doi:10.1093/brain/awv106 BRAIN 2015: 138; 2087–2102 | 2087

BRAIN
A JOURNAL OF NEUROLOGY

Brain structure–function associations in multi-generational families genetically enriched for bipolar disorder

Scott C. Fears,[1] Remmelt Schür,[2] Rachel Sjouwerman,[3] Susan K. Service,[1] Carmen Araya,[4] Xinia Araya,[4] Julio Bejarano,[4] Emma Knowles,[5] Juliana Gomez-Makhinson,[1] Maria C. Lopez,[6] Ileana Aldana,[1] Terri M. Teshiba,[1] Zvart Abaryan,[1] Noor B. Al-Sharif,[1] Linda Navarro,[1] Todd A. Tishler,[1] Lori Altshuler,[1] George Bartzokis,[1] Javier I. Escobar,[7] David C. Glahn,[5] Paul M. Thompson,[1] Carlos Lopez-Jaramillo,[6] Gabriel Macaya,[4] Julio Molina,[1,8] Victor I. Reus,[9] Chiara Sabatti,[10] Rita M. Cantor,[1,11] Nelson B. Freimer[1] and Carrie E. Bearden[1]

Recent theories regarding the pathophysiology of bipolar disorder suggest contributions of both neurodevelopmental and neurodegenerative processes. While structural neuroimaging studies indicate disease-associated neuroanatomical alterations, the behavioural correlates of these alterations have not been well characterized. Here, we investigated multi-generational families genetically enriched for bipolar disorder to: (i) characterize neurobehavioural correlates of neuroanatomical measures implicated in the pathophysiology of bipolar disorder; (ii) identify brain–behaviour associations that differ between diagnostic groups; (iii) identify neurocognitive traits that show evidence of accelerated ageing specifically in subjects with bipolar disorder; and (iv) identify brain–behaviour correlations that differ across the age span. Structural neuroimages and multi-dimensional assessments of temperament and neurocognition were acquired from 527 (153 bipolar disorder and 374 non-bipolar disorder) adults aged 18–87 years in 26 families with heavy genetic loading for bipolar disorder. We used linear regression models to identify significant brain–behaviour associations and test whether brain–behaviour relationships differed: (i) between diagnostic groups; and (ii) as a function of age. We found that total cortical and ventricular volume had the greatest number of significant behavioural associations, and included correlations with measures from multiple cognitive domains, particularly declarative and working memory and executive function. Cortical thickness measures, in contrast, showed more specific associations with declarative memory, letter fluency and processing speed tasks. While the majority of brain–behaviour relationships were similar across diagnostic groups, increased cortical thickness in ventrolateral prefrontal and parietal cortical regions was associated with better declarative memory only in bipolar disorder subjects, and not in non-bipolar disorder family members. Additionally, while age had a relatively strong impact on all neurocognitive traits, the effects of age on cognition did not differ between diagnostic groups. Most brain–behaviour associations were also similar across the age range, with the exception of cortical and ventricular volume and lingual gyrus thickness, which showed weak correlations with verbal fluency and inhibitory control at younger ages that increased in magnitude in older subjects, regardless of diagnosis. Findings indicate that neuroanatomical traits potentially impacted by bipolar disorder are significantly associated with multiple neurobehavioural domains. Structure–function relationships are generally preserved across diagnostic groups, with the notable exception of ventrolateral prefrontal and parietal association cortex, volumetric increases in which may be associated with cognitive resilience specifically in individuals with bipolar disorder. Although age impacted all neurobehavioural traits, we did not find any evidence of accelerated cognitive decline specific to bipolar disorder subjects. Regardless of diagnosis, greater global brain volume may represent a protective factor for the effects of ageing on executive functioning.

Received September 26, 2014. Revised February 9, 2015. Accepted February 14, 2015. Advance Access publication May 5, 2015

1 Department of Psychiatry and Biobehavioural Science, University of California, Los Angeles, California, USA
2 Academisch Medisch Centrum, Department of Paediatric Neurology/Emma Children's Hospital, Amsterdam, The Netherlands
3 University Medical Centre Utrecht, Neuroscience, Utrecht, The Netherlands
4 Cell and Molecular Biology Research Centre, Universidad de Costa Rica, San Pedro de Montes de Oca, Costa Rica
5 Department of Psychiatry, Yale University and Olin Neuropsychiatric Research Centre, Institute of Living, Hartford Hospital, Hartford, Connecticut, USA
6 Grupo de Investigación en Psiquiatría [Research Group in Psychiatry (GIPSI)], Departamento de Psiquiatría, Facultad de Medicina, Universidad de Antioquia. Medellín, Colombia
7 Department of Psychiatry and Family Medicine, Rutgers-Robert Wood Johnson Medical School, New Brunswick, New Jersey, USA
8 BioCiencias Laboratory, Guatemala, Guatemala
9 Department of Psychiatry, University of California, San Francisco, California, USA
10 Department of Health Research and Policy, Stanford University, Stanford, California, USA
11 Department of Human Genetics, University of California, Los Angeles, California, USA

Correspondence to: Scott C. Fears,
300 Medical Plaza,
Suite 2339,
Los Angeles,
CA 90095,
USA
E-mail: sfears@mednet.ucla.edu

Keywords: bipolar disorder; structural MRI; neurocognition; temperament; pedigrees; component phenotype

Introduction

Converging evidence from the fields of structural neuroimaging and cognitive neuroscience indicates that bipolar disorder impacts neural systems involved in various aspects of cognition and temperament, and highlights the importance of understanding structure–function relationships in bipolar disorder (Bearden *et al.*, 2001; Frangou, 2009). The most replicated neuroimaging findings in bipolar disorder involve decreased global brain volume and increased ventricular volume; more recent studies investigating regionally specific changes have observed cortical thinning of greatest magnitude in the prefrontal and temporal cortices (Rimol *et al.*, 2010; Houenou *et al.*, 2011; Fears *et al.*, 2014). In parallel to imaging studies, neurobehavioural analyses have shown that subjects with bipolar disorder have impairments in cognitive domains involving attention, speeded information-processing, and working and declarative memory, as well as elevated rates of impulsivity, even when in a euthymic mood state (Robinson *et al.*, 2006; Torres *et al.*, 2007; Arts *et al.*, 2008; Balanzá-Martínez *et al.*, 2008; Bora *et al.*, 2009; Sole *et al.*, 2012).

A limited number of brain structure–function relationships in bipolar disorder have been investigated to date, with the goal of identifying altered structure–function associations that may provide clues into the pathophysiology of the disorder. While the majority of brain–behaviour correlations appear to be of similar magnitude and direction in bipolar disorder and healthy subjects (Coffman *et al.*, 1990; Sax *et al.*, 1999; Killgore *et al.*, 2009; Hartberg *et al.*, 2011*a*, *b*; Avery *et al.*, 2014), some studies have found that subjects with bipolar disorder have reduced or inverse relationships relative to those observed in controls, including associations between: anterior cingulate and caudate volumes and performance on executive function measures (Zimmerman *et al.*, 2006; Kozicky *et al.*, 2013), as well as lateral prefrontal cortex volume and increased inhibitory control (Haldane *et al.*, 2008). Conversely, some studies have found brain–behaviour correlations in subjects with bipolar disorder that were not found in controls. Haldane *et al.* (2008) found that controls showed the expected positive correlation between prefrontal grey matter volume and executive function performance, whereas patients with bipolar disorder showed an abnormal association between parietal volume and executive function. Other small studies have additionally found anomalous relationships in patients with bipolar disorder between amygdala volume and long-term memory (Killgore *et al.*, 2009), corpus callosal area and impulsivity (Matsuo *et al.*, 2010), temporal pole thickness and working memory (Hartberg *et al.*, 2011*a*), and ventricular volumes with processing speed and executive functioning (Hartberg *et al.*, 2011*b*). The absence of a correlation in bipolar disorder participants that is otherwise present in controls is generally interpreted as a disruption of a normal structure–function relationship, whereas the presence of a significant correlation that is not observed in controls has been interpreted as a compensatory change that may provide some resilience (e.g. involvement of parietal cortical regions to augment executive functioning; Haldane *et al.*, 2008). However, the collective findings of previous research are difficult to interpret because, with the exception of a few recent studies (Hartberg *et al.*, 2011*a*, *b*), previous investigations have included small sample sizes and have focused on a few select brain and behavioural

measures, with little overlap between studies in choice of measures.

Additionally, a notable challenge for the investigation of brain–behaviour relationships in neuropsychiatric disorders is the effect of age, which is known to have a significant influence on both brain morphology and cognition (Caserta *et al.*, 2009). This is of particular relevance in the study of bipolar disorder, given emerging work suggesting that some aspects of the disorder are related to alterations of neurodevelopmental processes, whereas other consequences of the disorder such as cognitive impairment may be due to progressive brain changes that become more apparent with increasing age (Fries *et al.*, 2012; Schneider *et al.*, 2012; Budni *et al.*, 2013; Gama *et al.*, 2013). Some investigators postulate that toxicity accumulates over the course of the illness, resulting in accelerated neurodegeneration, which could explain observations of increasing clinical severity with disease progression (Budni *et al.*, 2013). Previous studies of brain–behaviour associations in bipolar disorder have not explicitly examined the effect of age on brain–behaviour relationships; therefore it is unclear whether there are differences in the magnitude of these associations across the age span.

The current study aimed to characterize brain–behaviour associations in extended families with heavy genetic loading for bipolar disorder ascertained from two closely related, genetically isolated populations in the Central Valley of Costa Rica (CVCR) and the Antioquia region of Colombia (ANT), which have been the focus of ongoing genetic investigations of bipolar disorder for several decades (Freimer *et al.*, 1996; Carvajal-Carmona *et al.*, 2003; Hong *et al.*, 2004; Herzberg *et al.*, 2006; Service *et al.*, 2006). The disorder is highly penetrant in these families, often affecting multiple first-degree relatives. Additionally, the disorder tends to be severe, with ~70% of subjects with bipolar disorder in our sample reporting a history of psychotic symptoms (Fears *et al.*, 2014). In our most recent study of these families, we reported on the initial analysis of an extensive set of brain and behavioural measures acquired using an intermediate phenotype approach to genetically dissect bipolar disorder. The phenotypic assessments included high-resolution structural neuroimaging and behavioural assessments covering a range of temperament and cognitive domains. In our most recent study, we found that about two-thirds of the measures were heritable and about one-third of the traits were associated with the disease (i.e. significantly different between participants with bipolar disorder and non-bipolar disorder family members). In particular, neuroimaging analysis showed bipolar disorder-associated global volume reduction and cortical thinning in the cortico-cognitive network of the dorsolateral and ventrolateral prefrontal cortex and the ventral-limbic system, involving the hippocampus, amygdala and orbitofrontal cortex. Analysis of the behavioural phenotypes identified bipolar disorder-associated impairments in processing speed, verbal learning and memory, category fluency and inhibitory control. We also found that there was a complex network of phenotypic correlations among the trait measures (Fears *et al.*, 2014). The current study leverages the unique opportunity this dataset presents to investigate brain structure–function relationships across a wide age range, in both affected and unaffected individuals with a homogeneous genetic and environmental background, to provide a more complete description of the brain structure–function network relative to previous studies. Due to the large number of traits investigated, we adopted a multiple-step approach to identify the strongest set of associations. Within this set we then investigated: (i) whether any of the correlations differ in bipolar disorder versus family members without bipolar disorder; (ii) whether individuals with bipolar disorder show evidence of accelerated neurodegeneration; and (iii) whether brain–behaviour relationships differ across the wide age range included in our sample (18 to 87 years). These tests were accomplished using interaction terms in linear regression models. The inclusion of a Brain × Diagnosis interaction term in a linear model was used to test whether a brain–behaviour association differed between family members with and without a bipolar disorder diagnosis, a Diagnosis × Age interaction term was used to test whether the effect of bipolar disorder diagnosis differed across the age span (i.e. accelerated ageing), and the inclusion of a Brain × Age interaction term tested for differences in brain–behaviour associations across age.

Materials and methods

Sample

Neuroimages and neurocognitive assessments were acquired from members of extended pedigrees with heavy genetic loading for bipolar disorder. Subjects were recruited from nuclear families within pedigrees that included at least one member with a confirmed bipolar disorder type I diagnosis, available parents, and at least two siblings without bipolar disorder. Diagnoses were based on DSM-IV criteria, and were established with a best estimate process using Spanish language versions of the Mini International Neuropsychiatric Interview and the Diagnostic Interview for Genetic Studies (DIGS), as described in Fears *et al.* (2014). The study sample included 153 subjects with severe bipolar disorder disorder (BP-1) and 374 of their non-bipolar disorder relatives, ranging in age from 18 to 87 years. The distribution of age, sex, diagnostic and education variables did not differ across sites (Table 1). Written informed consent was obtained from each participant, and the institutional review boards at participating institutions approved all study procedures.

Image acquisition

T_1-weighted structural neuroimages were acquired on 1.5 T scanners from 527 subjects (285 from Costa Rica and 242 from Colombia). In Costa Rica, images were acquired on a Siemens Magnetom Vision 1.5 T machine using a magnetization prepared rapid gradient echo (MPRAGE) sequence. In Colombia, images were acquired on a Philips Gyroscan

Table 1 Study demographics

	Subjects with bipolar disorder		Family members without bipolar disorder	
	CVCR	ANT	CVCR	ANT
n	77	76	208	166
Females	43	46	109	95
Age Range (years)	22–81	27–85	20–87	18–85
Mean Age (SD)	49.2 (12.6)	49.7 (14.1)	50 (15.1)	49.6 (17.8)
Mean years of education (SD)	7.9 (4.7)	8.7 (4.7)	7.8 (5.0)	8.2 (4.5)

Summary statistics for sex, age and education characteristics are listed for bipolar disorder and non-bipolar disorder subjects. ANT = Antioquia region of Colombia site; CVCR = Central Valley of Costa Rica site; SD = standard deviation.

Intera 1.5 T machine using a MPRAGE sequence. At the outset of the study, the two scanners were calibrated by acquiring images from three study personnel who travelled to each site and were scanned at each location. Images were aligned and adjusted for local or global scaling differences to ensure compatibility of sequences across sites. Additionally, during the study period, images were checked for quality control on an ongoing basis and feedback was provided to correct any identified problems. Given these quality assurance steps, site differences were minimal. The greatest difference we observed between the two scanners was due to reduced grey/white contrast in the images acquired in Costa Rica compared to Colombia. Given that the image-processing algorithm uses the grey/white contrast to segment cortical grey matter, it was not surprising that the Costa Rican images tended to have slightly lower values for cortical thickness measures relative to Colombia. On average, cortical thickness across all cortical regions was 3.9% lower in images acquired in Costa Rica. The scanner effect was uniform across age, sex and diagnostic category, providing confidence that including country as a covariate in the linear models used to adjust for site effects was reliable.

Phenotypes

Brain measures

Structural neuroanatomical measures were generated from T_1-weighted images using standard methods from the Freesurfer software package (http://surfer.nmr.mgh.harvard.edu). We implemented a quality assurance pipeline involving manual inspection of intermediate steps within the processing stream to correct any errors and ensure reliable final measures (e.g. manually correcting the white matter segmentation mask to provide a more accurate base to build the tessellated surface mesh). In the first step of the statistical analysis (described below), we adopted a strictly objective approach and included all structural MRI phenotypes derived from the image processing protocol, which included ~90 volume, surface area and cortical thickness measures. For subsequent steps, we reduced the number of tests by focusing on a subset of 37 brain traits that were selected for their relevance to the pathophysiology of bipolar disorder based on both the existing literature and findings in our sample. First, we included all brain measures that our previous analysis showed were significantly associated with bipolar disorder, including global measures (total cortical, total white matter and third ventricle volume), regional volumes (hippocampus, cerebellum, ventral diencephalon and the corpus callosum) and thickness measures from cortical regions that we found to be significantly thinner in subjects with bipolar disorder, including the majority of prefrontal and temporal regions (Fears *et al.*, 2014). Additionally, to provide a basis for comparison of our results to previous work, we included additional brain traits that have been reported to show different patterns of brain–behaviour correlations between subjects with bipolar disorder and healthy controls, specifically: the amygdala, caudate, cingulate gyrus and parietal association cortex (Zimmerman *et al.*, 2006; Haldane *et al.*, 2008; Killgore *et al.*, 2009; Kozicky *et al.*, 2013). The cortical regions selected for analysis are shown in Fig. 1, and include all anterior and posterior association cortices.

Behavioural measures

In addition to neuroimaging, participants were assessed across dimensions of temperament and neurocognition obtained from multiple instruments, including the computerized South Texas Assessment of Neurocognition (Glahn *et al.*, 2007), paper-and-pencil neurocognitive assessment measures and self-report questionnaires (Table 2). To reduce the overall number of tests, we applied the following strategy: (i) sets of variables that tapped into the same behavioural construct were identified; and (ii) if the variables identified in the first step were correlated $>r = 0.6$, we eliminated the variable that showed the weakest association with bipolar disorder. Using this strategy, 38 behavioural measures were selected for the current investigation.

Statistical analysis

Due to the large number of brain–behaviour pairs, we adopted a multi-step approach to identify significant associations.

(i) In the first step, correlations and their standard errors were estimated for all possible pairs of traits (37 brain traits and 38 behavioural traits for a total of 1406 brain–behaviour pairs) using the following criteria: for pairs of heritable traits, the phenotypic correlation (ρ_p) was estimated from their genetic (ρ_G) and environmental (ρ_E) correlations using the SOLAR 6.3.6 software package (Almasy and Blangero, 1998) as: $\rho_p = \rho_G\sqrt{h^2_{trait1}h^2_{trait2}} + \rho_E\sqrt{(1-h^2_{trait1})(1-h^2_{trait2})}$ (Almasy *et al.*, 1997), where $h^2_{trait\ i}$ represents the heritability of that trait. Because of potential problems related to convergence errors, Pearson's correlation coefficients were used for pairs in which at least one trait showed no evidence of heritability in this sample, and the standard error was estimated as: $se = \sqrt{(1-r^2)/(n-2)}$.

(ii) To select brain–behaviour pairs for further study, an approximate *P*-value for the null hypothesis of no correlation was obtained via a Gaussian approximation of the distribution of the ratio of estimated correlation to estimated standard error. A Benjamini-Hochberg false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995) was then applied to these *P*-values with a target level of 0.05.

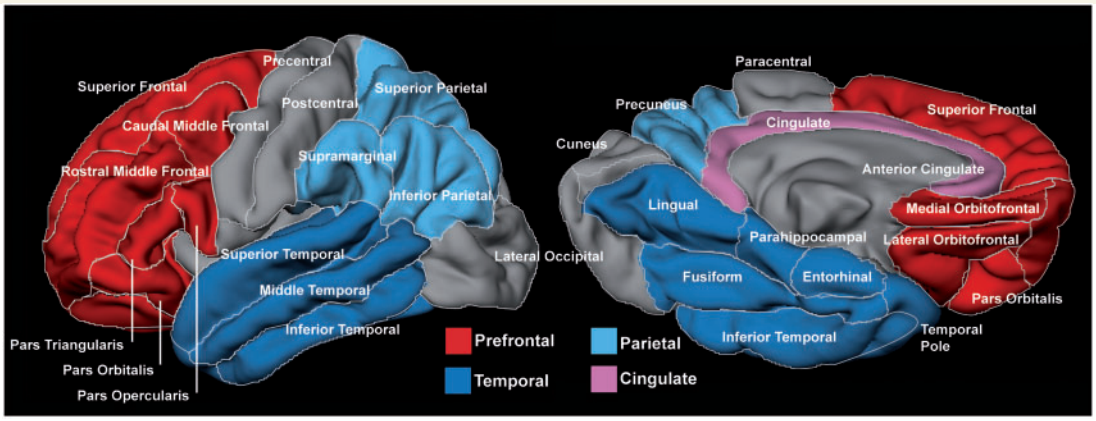

Figure 1 Cortical regions included in linear regression analysis. Cortical regions selected based on the relevance to bipolar disorder pathophysiology are coloured by region: red = prefrontal cortex; dark blue = temporal cortex; light blue = parietal cortex; purple = cingulate cortex. Grey regions were not included in the analysis.

(iii) For the reduced set of pairs identified by the correlation analysis, we then implemented linear regression to identify significant brain–behaviour associations, using the following model:

$$\text{Model 1}: behaviour = A_0 + A_1 brain + A_2 diagnosis + e$$

In this model, A_0 is the estimate of the mean value for the behavioural trait after accounting for covariates, A_1 estimates the association between the brain and behavioural trait and A_2 estimates the effect of diagnosis on the behavioural trait. To identify significant brain–behaviour associations (i.e. test A_1) we implemented a Benjamini-Hochberg false discovery rate (FDR) procedure applying a more stringent target level of 0.01 to minimize type I error.

(iv) Significant brain–behaviour associations identified using Model 1 were then tested with two additional models to determine (a) whether any of the correlations were different in individuals with bipolar disorder versus family members without bipolar disorder (Model 2); and (b) whether brain–behaviour relationships differed as a function of age (Model 3). Differences in the magnitude and direction of brain–behaviour associations between diagnostic groups were tested by including an interaction term (Brain × Diagnosis) in the linear regression model:

$$\text{Model 2}: behaviour = B_0 + B_1 brain + B_2 diagnosis + B_3(brain * diagnosis) + e$$

In this model, B_0 is the estimate of the mean value for the behavioural trait after accounting for covariates, B_1 represents the main effect of the brain measure on the behavioural measure, B_2 estimates the effect of bipolar disorder diagnosis on the behavioural measures and B_3 (interaction term) estimates whether there is a difference in the effect of the brain measure on the behavioural measure as a function of diagnostic group.

To assess the main effect of age on behavioural measures, and its interactions with bipolar disorder diagnosis and brain measures [Step (iv)b from above], the following model was used:

$$\text{Model 3}: behaviour = C_0 + C_1 brain + C_2 diagnosis + C_3 age + C_4(diagnosis * age) + C_5(brain * age) + e$$

In this model, C_0, C_1 and C_2 are analogous to B_0, B_1 and B_2 in Model 2, whereas C_3 estimates the main effect of age on the behavioural measure. The first interaction term, C_4 (Diagnosis × Age), tests whether the effect of age on behaviour differs across diagnostic groups and the second interaction term, C_5 (Brain × Age), tests whether the effect of brain on behaviour differs across the age range of the sample (18–87 years).

Before implementing Models 1 and 2, multiple regression was used to adjust the brain and behavioural measures for relevant covariates; specifically, brain volume measures were regressed on country, sex, age and intracranial volume, cortical thickness measures were regressed on country, sex and age. Additionally, given emerging evidence that obesity is associated with structural brain changes, all brain measures were adjusted for body weight (Walther *et al.*, 2010; Yokum *et al.*, 2012; Bond *et al.*, 2015; Willette and Kapogiannis, 2015). Behavioural measures were regressed on country, sex, age, and education. Before implementing Model 3, the traits were adjusted as in Models 1 and 2, except age was not included in the multiple regressions. Due to the non-independence of pedigree members, regressions were implemented in SOLAR, which uses the pedigree structure to account for the relatedness among individuals. To address the fact that some brain and behavioural traits deviate from the normal distribution, a rank-based procedure was used to inverse normal transform all phenotypes to guard against errors induced by skewed distributions (Van der Waerden, 1952). To compare the magnitude of effect attributable to each covariate for the behavioural traits, the proportion of variance accounted for by each covariate for each behavioural measure was estimated from the R-squared statistic by sequentially adding the individual covariates to a linear regression model.

The significance of each interaction term was evaluated by assuming that twice the difference in log-likelihoods between models with and without the term followed a Chi-squared distribution with one degree of freedom. To determine significance thresholds, we used a Bonferroni correction on each of the three separate families of tests, corresponding to the three types of studied interactions; B_3 in Model 2, and C_4 and C_5 in Model 3.

Table 2 Behavioural measures

Instrument	Subdomain assessed	Phenotype	Measure
Temperament			
TEMPS-A (Akiskal and Akiskal, 2005)	Affective temperament	TEMPS Anxiety	Total score on three anxiety items
		TEMPS Cyclothymia	Total score on 12 cyclothymia items
		TEMPS Depressive	Total score on eight depressive items
		TEMPS Hyperthymia	Total score on eight hyperthymia items
		TEMPS Irritability	Total score eight of irritability items
Aggression Questionnaire (Buss and Perry, 1992)	Impulsivity/Risk-taking	Aggression Questionnaire	Score on 12-item Likert-scale of aggressive traits/behaviours
Barratt Impulsivity Scale (Patton *et al.*, 1995)		Barratt Impulsivity Scale	Score on 30-item Likert-scale assessing frequency of impulsive behaviours
Sensation Seeking Scale (Kolin *et al.*, 1964; Zuckerman and Link, 1968)		Sensation Seeking Scale	Score on 40 items of sensory stimulation preferences
Barron Welsh Art Scale (Barron and Welsh, 1952; Srivastava *et al.*, 2010)	Perceptual creativity	Barron Welsh Art Scale Dislike	Preference rating on simple/symmetric figures of 86 total
		Barron Welsh Art Scale Like	Preference rating on complex/asymmetric figures of 86 total
Peters Delusion Inventory (Peters *et al.*, 2004)	Psychosis-proneness	Peters Delusion Inventory	Score on 40 items assessing delusional ideation and unusual perceptual experiences
Neurocognition			
Abstraction Inhibition and Working Memory (Glahn *et al.*, 2000)	Executive function, working memory	AIM Abstraction	Number of correctly matched shapes presented simultaneously
		AIM Abstraction plus Memory	Number of correctly matched shapes after delayed target presentation
California Verbal Learning Test	Long term memory	CVLT Total Trials 1–5	Number of items recalled over five repeated exposures of a 16 word list
		CVLT Delayed Recall	Number of items out of 16 word list recalled after a 20-min delay
Identical Pairs Continuous Performance Test	Working memory	IPCPT hits	Number of correctly identified pairs on continuous performance test
Penn Conditional Exclusion Test (Kurtz *et al.*, 2004)	Executive function	PCET categories achieved	Number of categories of achieved
		PCET number correct	Number of correctly identified non-matching objects
Spatial Capacity Delayed Response Test	Working memory	SCAP mean number correct all trials	Mean number of correct responses on all trials
Stop Signal Task	Executive function	SST Correct Go	Number of correct go trials
		SST Correct Stop	Number correct stop trials
Stroop Color-Word Interference Test (Stroop, 1992)	Executive function	Stroop Color Word Test Errors	Number of errors on Color-Word test
		Stroop Color Word Test Time	Time needed to complete test
Trail Making Test	Processing speed	Trailmaking Number-Letter Seq. Time	Time needed to connect alternating sequence of numbers and letters
Wechsler Abbreviated Scale of Intelligence	Executive function	WASI Vocabulary	Number of correctly named/defined objects/words
		Matrix Reasoning	Number of correctly completed patterns
Wechsler Memory Scale (Wechsler, 2008)	Long term memory	WMS Logical Memory Immediate	Memory score for auditory story immediately after presentation
		WMS Logical Memory Delay	Memory score for auditory story after 20-min delay
		WMS Visual Reproduction Immediate	Score for visuospatial memory immediately after figure presentation
		WMS Visual Reproduction Delay	Score for visuospatial memory after delay
Miscellaneous (Glahn *et al.*, 2010)	Processing speed, long term memory Verbal fluency, working memory	Digit Symbol Copy	Correctly identified digit-symbol pairs in 90 s
		Digit Symbol Recall	Number of digits recalled when presented with corresponding symbols
		Face Memory	Number of faces recalled from visual presentation after delay
		Verbal Letter Fluency	Words starting with a specific letter generated in 60 s
		Verbal Category Fluency	Animal names generated in 60 s
		VWM Digits Forward number correct	Correctly recalled digits strings in original order of presentation
		VWM Digits Backward number correct	Correctly recalled digits strings in reverse order of presentation
		VWM Letter-Number Seq. number correct	Correctly recalled number-letter strings, in alpha-numeric sequence

Details regarding the 55 behavioural measures acquired from the sample are listed including the instrument, domain, and description of the measure. AIM = Abstraction Inhibition and Memory Test; CVLT = California Verbal Learning Test; IPCPT = Identical Pairs Continuous Performance Test; PCET = Penn Conditional Exclusion Test; SCAP = Spatial Capacity Delayed Response Test; SST = Stop Signal Task; TEMPS = Temperament Evaluation of Memphis, Pisa, Paris and San Diego; TONI = Test of Non-verbal Intelligence; VWM = verbal working memory; WASI = Wechsler Abbreviated Scale of Intelligence; WMS = Wechsler Memory Scale.

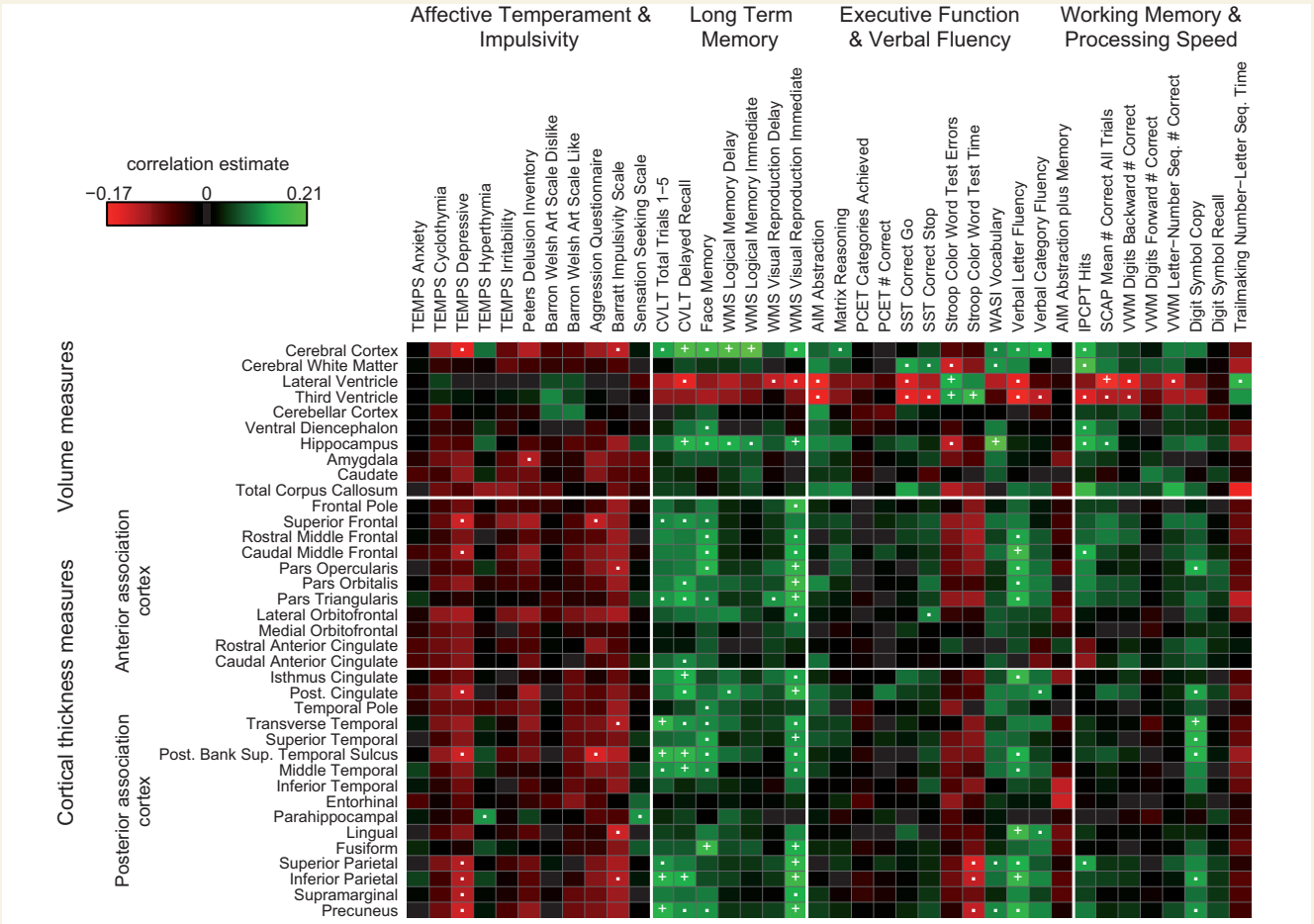

Figure 2 Heat map of brain–behaviour associations. Correlation coefficient estimates of each brain–behaviour association are represented as coloured squares in the heat map. Green indicates positive correlations and red indicates negative correlations. Brain–behaviour pairs that showed evidence of association in the initial filtering step using standard correlation analysis are marked with a white dot. These pairs were selected for follow-up analysis using linear regression models, and associations that were identified as significant using a FDR-corrected alpha of 0.01 are marked with a plus sign. Neuroanatomical measures are listed on the left axis and neurobehavioural measures are listed across the top. Vertical white lines separate the behavioural domains, which are labelled at the top of the plot. A thick horizontal line separates the volume measures (*top rows*) from cortical thickness measures (*bottom rows*) and a thin horizontal line separates the anterior and posterior association regions. AIM = Abstraction Inhibition and Memory test; CVLT = California Verbal Learning Test; IPCPT = Identical Pairs Continuous Performance Test; PCET = Penn Conditional Exclusion Test; SCAP = Spatial Capacity Delayed Response Test; SST = Stop Signal Task; TEMPS = Temperament Evaluation of Memphis, Pisa, Paris and San Diego; VWM = verbal working memory; WASI = Wechsler Abbreviated Scale of Intelligence; WMS = Wechsler Memory Scale.

Results

Analysis of overall brain–behaviour associations

One hundred and forty-seven brain–behaviour pairs were selected in the initial step of the analysis for subsequent testing in linear regression models, and are marked with a dot in Fig. 2. Thirty-two of the 147 pairs showed a significant brain–behaviour association in the linear regression analysis in Model 1 (Fig. 2 and Supplementary Table 1). Twelve of 38 behavioural traits (32%) were significantly associated with at least one brain measure. Overall, after accounting for covariates the correlation estimates were of low magnitude, ranging from 0.13 to 0.21, with a mean of 0.16 (Supplementary Table 1). The proportion of variance in the behavioural trait accounted for by the brain measures, as estimated by the R-squared statistic, is shown in Fig. 3 and was generally less than the variance accounted for by age and education, and roughly the same magnitude as bipolar disorder diagnosis.

For brain measures, greater volume/thickness predicted better performance on all behavioural tasks, whereas increased CSF space (i.e. greater lateral and third ventricle volume) predicted poorer performance. There was some overlap in the pattern of behavioural associations with the cerebral cortex and ventricles on measures of verbal

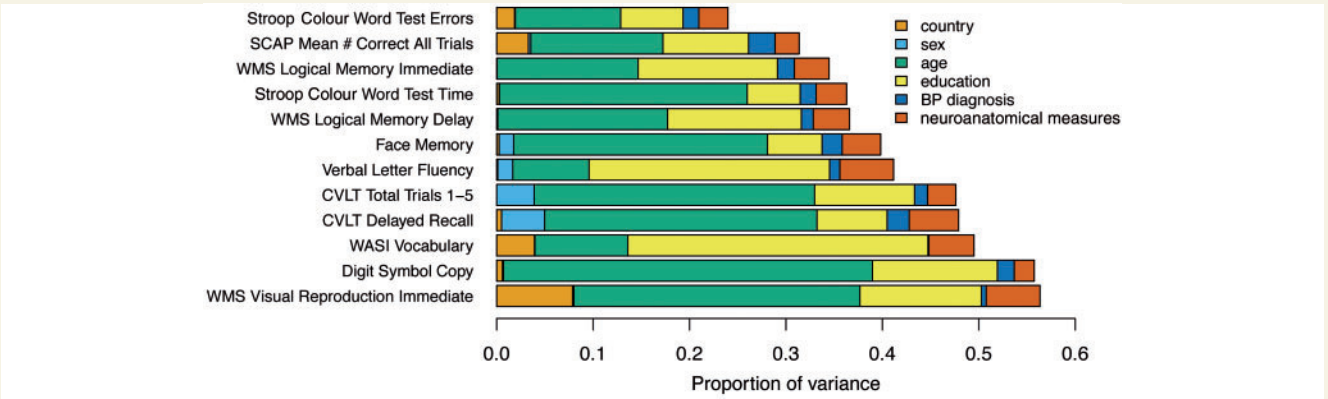

Figure 3 Proportion of variances in neurobehavioural measures accounted for by covariates. The proportion of variance accounted for by each covariate for the 12 behavioural measures that were significantly associated with at least one brain measure in the linear regression models is shown as a stacked bar plot. To determine the portion of variance accounted for by each covariate, linear regressions were used to estimate R-squared starting with a single covariate (country) and then adding additional covariates sequentially in the order shown in each bar plot [country, sex, age, education, bipolar disorder (BP) diagnosis and brain traits]. If more than one brain measure was associated with a behavioural phenotype, all associated brain measures were included in the regression model, so that the red bar represents the proportion of variance accounted for by all brain measures significantly associated with the behavioural phenotype.

fluency and sustained attention/working memory, but in general these two indices were correlated with measures from distinct cognitive domains. Cerebral cortical volume was more associated with long-term memory tasks, whereas ventricular volumes were inversely correlated with performance on working memory and executive function tasks. Hippocampal volume was associated with declarative memory measures, as well as verbal IQ (Wechsler Abbreviated Scale of Intelligence).

In general, cortical thickness measures showed a more restricted pattern of behavioural correlations, specifically with: verbal learning and delayed recall as assessed by the California Verbal Learning Test (CVLT), immediate visual reproduction, as assessed by the Wechsler Memory Scale, face memory, verbal letter fluency and processing speed, as measured by the Digit Symbol test. The declarative memory and verbal fluency measures appeared to associate with both anterior and posterior association regions, although the behavioural correlations appeared more robust in the posterior regions, as evidenced by a greater number of strong associations in the linear regression tests. These data suggest that processing speed is more specifically associated with variation in posterior association cortex volume (temporal and parietal regions).

The initial correlation analysis provided suggestive evidence that greater total cortical volume was inversely correlated with self-reported depressive symptoms on the TEMPS-A (Temperament Evaluation of Memphis, Pisa, Paris and San Diego-autoquestionnaire version) and reduced thickness in the superior and middle frontal gyri, posterior cingulate, posterior temporal sulcus, and all parietal regions were associated with higher TEMPS-A depression scores. Additionally, thinner cortex in adjacent temporal and parietal cortex (transverse temporal and inferior parietal regions) was suggestively associated with higher trait impulsivity on the BIS. However, these relationships did not survive correction for multiple comparisons at the more stringent $P = 0.01$ level.

Four behavioural traits were significantly associated with three or more brain measures in the linear regression tests (Fig. 2). Three of these highly connected behavioural measures were indices of long-term memory, specifically California Verbal Learning Test immediate and delayed recall, and Wechsler Memory Scale immediate visual reproduction. Verbal letter fluency was also associated with multiple neuroanatomical measures, including thickness from anterior and posterior association cortex.

Brain × Bipolar disorder interaction analysis

The 32 significant brain–behaviour associations were tested for differences in magnitude and direction between bipolar disorder and non-bipolar disorder family members using a linear regression, as described by Model 2 which included a Brain × Diagnosis interaction term (Supplementary Table 1). After Bonferroni correction, there were significant Brain × Diagnosis interactions for two pairs of traits: the association between thickness of the pars orbitalis in the inferior frontal gyrus and Wechsler Memory Scale immediate visual memory ($P = 1.6 \times 10^{-4}$) and the association between thickness of the supramarginal gyrus and Wechsler Memory Scale immediate visual memory ($P = 1.5 \times 10^{-3}$). For these pairs of traits, the correlation among participants without bipolar disorder was low, whereas the correlation for individuals with bipolar disorder was of greater magnitude. Details of the chi-square test are presented in Table 3, and the interaction is plotted in Fig. 4. We determined the effect sizes for the significant interaction terms by

Table 3 Brain × Diagnosis interaction analysis for the association between immediate visual memory and thickness measures from the pars orbitalis and supramarginal gryus

Behaviour	Brain	Model	A_0 (SD)	A_1 (SE)	A_2 (SE)	A_3 (SE)	LogLik	χ^2 (*P*-value)
WMS visual memory immediate	Pars orbitalis thickness	Base	0.09 (0.98)	0.14 (0.04)	−0.20 (0.09)		−228.3	
		Full	0.11 (0.97)	0.04 (0.05)	−0.14 (0.09)	0.33 (0.09)	−235.4	14.2 (1.6×10^{-4})
WMS visual memory immediate	Adjusted pars orbitalis thickness*	Base	0.09 (0.68)	0.05 (0.05)	−0.16 (0.06)		−39.28	
		Full	0.09 (0.67)	−0.04 (0.06)	−0.14 (0.06)	0.35 (0.11)	−33.96	10.7 (1.0×10^{-3})
WMS visual memory immediate	Supramarginal thickness	Base	0.10 (0.99)	0.13 (0.4)	−0.21 (0.09)		−236.1	
		Full	0.11 (0.98)	0.04 (0.5)	−0.19 (0.09)	0.28 (0.09)	−244.6	10.1 (1.5×10^{-3})
WMS visual memory immediate	Adjusted supramarginal thickness*	Base	0.09 (0.67)	0.006 (0.07)	−0.16 (.06)		−39.02	
		Full	0.09 (0.67)	−0.12 (0.08)	−0.18 (0.06)	0.39 (0.14)	−50.67	7.3 (6.97×10^{-3})

Parameter estimates for each analysis are shown for the versions of Model 2 used to test for Brain × Diagnosis interaction; the base model: $behaviour = B_0 + B_1 brain + B_2 diagnosis + e$, and the full model: $behaviour = B_0 + B_1 brain + B_2 diagnosis + B_3(Brain \times Diagnosis) + e$. The upper section shows analysis results for the pars orbitalis and the lower section shows results for the supramarginal gyrus. Within each section, the upper rows show the results of the analysis using the unadjusted thickness measures and the lower rows show the results of the analysis on the thickness measure that was adjusted by regressing out the effect of average thickness determined from all cortical regions. The log likelihood for each model is shown in the second-to-last column and the Chi-square and the uncorrected *P*-value comparing the base and full model with one degree of freedom is shown in the last column. WMS = Wechsler Memory Scale; SE = standard error; LogLik = log likelihood. *Thickness measures adjusted for global thickness using an average thickness for all cortical regions as a covariate in a linear regression model.

estimating the proportion of variance explained by each interaction term. The pars orbitalis × diagnosis term accounted for 1.9% and the supramarginal gyrus × diagnosis term accounted for 1.5% of the variance in the Wechsler Memory Scale immediate visual reproduction trait. We undertook a follow-up analysis to disentangle region-specific effects from a global thickness effect. A mean cortical thickness value was derived for each individual subject by averaging the thickness measures from all 33 cortical regions obtained from the Freesurfer package. The two thickness measures were regressed on the mean cortical thickness and the residualized trait was retested with the same linear model (Model 2). The *P*-value for the Brain × Diagnosis interaction was attenuated for both pairs, but these regions still showed a strong signal (Table 3). These findings indicate that although global thickness accounted for some of the associations, both cortical regions seem to have a specific association with visual memory independent of the global signal.

Age × Diagnosis and Brain × Age interaction analysis

As a main effect, age accounted for a significant proportion of the variance for many of the behavioural traits (Fig. 3). Model 3 extended the investigation of the age effect by including two interaction terms, Age × Diagnosis and Brain × Age. None of the behavioural measures included in this analysis showed significant or suggestive evidence for Age × Diagnosis interactions (Supplementary Table 1), indicating that the effect of bipolar disorder diagnosis on each of the traits did not differ across the age range in this sample. Similarly, we tested the neuroanatomical measures for evidence of accelerated decline using linear regression models, but none of the brain measures showed suggestive evidence for Age × Diagnosis interactions (data not shown).

Differences in brain–behaviour associations as a function of age were tested in linear regressions including a Brain × Age interaction term in Model 3. Six pairs of traits showed a significant Brain × Age interaction (Supplementary Table 1). Details of the Chi-square test for three examples are shown in Table 4. Cortical volume showed a weak correlation with verbal letter and category fluency in younger individuals, but was positively correlated with performance with increasing age ($P = 2.5 \times 10^{-4}$ and 1.3×10^{-3}, respectively). Third ventricle volume showed significant Brain × Age interactions for two measures of inhibitory control; the Stop Signal Task correct Go trials ($P = 4.0 \times 10^{-4}$) and Stroop Colour-Word Test Errors ($P = 4.7 \times 10^{-4}$). Lateral ventricle volume also showed a significant Brain × Age interaction with Stroop Colour-Word Test errors ($P = 1.2 \times 10^{-5}$). For these three pairs, the magnitude of the brain–behaviour correlation was low in younger patients but increased with age, such that in older patients, greater ventricular volume predicted poorer performance on these inhibitory control tasks. Cortical thickness in the lingual gyrus also showed a significant Brain × Age interaction with verbal letter fluency, with younger participants showing weak correlations that increased in older patients ($P = 8.6 \times 10^{-4}$). The effect size of interaction terms as measured by the R-squared estimate for each model ranged between 1.2% and 3.3% of the behavioural variance. To demonstrate the difference in these relationships between ages, the brain–behaviour correlations are plotted separately for younger (<55 years) and older (>55 years)

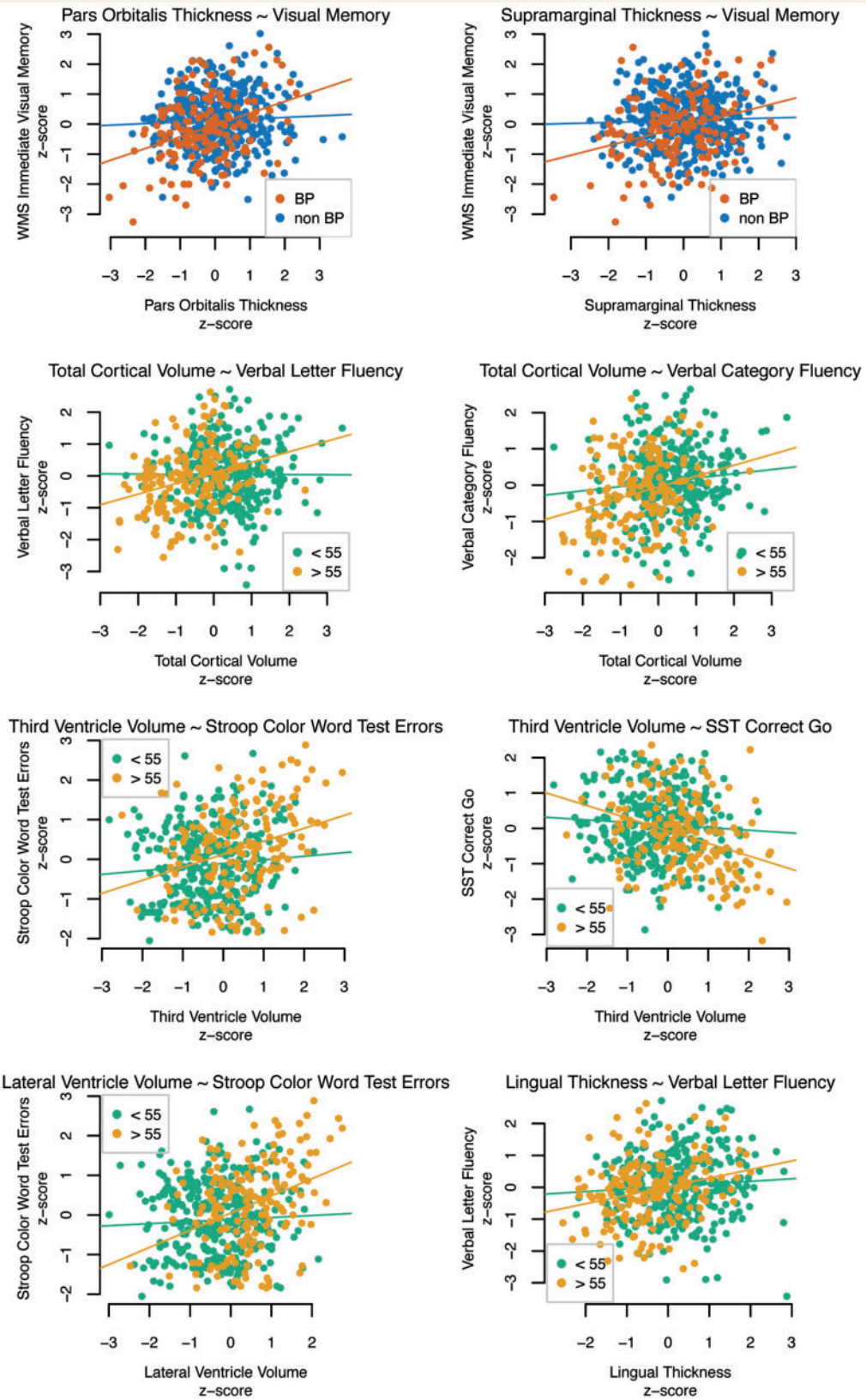

Figure 4 Scatterplots of Brain × Diagnosis and Brain × Age interactions. The upper two panels shows the difference in magnitude of the correlation between bipolar disorder (BP) and non-bipolar disorder (non-BP) family members for Wechsler Memory Scale (WMS) Visual Reproduction (immediate memory) and cortical thickness within the pars orbitalis (*top left*) and supramarginal gyrus (*top right*). The six lower panels show the difference in magnitude of correlation for the six pairs of traits that had significant Brain × Age interactions. To heuristically demonstrate the significant Brain × Age interactions, the brain–behaviour associations were plotted after dividing the subjects into two groups based on age, under 55 (green) and over 55 (yellow) years. SST = Stop Signal Task.

Table 4 Estimates and summary statistics of the Brain × Age interaction analysis

Behaviour	Brain	Model	B_0 (SE)	B_1 (SE)	B_2 (SE)	B_3 (SE)	B_4 (SE)	B_5 (SE)	LogLik	χ^2 (*P*-value)
Verbal letter fluency	Total cortical volume	Base	0.07 (0.78)	0.14 (0.07)	−0.25 (0.08)	1×10^{-3} (2×10^{-3})			−121.39	
		Full	0.15 (0.77)	0.15 (0.07)	−0.29 (0.08)	1×10^{-3} (0.03)	3×10^{-4} (5×10^{-3})	0.01 (2×10^{-3})	−114.23	14.31 (5.5×10^{-4})
Stroop CWT errors	Third ventricle volume	Base	−0.07 (0.84)	0.16 (0.05)	0.25 (0.08)	5×10^{-3} (3×10^{-3})			−154.64	
		Full	−0.13 (0.83)	0.15 (0.05)	0.26 (0.08)	3×10^{-3} (3×10^{-3})	5×10^{-5} (6×10^{-3})	9×10^{-3} (3×10^{-3})	−148.43	12.41 (4.3×10^{-4})
SST correct Go's	Third ventricle volume	Base	0.05 (0.88)	−0.16 (0.05)	−0.32 (0.09)	-4×10^{-3} (3×10^{-3})			−182.92	
		Full	0.11 (0.87)	−0.14 (0.05)	−0.33 (0.09)	-9×10^{-4} (3×10^{-3})	-6×10^{-3} (6×10^{-3})	−0.01 (3×10^{-3})	−176.19	13.46 (2.4×10^{-4})

Details for three examples of the linear regression modelling of age and interaction terms are shown. Parameter estimates for each analysis are shown for the two versions of Model 3 used to test for Diagnosis × Age and Brain × Age interactions; the base model: $behaviour = C_0 + C_1 brain + C_2 diagnosis + C_3 age + e$, and the full model: $behaviour = C_0 + C_1 brain + C_2 diagnosis + C_3 age + C_4(Diagnosis \times Age) + C_5(Brain \times Age) + e$. The log likelihood for each model is shown in the second-to-last column and the Chi-square and the *P*-value comparing the base and full model with one degree of freedom is shown in the last column. CWT = Colour-Word Test; SST = Stop Signal Task; LogLik = log likelihood.

subjects in Fig. 4. Note that this representation of the data is not derived from the linear regression model, which treated age as a continuous variable, but provides a heuristic visualization of the difference in brain–behaviour correlation as a function of age. For these six pairs, a secondary analysis was performed to test whether the Brain × Age interaction differed for bipolar disorder and non-bipolar disorder groups by repeating the linear regression including a three-way interaction term, Brain × Age × Diagnosis, none of which were significant.

Discussion

Here, in a large sample ($n = 527$) ascertained from 26 families with heavy genetic loading for bipolar illness, we found an extensive network of behavioural correlates of brain traits relevant to the pathophysiology of bipolar disorder, which includes a broad range of neurocognitive and temperament traits. The network of structure–function associations in the bipolar disorder pedigrees supports the emerging picture of a distributed set of brain regions contributing to complex behaviour. Many brain areas predicted multiple behaviours across several domains; conversely, many behavioural measures were influenced by multiple brain regions. Global measures of cortical and ventricular volume had more robust associations with behavioural traits relative to local volume and thickness measures, and had distinct cognitive correlates. Specifically, whereas cortical volume was (positively) associated with declarative memory measures, ventricular volume was (inversely) associated with working memory and executive function. Prefrontal, temporal, and parietal grey matter thickness measures had fewer behavioural correlates compared to volume indices, and were more specific to declarative memory, letter fluency, and processing speed. The majority of the behavioural correlates were similar between diagnostic groups, with the exception of ventrolateral prefrontal and supramarginal gyrus cortical thickness, which showed correlations with visual memory that were specific to subjects with bipolar disorder. Our analysis of the impact of age across the broad age range represented in our sample showed that, while age had the expected large effect on all behavioural measures (Fig. 3), there was no evidence that effects of age were different in the participants with bipolar disorder compared to non-bipolar disorder family members. Our analysis of brain–behaviour associations as a function of age (Brain × Age interaction) showed that most brain–behaviour correlations were similar across age groups, with the exception of associations between cortical and ventricular volume and lingual thickness with several measures of executive functioning, which had low correlations in the younger subjects that increased with age.

The results support previous findings in both healthy and clinical populations, indicating that greater grey matter volume and thickness predict better performance, whereas greater ventricular volumes predict poorer performance

(Sullivan *et al.*, 1996; Gur *et al.*, 2000; Antonova *et al.*, 2004; McDaniel, 2005; Hartberg *et al.*, 2010). Our investigation also supports previous studies that have shown that most brain–behaviour correlations are similar between bipolar disorder and non-bipolar disorder subjects, suggesting that overall brain structure–function relationships are similar between bipolar disorder and non-bipolar disorder individuals (Coffman *et al.*, 1990; Hartberg *et al.*, 2011*a*, *b*; Avery *et al.*, 2014). We found little support for prior findings of differences in brain–behaviour correlations between bipolar disorder and healthy controls. Additionally, although our study provides support for the previously identified association between third ventricle volume and inhibitory control measures (Hartberg *et al.*, 2011*b*), we did not find any evidence that the correlation differed between subjects with and without bipolar disorder. However, the discrepancy may be explained by our finding of a significant Brain × Age interaction for these associations, which was not explicitly analysed in the Hartberg *et al.* (2011*b*) study. Finally, in contrast to some previous studies, we did not identify any correlations that were present in non-bipolar disorder individuals, but were absent or reduced in bipolar disorder participants (Zimmerman *et al.*, 2006; Haldane *et al.*, 2008; Kozicky *et al.*, 2013).

Recent theories regarding the pathophysiology of bipolar disorder suggest that the disorder may involve accelerated neurodegenerative processes, highlighting the importance of characterizing the pattern of structure–function relationships across the age span (Fries *et al.*, 2012; Schneider *et al.*, 2012; Budni *et al.*, 2013; Gama *et al.*, 2013). These theories postulate that the cyclic repetition of mood episodes during the course of the illness results in an increasing toxic burden (e.g. oxidative stress) that causes accelerated neurodegeneration. Such theories predict that some cognitive functions would show increased rate of decline with increasing age in subjects with bipolar disorder relative to those without. We tested this hypothesis by including a Diagnosis × Age interaction term in Model 3; bearing in mind the limitations of our cross-sectional design (discussed below), we found no evidence that the effect of the disorder on cognition or brain measures (data not shown) increased in older ages.

It is well established that ageing has a strong effect on brain measures (DeCarli *et al.*, 2005; McDaniel, 2005; Narr *et al.*, 2007; Luders *et al.*, 2009). In contrast to the view espoused by some investigators that bipolar disorder may be associated with accelerated neurodegeneration (Fries *et al.*, 2012; Schneider *et al.*, 2012; Budni *et al.*, 2013; Gama *et al.*, 2013), however, we did not detect evidence of accelerated ageing in individuals with bipolar disorder. We found instead that there is a changing relationship between brain volume/thickness and some aspects of cognitive functioning in different age groups, supporting findings from previous work (Zimmerman *et al.*, 2006; Gautam *et al.*, 2011). Greater total cortical volume, greater lingual thickness, and smaller ventricular volumes predicted better executive functioning in older subjects, regardless of diagnosis. These results are consistent with the hypothesis that greater volumes and/or less atrophy are associated with greater functional capacity, which may provide a buffer against cognitive decline during ageing (Brickman *et al.*, 2006; Zimmerman *et al.*, 2006; Gautam *et al.*, 2011). The Brain × Age interaction appeared similar in participants regardless of bipolar disorder diagnosis, suggesting that the advantages of greater brain volume on executive function are not specific to the disorder. However, in our sample, participants with bipolar disorder tended to have lower cortical and larger ventricular volume compared to non-bipolar disorder family members, suggesting that, on average, individuals with bipolar disorder will have a lower level of functioning.

Previous studies have demonstrated significant variability in functioning of bipolar disorder patients, with a substantial proportion (30–40%) of patients showing little or no evidence of neurocognitive impairment (Altshuler *et al.*, 2004; Burdick *et al.*, 2014). This finding is reflected in our study by the fact that the distributions of neurocognitive measures for bipolar disorder participants show considerable overlap with the distributions from non-bipolar disorder subjects (e.g. overlap of red and blue points for visual memory along the *y*-axis of the upper two panels in Fig. 4). Additionally, the distribution of brain measures for many of the individuals with bipolar disorder fell within the same range as the non-bipolar disorder individuals (e.g. overlap of red and blue points along the *x*-axis for pars orbitalis thickness and supramarginal gyrus in the upper two panels of Fig. 4). Thus, although on average some brain and behavioural measures differed between diagnostic groups, our study demonstrates the variability of brain structure and behavioural functioning within bipolar disorder subjects that may contribute to the diversity of functional outcomes in individuals with the disorder. An additional factor that may be relevant to the functional heterogeneity within individuals with bipolar disorder is the finding that a relatively small proportion of variance in neurobehavioural measures was accounted for by bipolar disorder diagnosis and neuroanatomical traits (Fig. 3). This finding may not be surprising given that individuals with bipolar disorder show less cognitive impairment and brain volume reduction relative to other neuropsychiatric disorders like schizophrenia (Rimol *et al.*, 2010, 2012; De Peri *et al.*, 2012; Ivleva *et al.*, 2013; Anderson *et al.*, 2013). Our findings also highlight the possibility that factors like education, which explain more variance in cognitive functioning in our analysis than do brain and diagnostic measures, may have the potential to compensate for any functional deficits due to bipolar disorder-associated brain changes.

Novel aspects of this study include the largest sample size to date for the analysis of brain–behaviour correlations in bipolar disorder. The multidimensional assessment we employed is the broadest range of temperament and neurocognitive measures ever analysed, to our knowledge, for brain–behaviour associations for any psychiatric disorder.

We identified substantially more correlations compared to previous studies, although the magnitude of correlations identified in this study (once adjusted for confounding variables) was generally lower compared to previous studies. Additionally, the current design is unique relative to previous studies because it was explicitly designed for genetic studies, and will allow for future investigations of the genetic factors that contribute to the brain–behaviour associations. It has been notoriously difficult to identify correlations between behavioural measures and MRI morphometric measures (Boekel *et al.*, 2015). Previous studies have been limited by small sample sizes and generally focused on a limited set of brain regions and behavioural associations. In contrast, the current study leveraged our relatively large sample size and adopted a more agnostic approach to characterize a large matrix of brain–behaviour pairs (Fig. 2). Although the large number of tests inherently limits the power of our approach (as discussed in more detail below), the consistent patterns of associations we identified (e.g. multiple assessments of long-term memory have similar patterns of correlation among cortical regions) provide confidence that we have identified biologically relevant structure–function associations.

It is important to underscore that our study examined a large number of possible interactions in an attempt to follow an objective, unbiased approach. While, for the primary analysis, we selected 147 brain–behaviour pairs that appear to be strongly correlated in our data, we did not use any other prior information to restrict our domain of analysis. In identifying statistically significant results, we increased the power of our approach by using a multi-step design to reduce the number of tests at each subsequent step. This should be taken into account when comparing our results with those of other studies, which focused on a smaller subset of brain and behavioural measures and did not have to contend with similar multiple testing problems. For example, here we did not find a significant association between whole brain volume and IQ, which prior studies have found to be correlated with a magnitude of 0.10 to 0.35 (Frangou *et al.*, 2004; McDaniel, 2005; Narr *et al.*, 2007; Luders *et al.*, 2009). Our analysis tested total cerebral volume for association with two measures representative of IQ; the Matrix Reasoning and Vocabulary scores from the Wechsler Abbreviated Scale of Intelligence. Neither association was identified as significant in the Model 1 linear regressions; however, both of these associations passed the initial filter steps with correlation estimates of $\sim$0.13 (P-values of 0.003 and 0.007, respectively) and would have been considered significant had we focused specifically on these pairs.

A potential limitation of our study is that the ascertainment strategy may have enriched for correlations that are more specific to these bipolar disorder families relative to the general population, and may therefore not generalize to other samples. This concern is attenuated by the fact that the pattern of bipolar disorder-associated brain and behavioural differences in these families is very similar to case-control investigations of independent subjects (Fears *et al.*, 2014). An additional issue that must be considered regarding our ascertainment strategy is that some non-bipolar disorder family members meet criteria for disorders other than bipolar disorder. The most common non-bipolar disorder diagnosis in the families is major depressive disorder, and in the current sample, 73 of 374 non-bipolar disorder family members meet criteria for major depressive disorder. Analysis of the data set excluding the 73 individuals with major depressive disorder (data not shown) was essentially identical to the complete data set indicating that the presence of these family members did not substantially influence the investigation. Compared to case-control designs, the enrichment of other psychiatric disorders in our sample likely reduces power to identify differences between the bipolar disorder and non-bipolar disorder individuals. At the same time, we can have increased confidence that the identified differences are bipolar disorder-specific (and not reflective of more general psychopathology).

Additionally, the large number of potential brain–behaviour pairs in our sample may have limited power to identify Brain $\times$ Diagnosis interactions. Nevertheless, by restricting the analysis of interaction effects to the subset of pairs showing the strongest associations ($n = 37$), we were able to identify significant interactions of moderate effect (i.e. accounting for 1.2–3.3% of the model variance). The multi-step approach we used may have eliminated brain–behaviour pairs in the initial steps of the analysis that may have shown a different pattern of correlation between diagnostic groups in subsequent steps. However, given the relatively large number of traits that did survive the initial steps, our results suggest that altered brain structure–function associations are not a prominent feature of bipolar disorder.

It is also important to note that inferences regarding age-associated changes are limited by the cross-sectional study design. To confirm our finding that there is no evidence of accelerated ageing in subjects with bipolar disorder would require a prospective longitudinal design. Additionally, we identified brain–behaviour correlations that varied as a function of age; however, these differences may not be related to age *per se*, but rather other environmental factors that were shared between cohorts of similarly aged individuals. Similarly, our study design does not allow us to draw inferences regarding causal relationships. Although neuroanatomical measures are generally considered to be 'upstream' of behaviours, these relationships are likely bi-directional. For example, behaviours associated with the disorder (e.g. impulsiveness or social isolation) may lead to environmental exposures that may in turn impact brain measures.

Taken as a whole, and keeping in mind the limitations of our study design, our findings indicate that typical brain structure–function relationships are largely preserved in individuals with bipolar disorder, suggesting that efforts to characterize the pathophysiology of the disorder should focus on delineating impairment of typical brain functions,

rather than identifying anomalous processes unique to bipolar disorder individuals. Furthermore, despite a body of research speculating on differential effects of ageing on the brain in individuals with bipolar disorder (Fries *et al.*, 2012; Schneider *et al.*, 2012; Budni *et al.*, 2013; Gama *et al.*, 2013), within this large sample we did not find evidence for such effects. These findings suggest that, from a clinical staging perspective (Frank *et al.*, 2014; Kapczinski *et al.*, 2014) factors other than chronological age may be more relevant to real-world functioning for patients with bipolar disorder. Ultimately, our aim is to use genetic methods to elucidate the causal biological connections between brain structure and behaviour. In ongoing work we are investigating genotypes and whole genome sequence information to identify both common and rare genetic variants associated with the brain and behavioural phenotypes. Once identified, this information can be used to begin disentangling the complex causal pathways that contribute to the development and manifestations of bipolar disorder (Didelez and Sheehan, 2007; Ebrahim and Davey Smith, 2008).

Funding

This research was supported by National Institute of Health Grants R01MH075007, R01MH095454, P30NS062691 (N.B.F.), K23MH074644-01 (C.E.B.) R01HG006695 (C.S.), and K08MH086786 (S.C.F.), Colciencias and Codi-University Of Antioquia (C.L-J.).

Supplementary material

Supplementary material is available at *Brain* online.

References

Akiskal HS, Akiskal KK. TEMPS: temperament evaluation of Memphis, Pisa, Paris and San Diego. J Affect Disord 2005; 85: 1.

Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 1998; 62: 1198–211.

Almasy L, Dyer TD, Blangero J. Bivariate quantitative trait linkage analysis: pleiotropy versus co-incident linkages. Genet Epidemiol 1997; 14: 953–8.

Altshuler LL, Ventura J, van Gorp WG, Green MF, Theberge DC, Mintz J. Neurocognitive function in clinically stable men with bipolar I disorder or schizophrenia and normal control subjects. Biol Psychiatry 2004; 56: 560–69.

Anderson D, Ardekani BA, Burdick KE, Robinson DG, John M, et al. Overlapping and distinct gray and white matter abnormalities in schizophrenia and bipolar I disorder. Bipolar Disord 2013; 15: 680–93.

Antonova E, Sharma T, Morris R, Kumari V. The relationship between brain structure and neurocognition in schizophrenia: a selective review. Schizophr Res 2004; 70: 117–45.

Arts B, Jabben N, Krabbendam L, van Os J. Meta-analyses of cognitive functioning in euthymic bipolar patients and their first-degree relatives. Psychol Med 2008; 38: 771–85.

Avery S, Williams L, Woolard A, Heckers S. Relational memory and hippocampal function in psychotic bipolar disorder. Eur Arch Psychiatry Clin Neurosci 2014; 264: 199–211.

Balanzá-Martínez V, Rubio C, Selva-Vera G, Martinez-Aran A, Sánchez-Moreno J, Salazar-Fraile J, et al. Neurocognitive endophenotypes (endophenocognitypes) from studies of relatives of bipolar disorder subjects: a systematic review. Neurosci Biobehav Rev 2008; 32: 1426–38.

Barron F, Welsh GS. Artistic perception as a possible factor in personality style: Its measurement by a figure preference test. J Psychol 1952; 33: 199–203.

Bearden CE, Hoffman KM, Cannon TD. The neuropsychology and neuroanatomy of bipolar affective disorder: a critical review. Bipolar Disord 2001; 3: 106–50; discussion 151–3.

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

Boekel W, Wagenmakers E, Belay L, Verhagen J, Brown S, Forstmann BU. A purely confirmatory replication study of structural brain-behavior correlations. Cortex 2015; 66: 115–33.

Bond DJ, Ha TH, Lang DJ, Su W, Torres IJ, Honer WG, et al. Body mass index-related regional gray and white matter volume reductions in first-episode mania patients. Biol Psychiatry 2015; 76: 138–45.

Bora E, Yucel M, Pantelis C. Cognitive endophenotypes of bipolar disorder: a meta-analysis of neuropsychological deficits in euthymic patients and their first-degree relatives. J Affect Disord 2009; 113: 1–20.

Brickman AM, Zimmerman ME, Paul RH, Grieve SM, Tate DF, Cohen RA, et al. Regional white matter and neuropsychological functioning across the adult lifespan. Biol Psychiatry 2006; 60: 444–53.

Budni J, Valvassori SS, Quevedo J. Biological mechanisms underlying neuroprogression in bipolar disorder. Rev Bras Psiquiatr 2013; 35: 1–2.

Burdick KE, Russo M, Frangou S, Mahon K, Braga RJ, Shanahan M, Malhotra AK. Empirical evidence for discrete neurocognitive subgroups in bipolar disorder: clinical implications. Psychol Med 2014; 44: 1–14.

Buss AH, Perry M. The aggression questionnaire. J Pers Soc Psychol 1992; 63: 452–9.

Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, et al. Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. Hum Genet 2003; 112: 534–41.

Caserta MT, Bannon Y, Fernandez F, Giunta B, Schoenberg MR, Tan J. Normal brain aging clinical, immunological, neuropsychological, and neuroimaging features. Int Rev Neurobiol 2009; 84: 1–19.

Coffman JA, Bornstein RA, Olson SC, Schwarzkopf SB, Nasrallah HA. Cognitive impairment and cerebral structure by MRI in bipolar disorder. Biol Psychiatry 1990; 27: 1188–96.

DeCarli C, Massaro J, Harvey D, Hald J, Tullberg M, Au R, et al. Measures of brain morphology and infarction in the framingham heart study: establishing what is normal. Neurobiol Aging 2005; 26: 491–510.

De Peri L, Crescini A, Deste G, Fusar-Poli P, Sacchetti E, Vita A. Brain structural abnormalities at the onset of schizophrenia and bipolar disorder: a meta-analysis of controlled magnetic resonance imaging studies. Curr Pharm Des 2012; 18: 486–94.

Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Stat Methods Med Res 2007; 16: 309–30.

Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? Hum Genet 2008; 123: 15–33.

Fears SC, Service SK, Kremeyer B, Araya C, Araya X, Bejarano J, et al. Multisystem component phenotypes of bipolar disorder for

genetic investigations of extended pedigrees. JAMA Psychiatry 2014; 71: 375–87.
Frangou S. Risk and resilience in bipolar disorder: rationale and design of the Vulnerability to Bipolar Disorders Study (VIBES). Biochem Soc Trans 2009; 37: 1085–9.
Frangou S, Chitins X, Williams SC. Mapping IQ and gray matter density in healthy young people. Neuroimage 2004; 23: 800–5.
Frank E, Nimgaonkar VL, Phillips ML, Kupfer DJ. All the world's a (clinical) stage: rethinking bipolar disorder from a longitudinal perspective. Mol Psychiatry 2014; 20: 23–31.
Freimer NB, Reus VI, Escamilla M, Spesny M, Smith L, Service S, et al. An approach to investigating linkage for bipolar disorder using large Costa Rican pedigrees. Am J Med Genet 1996; 67: 254–63.
Fries GR, Pfaffenseller B, Stertz L, Paz AV, Dargél AA, Kunz M, et al. Staging and neuroprogression in bipolar disorder. Curr Psychiatry Rep 2012; 14: 667–75.
Gama CS, Kunz M, Magalhães PV, Kapczinski F. Staging and neuroprogression in bipolar disorder: a systematic review of the literature. Rev Bras Psiquiatr 2013; 35: 70–4.
Gautam P, Cherbuin N, Sachdev PS, Wen W, Anstey KJ. Relationships between cognitive function and frontal grey matter volumes and thickness in middle aged and early old-aged adults: the PATH Through Life Study. Neuroimage 2011; 55: 845–55.
Glahn DC, Almasy L, Barguil M, Hare E, Peralta JM, Kent JW, et al. Neurocognitive endophenotypes for bipolar disorder identified in multiplex multigenerational families. Arch Gen Psychiatry 2010; 67: 168–77.
Glahn DC, Almasy L, Blangero J, Burk GM, Estrada J, Peralta JM, et al. Adjudicating neurocognitive endophenotypes for schizophrenia. Am J Med Genet B Neuropsychiatr Genet 2007; 144B: 242–9.
Glahn DC, Cannon TD, Gur RE, Ragland JD, Gur RC. Working memory constrains abstraction in schizophrenia. Biol Psychiatry 2000; 47: 34–42.
Gur RE, Cowell PE, Latshaw A, Turetsky BI, Grossman RI, Arnold SE, et al. Reduced dorsal and orbital prefrontal gray matter volumes in schizophrenia. Arch Gen Psychiatry 2000; 57: 761–8.
Haldane M, Cunningham G, Androutsos C, Frangou S. Structural brain correlates of response inhibition in Bipolar Disorder I. J Psychopharmacol 2008; 22: 138–43.
Hartberg CB, Lawyer G, Nyman H, Jönsson EG, Haukvik UK, Saetre P, et al. Investigating relationships between cortical thickness and cognitive performance in patients with schizophrenia and healthy adults. Psychiatry Res 2010; 182: 123–33.
Hartberg CB, Sundet K, Rimol LM, Haukvik UK, Lange EH, Nesvåg R, et al. Brain cortical thickness and surface area correlates of neurocognitive performance in patients with schizophrenia, bipolar disorder, and healthy adults. J Int Neuropsychol Soc 2011a; 17: 1080–93.
Hartberg CB, Sundet K, Rimol LM, Haukvik UK, Lange EH, Nesvåg R, et al. Subcortical brain volumes relate to neurocognition in schizophrenia and bipolar disorder and healthy controls. Prog Neuropsychopharmacol Biol Psychiatry 2011b; 35: 1122–30.
Herzberg I, Jasinska A, García J, Jawaheer D, Service S, Kremeyer B, et al. Convergent linkage evidence from two Latin-American population isolates supports the presence of a susceptibility locus for bipolar disorder in 5q31-34. Hum Mol Genet 2006; 15: 3146–53.
Hong KS, McInnes LA, Service SK, Song T, Lucas J, Silva S, et al. Genetic mapping using haplotype and model-free linkage analysis supports previous evidence for a locus predisposing to severe bipolar disorder at 5q31-33. Am J Med Genet B Neuropsychiatr Genet 2004; 125B: 83–6.
Houenou J, Frommberger J, Carde S, Glasbrenner M, Diener C, Leboyer M, et al. Neuroimaging-based markers of bipolar disorder: evidence from two meta-analyses. J Affect Disord 2011; 132: 344–55.
Ivleva EI, Bidesi AS, Keshavan MS, Pearlson GD, Meda SA, Dodig D, et al. Gray matter volume as an intermediate phenotype for psychosis: Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). Am J Psychiatry 2013; 170: 1285–96.
Kapczinski F, Magalhães PV, Balanzá-Martinez V, Dias VV, Frangou S, Gama CS, et al. Staging systems in bipolar disorder: an International Society for Bipolar Disorders Task Force Report. Acta Psychiatr Scand 2014; 130: 354–63.
Killgore WD, Rosso IM, Gruber SA, Yurgelun-Todd DA. Amygdala volume and verbal memory performance in schizophrenia and bipolar disorder. Cogn Behav Neurol 2009; 22: 28–37.
Kolin EA, Price L, Zoob I. Development of a Sensation-Seeking Scale. J Consult Psychol 1964; 28: 477–82.
Kozicky JM, Ha TH, Torres IJ, Bond DJ, Honer WG, Lam RW, et al. Relationship between frontostriatal morphology and executive function deficits in bipolar I disorder following a first manic episode: data from the Systematic Treatment Optimization Program for Early Mania (STOP-EM). Bipolar Disord 2013; 15: 657–68.
Kurtz MM, Ragland JD, Moberg PJ, Gur RC. The Penn Conditional Exclusion Test: a new measure of executive-function with alternate forms of repeat administration. Arch Clin Neuropsychol 2004; 19: 191–201.
Luders E, Narr KL, Thompson PM, Toga AW. Neuroanatomical correlates of intelligence. Intelligence 2009; 37: 156–63.
Matsuo K, Nielsen N, Nicoletti MA, Hatch JP, Monkul ES, Watanabe Y, et al. Anterior genu corpus callosum and impulsivity in suicidal patients with bipolar disorder. Neurosci Lett 2010; 469: 75–80.
McDaniel M. Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. Intelligence 2005; 33: 337–46.
Narr KL, Woods RP, Thompson PM, Szeszko P, Robinson D, Dimtcheva T, et al. Relationships between IQ and regional cortical gray matter thickness in healthy adults. Cereb Cortex 2007; 17: 2163–71.
Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. J Clin Psychol 1995; 51: 768–74.
Peters E, Joseph S, Day S, Garety P. Measuring delusional ideation: the 21-item Peters et al. Delusions Inventory (PDI). Schizophr Bull 2004; 30: 1005–22.
Rimol LM, Hartberg CB, Nesvåg R, Fennema-Notestine C, Hagler DJ, Pung CJ, et al. Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder. Biol Psychiatry 2010; 68: 41–50.
Rimol LM, Nesvåg R, Hagler DJ, Bergmann O, Fennema-Notestine C, Hartberg CB, Haukvik UK, et al. Cortical volume, surface area, and thickness in schizophrenia and bipolar disorder. Biol Psychiatry 2012; 71: 552–60.
Robinson LJ, Thompson JM, Gallagher P, Goswami U, Young AH, Ferrier IN, et al. A meta-analysis of cognitive deficits in euthymic patients with bipolar disorder. J Affect Disord 2006; 93: 105–15.
Sax KW, Strakowski SM, Zimmerman ME, DelBello MP, Keck PE, Hawkins JM. Frontosubcortical neuroanatomy and the continuous performance test in mania. Am J Psychiatry 1999; 156: 139–41.
Schneider MR, DelBello MP, McNamara RK, Strakowski SM, Adler CM. Neuroprogression in bipolar disorder. Bipolar Disord 2012; 14: 356–74.
Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, et al. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nat Genet 2006; 38: 556–60.
Sole B, Bonnin CM, Torrent C, Martinez-Aran A, Popovic D, Tabarés-Seisdedos R, et al. Neurocognitive impairment across the bipolar spectrum. CNS Neurosci Ther 2012; 18: 194–200.
Srivastava S, Childers ME, Baek JH, Strong CM, Hill SJ, Warsett KS, et al. Toward interaction of affective and cognitive contributors to creativity in bipolar disorders: a controlled study. J Affect Disord 2010; 125: 27–34.
Stroop JR. Studies of interference in serial verbal reactions. Journal of Experimental Psychology. General 1992; 121: 15.

Sullivan EV, Shear PK, Lim KO, Zipursky RB, Pfefferbaum A. Cognitive and motor impairments are related to gray matter volume deficits in schizophrenia. Biol Psychiatry 1996; 39: 234–40.

Torres IJ, Boudreau VG, Yatham LN. Neuropsychological functioning in euthymic bipolar disorder: a meta-analysis. Acta Psychiatr Scand Suppl 2007; 17–26.

Van der Waerden BL. Order tests for the two-sample problem and their power. Indagationes Mathematicae 1952; 14: 458.

Walther K, Birdsill AC, Glisky EL, Ryan L. Structural brain differences and cognitive functioning related to body mass index in older females. Hum Brain Mapp 2010; 31: 1052–64.

Wechsler D. Wechsler Adult Intelligence Scale® 4th Edition (WAIS®-IV). Harcourt Assessment, San Antonio, TX, 2008.

Willette AA, Kapogiannis D. Does the brain shrink as the waist expands? Ageing Res Rev 2015; 20C: 86–97.

Yokum S, Ng J, Stice E. Relation of regional gray and white matter volumes to current BMI and future increases in BMI: a prospective MRI study. Int J Obes (Lond) 2012; 36: 656–64.

Zimmerman ME, Brickman AM, Paul RH, Grieve SM, Tate DF, Gunstad J, et al. The relationship between frontal gray matter volume and cognition varies across the healthy adult lifespan. Am J Geriatr Psychiatry 2006; 14: 823–33.

Zimmerman ME, DelBello MP, Getz GE, Shear PK, Strakowski SM. Anterior cingulate subregion volumes and executive function in bipolar disorder. Bipolar Disord 2006; 8: 281–8.

Zuckerman M, Link K. Construct validity for the sensation-seeking scale. J Consult Clin Psychol 1968; 32: 420–26.