]); caudate nuclei (L: [−18, 12, 21], [−18, 27, 0]; R: [18, 0, 24], [21, 15, 12]); and thalamus (R: [15, −18, 18]), with t-contrast score differences ranging between −0.02 and −0.11 per year. In the prefrontal cortex, the activation seems to move further forward with age, although the positive regression coefficients did not reach significance.

Higher accuracy on the 2-back condition predicted slightly higher activation in several regions. The effect was most pronounced in middle frontal gyrus (R: [30, 6, 60], [36, 6, 48], [42, 9, 48]); inferior frontal gyrus (pars opercularis) (R: [42, 15, 27], [48, 9, 21]); anterior ventral bank of intraparietal sulcus (R: [36, −42, 42], [39, −51, 48]); superior parietal lobule (R: [9, −63, 57], [15, −63, 57], [24, −66, 48], [27, −57, 57]); precuneus (R: [3, −57, 48], [12, −57, 42]); inferior temporal gyrus (R: [51, −51, −9]);

middle temporal gyrus (R: [39, −57, 12], [45, −63, 12], [48, −60, 6], [51, −57, 9]); middle occipital gyrus (R: [30, −66, 33]); superior occipital gyrus (R: [30, −75, 39]); and caudate nucleus (L: [−18, 15, 12], [−15, 24, 3], [−15, −12, 21]), with a 0.004–0.009 increase in t-contrast score difference per percent increase in performance accuracy. Sex differences in BOLD response may partly account for these response accuracy effects, as males, on average, performed better on the task. Compared with the other covariates, the effect of FIQ on WM task-related activation was small, with only positive effects surviving cluster thresholding. Higher FIQ predicted slightly higher activation in the thalamus (L: [−15, −27, 12]) and hippocampus (cornu ammonis; CA) (L: [−30, −33, 0], [−21, −42, 6]), at 0.004–0.012 increase in t-contrast score difference per IQ point. These findings suggest that better task performance and higher IQ do not necessarily imply greater efficiency in the use of brain resources, which is consistent with the work of Klingberg et al. (2002).

Discussion

The heritability of task-related brain activation has not been estimated with any degree of certainty before, as the few previous twin studies have been limited in terms of sample size and analysis. Here we demonstrate for the first time, in the largest twin sample to date, using a voxel-level analysis, and consistent with our preliminary ROI analysis on a small subsample (Blokland et al., 2008), that there is a significant and substantial genetic influence on WM task-related activation across the brain, with genes accounting for up to 65% of the variance, averaging ~33%. Heritable areas include the dorsolateral prefrontal cortex, an area whose function is strongly affected in schizophrenia (Callicott et al., 1998) and other mental disorders. Our heritability estimates are comparable in magnitude to a small-sample ROI study in which genetic influences accounted for 38% of the variance in activation of the dorsal anterior cingulate cortex during an interference processing task (Matthews et al., 2007), as well as a recent family study where up to ~40% of the variance in resting-state functional brain connectivity was due to genes (Glahn et al., 2010). Importantly, we found significant heritability for the task component of interest, WM, whereas Koten et al. (2009) reported a small genetic influence on activation related to an unconstrained distraction phase of the n-back WM task, and in a very small sample (10 MZ twin pairs and their non-twin siblings). Additionally, while considerably lower than the heritability reported for large brain structure, our heritability estimates for task-related activation are of similar magnitude to those for small brain volumes (e.g., hippocampus) (Schmitt et al., 2007). Our findings are also in line with imaging studies showing an association between WM brain function and specific genetic polymor-
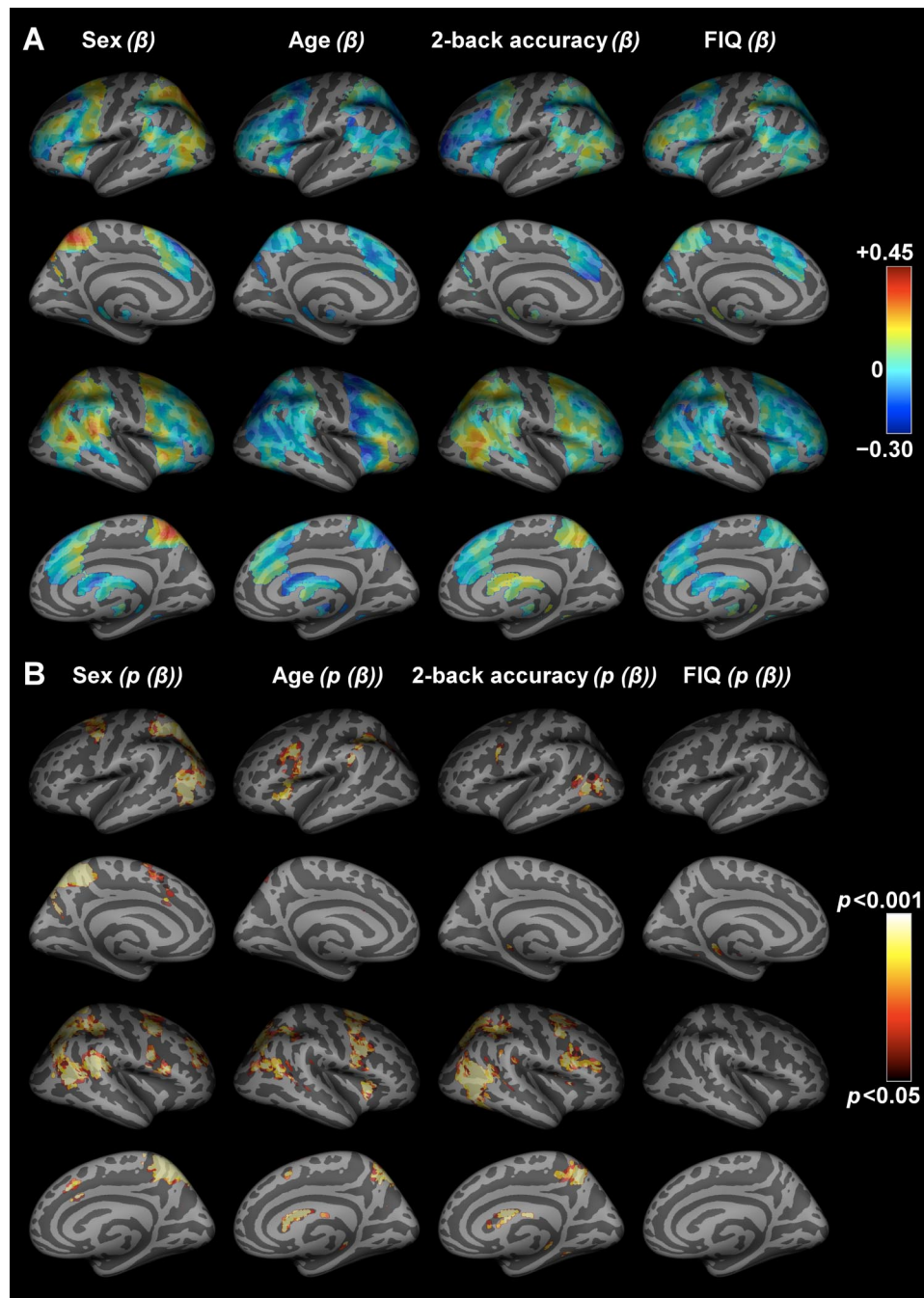
Figure 4. Covariate effect estimates for sex, age, 2-back performance accuracy, and FIQ. ***A***, Standardized regression coefficients (β values) obtained from multiple regression of task-related activation on sex, age, 2-back performance accuracy, and FIQ in Mx (Neale et al., 2002). Positive effects (i.e., greater activation in males, in older participants, and in participants who performed better on the *n*-back task or with higher FIQ) are represented by hot colors and negative effects (i.e., greater activation in females, in younger participants, and in participants who performed worse on the *n*-back task or with lower FIQ) are represented by cold colors. ***B***, Height-thresholded ($p < 0.05$) and cluster-thresholded (>147 voxels) *p* value maps corresponding to the regression coefficient maps of sex, age, 2-back performance accuracy, and FIQ.

phisms, such as GRM3, COMT, and TPH2 (Egan et al., 2004; Tan et al., 2007; Reuter et al., 2008), and are congruent with behavioral studies where a significant proportion of the variance in WM performance has been attributed to genes (Ando et al., 2001; Luciano et al., 2001), with the heritability estimates found here for brain activation being comparable to those for performance (20–65% vs 34–57%).

An alternative explanation that MZ co-twins have more similar task-related brain activation because they are more likely to use the same task strategy than DZ co-twins, as suggested by the

differential association between WM performance and activation of brain areas (Kirchhoff and Buckner, 2006), is unlikely. In a *post hoc* analysis, we assessed strategy use from self-report for a subsample of 139 twins, with results as follows: ~48% used a numeric strategy, 19% used a spatial strategy, and the remainder used a combination of both or neither strategy. This distribution was approximately the same for MZ and DZ twins, but more importantly, MZ co-twins were no more alike in their strategy use ($r = 0.21, p > 0.05$) than DZ co-twins ($r = 0.37, p < 0.01$). Thus, these preliminary data provide no indication that WM brain ac-

tivation from MZ twins is more similar because they adopt the same strategy. Strategy does not appear to be genetic, while there is a strong genetic influence on accuracy and RT. We controlled for possible effects of task performance (i.e., 2-back performance accuracy) on brain activation by including it as a covariate, along with age, sex, and FIQ, although excluding task performance as a covariate had a negligible effect on the heritability estimates and their significance. Because of the low phenotypic correlation between task performance and task-related brain activation, and similarly for FIQ and whole-brain gray matter volume (supplemental Figs. 4, 5, available at www.jneurosci.org as supplemental material), we did not explore whether a common set of genes could explain any covariation among these phenotypes. In the future, it would be interesting to extend the analysis further, in particular to investigate whether there is a genetic covariation between a voxel-level gray matter measure and voxel-level brain activation.

The spatial, numeric version of the *n*-back task used in this study recruits some specific brain regions that other *n*-back task versions may not (Owen et al., 2005). This means that heritability estimates found here may not be generalizable to other (e.g., verbal) *n*-back task versions or to other cognitive paradigms. We could theorize that spatial abilities may be evolutionarily older than verbal abilities and could therefore be more hard-wired in the brain and thus more strongly genetically influenced than verbal abilities. Also, while our twin correlations hinted at the possible presence of genetic dominance in some brain areas, we did not have sufficient power to estimate D separately, so A includes both additive and dominant genetic influences. In an even larger sample, it might be possible to estimate these influences separately. Congruent with studies of cognitive ability that show there is little common environmental variance in adults (Plomin and Spinath, 2004), our fMRI data did not provide any indication for common (shared) environmental influences.

Given the large unique environmental variance estimated in this and the few prior twin fMRI studies, we ascertained the extent to which this variance might be due to measurement error, which places a ceiling on heritability estimates. Prior studies reported mixed fMRI test-retest reliabilities, depending on sample composition, paradigm, brain region, magnetic field strength, scan interval, and analysis methods (Caceres et al., 2009). Our high test-retest reliability suggests it is unlikely that the large E variance component reflects only measurement error; unique environmental factors appear to be involved. Twins were screened for trauma, illness, and drug and medication use before inclusion, so these factors can largely be excluded. Unique environmental factors may include stochastic biological effects or idiosyncratic experiences, such as job stress or tertiary educational course. These factors could have a direct effect on the variance in activation, or an indirect effect, for example by influencing motivation, concentration, or nervousness during fMRI. New synapses may develop in response to unique environmental conditions, in addition to strengthening or discarding existing synapses. Heritability estimates were highest in areas with high test-retest reliabilities, showing the importance of reliable measures when searching for genetic effects. On a related note, it has been suggested that brain regions activated in some individuals only might be better candidates for genetic analysis, because genetic influences on brain activation in areas that are activated similarly among individuals might be underestimated (Koten et al., 2009). However, here we show in our large sample that individual variability within the group activation area is high, with voxel activation intensities approximating normal distributions. Regions outside the group activation area that are activated in just one or a few individuals would pose problems for genetic analysis, since the phenotypic variance in these regions would be too close to zero.

This voxel-based analysis can inform ROI studies. Averaging across an ROI has the advantage of increased signal-to-noise ratio relative to the voxel level; however, the spatial variability of interest may be missed when drawing anatomically defined ROIs, as this assumes that functional divisions and heritability patterning follow anatomical divisions. Future heritability studies will be able to estimate regional heritability more accurately when averaging activation across regions identified here as heritable (and reliable). Genome-wide association studies can furthermore benefit from the reduction in statistical tests associated with analyzing ROIs compared with voxel-level phenotypes.

As sex and age may mediate individual variation, we also quantified their effects on the BOLD response. Our finding of relatively strong sex differences in the brain's response to a WM task confirms the results of Schweinsburg et al. (2005) and Bell et al. (2006), but contradicts those of Schmidt et al. (2009). One theory is that differences in activation are due to gender-specific processing strategies used to accomplish the same cognitive task (Schweinsburg et al., 2005; Clements-Stephens et al., 2009). Another theory, supported by a recent near-infrared spectroscopy study in the prefrontal cortex, is that females possess more efficient hemodynamics during WM (Li et al., 2010). Possibly, different sets of genes explain variance in brain activation for males and females. In a larger sample, heritability could be estimated separately for each sex. Furthermore, areas with significant sex effects partially overlapped with those found for age (age effects were much smaller than sex effects), consistent with prior studies showing differential developmental patterns for males and females for neural activation (Schweinsburg et al., 2005; Clements-Stephens et al., 2009), possibly reflecting the development of gender-specific task strategies. Other possible explanations for age effects include greater neural efficiency in older participants (Schweinsburg et al., 2005), early age-related changes in cerebrovascular characteristics (Kannurpatti et al., 2010), or age-related gray matter decrease (Sowell et al., 2003). The relative influences of genes and environment may differ with age, as shown in a twin study of brain structure (Lenroot and Giedd, 2008), and it may eventually be possible to estimate them separately for different age groups.

In summary, this is the first fMRI study to investigate genetic and environmental influences on task-related brain activation in a combined MZ-DZ twin sample voxel by voxel. It is a substantive extension on our first analysis (Blokland et al., 2008), both in terms of complexity of the analyses and sample size, and provides the strongest support so far for the hypothesis that WM brain activation is heritable. By establishing the heritability of cognitive brain function, this study provides the necessary evidence required before task-related brain activation can be considered as an endophenotype (Gottesman and Gould, 2003) for brain disorders. As the size of our twin sample is increased over the next several years, these genetic brain maps should facilitate discovery of gene variants influencing cognitive brain function through genome-wide association studies, which may provide us with even greater insight into human neurobiology and cognition, and could open up new avenues in the diagnosis and treatment of brain disorders.

References

Ando J, Ono Y, Wright MJ (2001) Genetic structure of spatial and verbal working memory. Behav Genet 31:615–624.

Ashburner J, Friston KJ (1999) Nonlinear spatial normalization using basis functions. Hum Brain Mapp 7:254–266.

Bell EC, Willson MC, Wilman AH, Dave S, Silverstone PH (2006) Males and females differ in brain activation during cognitive tasks. Neuroimage 30:529–538.

Blokland GAM, McMahon KL, Hoffman J, Zhu G, Meredith M, Martin NG,

Thompson PM, de Zubicaray GI, Wright MJ (2008) Quantifying the heritability of task-related brain activation and performance during the N-back working memory task: a twin fMRI study. Biol Psychol 79:70–79.

Caceres A, Hall DL, Zelaya FO, Williams SC, Mehta MA (2009) Measuring fMRI reliability with the intra-class correlation coefficient. Neuroimage 45:758–768.

Callicott JH, Ramsey NF, Tallent K, Bertolino A, Knable MB, Coppola R, Goldberg T, van Gelderen P, Mattay VS, Frank JA, Moonen CT, Weinberger DR (1998) Functional magnetic resonance imaging brain mapping in psychiatry: methodological issues illustrated in a study of working memory in schizophrenia. Neuropsychopharmacology 18:186–196.

Callicott JH, Egan MF, Mattay VS, Bertolino A, Bone AD, Verchinksi B, Weinberger DR (2003) Abnormal fMRI response of the dorsolateral prefrontal cortex in cognitively intact siblings of patients with schizophrenia. Am J Psychiatry 160:709–719.

Clements-Stephens AM, Rimrodt SL, Cutting LE (2009) Developmental sex differences in basic visuospatial processing: differences in strategy use? Neurosci Lett 449:155–160.

Côté C, Beauregard M, Girard A, Mensour B, Mancini-Marïe A, Pérusse D (2007) Individual variation in neural correlates of sadness in children: a twin fMRI study. Hum Brain Mapp 28:482–487.

Drapier D, Surguladze S, Marshall N, Schulze K, Fern A, Hall MH, Walshe M, Murray RM, McDonald C (2008) Genetic liability for bipolar disorder is characterized by excess frontal activation in response to a working memory task. Biol Psychiatry 64:513–520.

Egan MF, Straub RE, Goldberg TE, Yakub I, Callicott JH, Hariri AR, Mattay VS, Bertolino A, Hyde TM, Shannon-Weickert C, Akil M, Crook J, Vakkalanka RK, Balkissoon R, Gibbs RA, Kleinman JE, Weinberger DR (2004) Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. Proc Natl Acad Sci U S A 101:12604–12609.

Eickhoff SB, Paus T, Caspers S, Grosbras MH, Evans AC, Zilles K, Amunts K (2007) Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. Neuroimage 36:511–521.

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th Ed. Harlow, Essex, UK: Longmans Green.

Fischl B, Sereno MI, Dale AM (1999) Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9:195–207.

Fletcher PC, Henson RN (2001) Frontal lobes and human memory: insights from functional neuroimaging. Brain 124:849–881.

Freire L, Roche A, Mangin JF (2002) What is the best similarity measure for motion correction in fMRI time series? IEEE Trans Med Imaging 21:470–484.

Glahn DC, Winkler AM, Kochunov P, Almasy L, Duggirala R, Carless MA, Curran JC, Olvera RL, Laird AR, Smith SM, Beckmann CF, Fox PT, Blangero J (2010) Genetic control over the resting brain. Proc Natl Acad Sci U S A 107:1223–1228.

Gottesman II, Gould TD (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. Am J Psychiatry 160:636–645.

Harrell TH, Honaker LM, Hetu M, Oberwager J (1987) Computerized versus traditional administration of the multidimensional aptitude battery-verbal scale: an examination of reliability and validity. Comput Hum Behav 3:129–137.

Jackson DN (1984) Multidimensional Aptitude Battery. Port Huron, MI: SIGMA Assessment Systems, Inc.

Kannurpatti SS, Motes MA, Rypma B, Biswal BB (2010) Neural and vascular variability and the fMRI-BOLD response in normal aging. Magn Reson Imaging 28:466–476.

Kirchhoff BA, Buckner RL (2006) Functional-anatomic correlates of individual differences in memory. Neuron 51:263–274.

Klingberg T, Forssberg H, Westerberg H (2002) Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. J Cogn Neurosci 14:1–10.

Koten JW Jr, Wood G, Hagoort P, Goebel R, Propping P, Willmes K, Boomsma DI (2009) Genetic contribution to variation in cognitive function: an fMRI study in twins. Science 323:1737–1740.

Lenroot RK, Giedd JN (2008) The changing impact of genes and environment on brain development during childhood and adolescence: initial findings from a neuroimaging study of pediatric twins. Dev Psychopathol 20:1161–1175.

Li T, Luo Q, Gong H (2010) Gender-specific hemodynamics in prefrontal cortex during a verbal working memory task by near-infrared spectroscopy. Behav Brain Res 209:148–153.

Luciano M, Wright MJ, Smith GA, Geffen GM, Geffen LB, Martin NG (2001) Genetic covariance among measures of information processing speed, working memory, and IQ. Behav Genet 31:581–592.

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

Macey PM, Macey KE, Kumar R, Harper RM (2004) A method for removal of global effects from fMRI time series. Neuroimage 22:360–366.

Matsuo K, Glahn DC, Peluso MA, Hatch JP, Monkul ES, Najt P, Sanches M, Zamarripa F, Li J, Lancaster JL, Fox PT, Gao JH, Soares JC (2007) Prefrontal hyperactivation during working memory task in untreated individuals with major depressive disorder. Mol Psychiatry 12:158–166.

Matthews SC, Simmons AN, Strigo I, Jang K, Stein MB, Paulus MP (2007) Heritability of anterior cingulate response to conflict: an fMRI study in female twins. Neuroimage 38:223–227.

McMillan J, Beavis A, Jones FL (2009) The AUSEI06: a new socioeconomic index for Australia. J Sociol 45:123–149.

Neale MC, Maes HHM (2004) Methodology for genetic studies of twins and families. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Neale MC, Baker SM, Xie G, Maes HH (2002) Mx: statistical modeling, 6th Ed. Richmond, VA: Department of Psychiatry, University of Virginia.

Owen AM, McMillan KM, Laird AR, Bullmore E (2005) N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum Brain Mapp 25:46–59.

Plomin R, Spinath FM (2004) Intelligence: genetics, genes, and genomics. J Pers Soc Psychol 86:112–129.

Polderman TJ, Gosso MF, Posthuma D, Van Beijsterveldt TC, Heutink P, Verhulst FC, Boomsma DI (2006) A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence. Acta Neurol Belg 106:191–207.

Polk TA, Park J, Smith MR, Park DC (2007) Nature versus nurture in ventral visual cortex: a functional magnetic resonance imaging study of twins. J Neurosci 27:13921–13925.

Reuter M, Esslinger C, Montag C, Lis S, Gallhofer B, Kirsch P (2008) A functional variant of the tryptophan hydroxylase 2 gene impacts working memory: a genetic imaging study. Biol Psychol 79:111–117.

Schmidt H, Jogia J, Fast K, Christodoulou T, Haldane M, Kumari V, Frangou S (2009) No gender differences in brain activation during the N-back task: an fMRI study in healthy individuals. Hum Brain Mapp 30:3609–3615.

Schmitt JE, Eyler LT, Giedd JN, Kremen WS, Kendler KS, Neale MC (2007) Review of twin and family studies on neuroanatomic phenotypes and typical neurodevelopment. Twin Res Hum Genet 10:683–694.

Schweinsburg AD, Nagel BJ, Tapert SF (2005) fMRI reveals alteration of spatial working memory networks across adolescence. J Int Neuropsychol Soc 11:631–644.

Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–428.

Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW (2003) Mapping cortical change across the human life span. Nat Neurosci 6:309–315.

Tan HY, Chen Q, Sust S, Buckholtz JW, Meyers JD, Egan MF, Mattay VS, Meyer-Lindenberg A, Weinberger DR, Callicott JH (2007) Epistasis between catechol-O-methyltransferase and type II metabotropic glutamate receptor 3 genes on working memory brain function. Proc Natl Acad Sci U S A 104:12536–12541.

Vaughan J (1999) Radiofrequency coils for imaging and spectroscopy. U.S. Patent 5,886:596.

Visscher PM (2004) Power of the classical twin design revisited. Twin Res 7:505–512.

Wishart HA, Saykin AJ, McDonald BC, Mamourian AC, Flashman LA, Schuschu KR, Ryan KA, Fadul CE, Kasper LH (2004) Brain activation patterns associated with working memory in relapsing-remitting MS. Neurology 62:234–238.

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

Zaitsev M, Steinhoff S, Shah NJ (2003) Error reduction and parameter optimization of the TAPIR method for fast T1 mapping. Magn Reson Med 49:1121–1132.

Zeng H, Constable RT (2002) Image distortion correction in EPI: comparison of field mapping with point spread function mapping. Magn Reson Med 48:137–146.