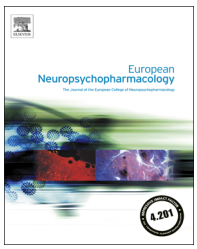
European Neuropsychopharmacology (2015) **25**, 2197–2209

www.elsevier.com/locate/euroneuro



Contribution of genes and unique environment to cross-sectional and longitudinal measures of subcortical volumes in bipolar disorder

Florian Bootsman[a,*], Rachel M. Brouwer[a], Sanne M. Kemner[a], Hugo G. Schnack[a], Astrid C. van der Schot[b], Ronald Vonk[c], Manon H.J. Hillegers[a], Dorret I. Boomsma[d], Hilleke E. Hulshoff Pol[a], Willem A. Nolen[e], René S. Kahn[a], Neeltje E.M. van Haren[a]

[a]*University Medical Center Utrecht Brain Center Rudolf Magnus, Utrecht, The Netherlands*
[b]*Alzheimer Nederland, Amersfoort, The Netherlands*
[c]*Reinier van Arkel Group, 's-Hertogenbosch, The Netherlands*
[d]*Free University Amsterdam, Department of Biological Psychology, Amsterdam, The Netherlands*
[e]*University of Groningen, University Medical Center Groningen, Department of Psychiatry, Groningen, The Netherlands*

Received 12 June 2015; received in revised form 19 August 2015; accepted 28 September 2015

KEYWORDS
Longitudinal twin study;
Bipolar disorder;
Magnetic resonance imaging;
Subcortical brain volumes;
Heritability

Abstract
The influence of genes and environment on the association between bipolar disorder (BD) and volumes of subcortical brain regions involved in emotion processing has rarely been studied. Furthermore, as far as we know, longitudinal twin studies of subcortical brain volume change in BD have not been carried out at all. In this study, we focused on the genetic and environmental contributions to cross-sectional and longitudinal measures of subcortical brain volumes in BD.
A total of 99 twins from monozygotic and dizygotic pairs concordant or discordant for BD and 129 twins from monozygotic and dizygotic healthy control pairs underwent magnetic resonance imaging at baseline. Longitudinal assessment was carried out in 48 twins from monozygotic and dizygotic patient pairs and 52 twins from monozygotic and dizygotic control pairs. Subcortical volume measures were obtained with Freesurfer software and analyzed with structural equation modeling software OpenMx. At baseline, BD was phenotypically and genetically associated with smaller volumes of the thalamus, putamen and nucleus accumbens. BD was not associated with subcortical brain volume change over time in any of the examined regions. Heritability of subcortical volumes at baseline was high, whereas subcortical volume change had low heritability.

*Corresponding author. Internal address: A.01.126, PO Box: 85500, 3508 GA, Utrecht, The Netherlands. Tel.: +31 88 755 8180; fax: +31 88 755 5466.
E-mail address: f.bootsman@umcutrecht.nl (F. Bootsman).

http://dx.doi.org/10.1016/j.euroneuro.2015.09.023

Genes contributing to BD showed overlap with those associated with smaller volumes of the thalamus, putamen and nucleus accumbens at baseline. Further evaluation of genetic contributions to abnormalities in subcortical brain regions assumed to be involved in emotion processing is recommended.

1. Introduction

Abnormal processing of emotion is considered a key feature of bipolar disorder (BD) (Goodwin et al., 2007). In BD, particular attention has been given to abnormalities in subcortical brain regions that are part of or associated with the cortico-striato-thalamic and limbic networks involved in emotion processing, including the amygdala, hippocampus, striatum and thalamus (Phillips et al., 2003, 2008; Emsell and McDonald, 2009; Marchand and Yurgelun-Todd, 2010; Aldhafeeri et al., 2012; Blond et al., 2012; Phillips and Swartz, 2014). However, neuroimaging studies disagree on the extent and variety of morphological abnormalities in BD in some of these regions. For example, both smaller and larger volumes of the amygdala, hippocampus and striatum have been reported (Emsell and McDonald, 2009; Savitz and Drevets, 2009; Rimol et al., 2010; Hajek et al., 2012; Phillips and Swartz, 2014). Furthermore, smaller volumes of the thalamus and nucleus accumbens have been shown in BD (Rimol et al., 2010) although the majority of studies investigating the thalamus report no differences in volume between BD patients and healthy controls (Emsell and McDonald, 2009).

In contrast to the relatively large number of cross-sectional neuroimaging studies investigating subcortical volume, there have only been a few longitudinal studies assessing subcortical volume change over time in BD. Here too, findings are inconclusive. For example, volume preservation, increases and decreases have all been demonstrated in the amygdala and hippocampus (see review by Lim et al. (2013)). There is some evidence of volume increase over time in the caudate nucleus and thalamus (Lisy et al., 2011), although there is limited data available. Moreover, measures of brain volume and brain volume change in BD are influenced by lithium use (often resulting in larger volumes in patients), age, familial load, mood status and variability in imaging methodology (Emsell and McDonald, 2009; Savitz and Drevets, 2009; Hallahan et al., 2011; Hajek et al., 2012), which complicates reliable assessment of subcortical abnormalities associated with the disease.

The heritability of BD has been estimated to be up to 85% (McGuffin et al., 2003). The degree to which genes and environmental factors contributing to BD are associated with subcortical brain volumes has not been studied extensively, and, as far as we know, volume change over time in subcortical structures has not been investigated at all. In one cross-sectional study, the genetic risk for BD was associated with smaller volume of the ventral striatum (McDonald et al., 2004). A different study found larger caudate nuclei in discordant monozygotic BD twin pairs compared to monozygotic healthy control twin pairs (Noga et al., 2001), which suggests that genes associated with BD influence volume of this region. Another group investigating subcortical brain regions in a genetically informative cohort found no association between liability to BD and grey matter of the amygdala-hippocampal complex or thalamus (McIntosh et al., 2006).

As far as we know, it has not been studied whether cross-sectional subcortical volume deficits in twins with BD show progressive change over time and to what extent subcortical volume change is related to genetic and environmental factors associated with the disease. Therefore, in this twin study, we set out to assess whether baseline volume and volume change over time in subcortical brain regions is associated with BD in monozygotic (MZ) and dizygotic (DZ) twin pairs concordant and discordant for the disease. In addition, the degree to which genes and environment influence the association between BD and the subcortical measures at baseline and over time is estimated.

2. Experimental procedures

2.1. Subjects

A total of 99 twins from pairs concordant or discordant for BD (MZ: 15 discordant and 9 concordant pairs; DZ: 20 discordant and 4 concordant pairs, and 1 patient and 2 co-twins from incomplete pairs) and 129 twins from healthy control pairs (MZ: 37 pairs and 2 twins from incomplete pairs; DZ: 25 pairs and 3 twins from incomplete pairs) were included at baseline. BD patients, their co-twins, and 49 healthy control twins were originally recruited by van der Schot et al. (2009), except for 1 DZ bipolar twin pair that was presently included for the first time. Of the remainder of 80 healthy control twins, 18 twins were taken from the cohort that was included by Brans et al. (2008) and 62 twins were taken from the cohort that was included by Baaré et al. (2001). These healthy twins were originally recruited from the (healthy) twin sample of the department of Psychiatry of the University Medical Center Utrecht and the Netherlands Twin Registry (Boomsma, 1998). Between 2010 and 2013 we conducted follow-up measurements with subjects from this total sample. Ultimately, longitudinal assessment of subcortical brain volume change was carried out in 48 twins from patient pairs (MZ: 10 discordant and 2 concordant pairs, and 1 patient and 2 co-twins from incomplete pairs; DZ: 6 discordant and 2 concordant pairs, and 5 co-twins) and 52 twins from control pairs (MZ: 13 pairs and 6 twins from incomplete pairs; DZ: 8 pairs and 4 twins from incomplete pairs). All twins were raised together, except for one control pair who were separated at 12 years of age when both parents died. The subjects were between 18 and 60 years of age at baseline. Cross-sectional and longitudinal demographic information is presented in Tables 1a and 1b. Clinical diagnosis of axis I psychiatric disorders was confirmed using the Structured Clinical Interview for DSM-IV (SCID) (First et al., 1997), for axis II personality disorders using the Structured *Interview For DSM-IV Personality* (SIDP and SCID-II) (Pfohl et al., 1997; First et al., 1997), and for both through available medical records. At both measurements, current mood state was assessed using the Young Mania Rating Scale (YMRS) (Young et al., 1978), Inventory for Depres-

Table 1a Demographic and clinical characteristics of the bipolar and matched healthy control twin pairs at baseline.

	Bipolar patients and their co-twins (n=99)		**Matched control twins** (n=129)	
	MZ[a] (n=48)	**DZ** (n=51)	**MZ**[b] (n=76)	**DZ** (n=53)
Gender, m/f	14/34	18/33	31/45	24/29
Age, y[c]	37.9 (10.5)	43.5 (8.1)	39.0 (9.8)	39.4 (8.0)
Parental educ., y[d]	10.9 (3.5)	11.4 (3.8)	11.4 (3.3)	11.3 (3.5)
Education, y[e]	12.0 (2.1)	13.3 (2.8)	13.7 (2.6)	13.0 (2.7)
Handedness (right/left or both)	34/14	44/7	62/14	44/9
	Patient (MZ)	**Co**	**Patient** (DZ)	**Co**
Onset, age	26.5 (8.9)	-	30.8 (9.8)	-
Birth order, 1st/2nd	15/18	9/6	13/16	13/9
Lithium, *on/off on day MRI, No.*	26/7	0/15	18/11	0/22
Antipsychotics, *on/off on day of MRI*	3/30	0/15	6/23	1/21
Antidepressants, *on/off on day of MRI*	9/24	0/15	8/21	1/21
YMRS score[f]	1.1 (1.5)	0.5 (0.8)	1.0 (3.3)	0.1 (0.7)
IDS score[g]	6.1 (6.6)	2.9 (4.2)	6.0 (8.6)	2.4 (3.7)
Psychotic sympt.	14	1	18	1

Abbreviations: MZ, Monozygotic; DZ, Dizygotic; n=number of individuals.

[a]15 MZ discordant pairs, 9 MZ concordant pairs; 20 DZ discordant pairs, 4 DZ concordant pairs; 1 DZ patient and 2 DZ co-twins from incomplete pairs.

[b]37 MZ healthy control pairs, 2 MZ healthy controls from incomplete pairs; 25 DZ healthy control pairs, 3 DZ healthy controls from incomplete pairs.

[c]Significant effect of group [$F(3,224)=3.7$, $p=0.013$]. Post-hoc test: DZ patient pairs were significantly older than MZ patient pairs [$p=0.016$] and MZ control pairs [$p=0.045$].

[d]Years of parental education could not be determined for 2 bipolar twin pairs (1 MZ, 1 DZ).

[e]Significant effect of group [$F(3,224)=4.8$, $p=0.03$]. Post-hoc test: MZ control pairs had significantly more years of education than MZ patient pairs [$p=0.001$].

[f]YMRS score was not assessed for 2 MZ patients, 2 MZ co-twins, 2 DZ patients and 1 DZ co-twin.

[g]IDS score was not assessed for 2 MZ patients, 2 MZ co-twins, 2 DZ patients and 1 DZ co-twin.

sive Symptomatology (IDS) (Beck et al., 1961) and the Hamilton Depression Rating Scale (HDRS, at follow-up only) (Hamilton, 1960). At baseline, all patients were euthymic (ie, were not in a depressive, manic, or hypomanic episode) or were in an episode in partial remission with a YMRS score of 4 or less and an IDS score of 12 or less, except for two patients who were (hypo)manic (YMRS scores of 5 and 17, respectively) and eight patients who were mildly to severely depressed (IDS scores of 14, 14, 15, 17, 20, 22, 29, 38, respectively). At follow-up, the majority of patients were euthymic except for one patient who was mildly hypomanic (YMRS score of 14) and two other patients who were mildly depressed (HDRS scores of 12 and 18 respectively).

At baseline, the twin pairs had no history of drug or alcohol dependency for the last six months. At follow-up, three patients met diagnostic criteria for alcohol abuse and/or dependence (one of them also for abuse of cannabis, sedatives and morphine), one patient had a cocaine dependency and two co-twins of patients met diagnostic criteria for alcohol abuse in the six months prior follow-up measurement. None had severe medical illness, verified with a medical history inventory.

The healthy control pairs were matched to the bipolar pairs for zygosity, sex, age, parental education, and birth order. At baseline, healthy control pairs had no history of axis I psychiatric disorder or axis II personality disorder according to DSM-IV criteria (SCID and SIDP, respectively) and no history of severe medical illness. Furthermore, they had no first-degree relative with a history of a major axis I psychiatric disorder (DSM-IV) such as schizophrenia, psychotic disorder, mood disorder, anxiety disorder, or substance-related disorder. However, at follow-up, two control subjects met diagnostic criteria for a major depressive episode, two control subjects were diagnosed with an adjustment disorder (one of which in full remission), one control subject was diagnosed with a specific phobia and three control subjects met diagnostic criteria for alcohol abuse. All clinical ratings were carried out by trained and experienced clinicians (psychologists and psychiatrists) and discussed in weekly meetings until consensus was obtained.

The family histories of both the affected and control twins were obtained via the Family Interview Genetic Studies (FIGS) (Nurnberger jr. et al., 1994) performed with both the proband and co-twin. Zygosity was determined by DNA fingerprinting using high polymorphic microsatellite markers 9-11 in the laboratory of the Division Biomedical Genetics, University Medical Center Utrecht. The study was approved by the medical ethics review board of the University Medical Center Utrecht and all participants gave written informed consent after full explanation of the study aims and procedures.

2.2. Brain imaging

Magnetic resonance images were acquired on Philips 1.5 Tesla scanners (at baseline Intera, at follow-up Achieva, Philips, the

Table 1b Demographic and clinical characteristics of the bipolar and matched healthy control twin pairs who were measured at both baseline and follow-up.

	Bipolar patients and their co-twins (n=48)				**Matched control twins** (n=52)	
	MZ[a] (n=27)	**DZ** (n=21)			**MZ**[b] (n=32)	**DZ** (n=20)
Gender, m/f	7/20	10/11			9/23	8/12
Age at baseline, y	37.9 (12.3)	41.0 (7.2)			39.6 (8.6)	39.6 (6.4)
Interval $_{MRI\ t0\text{-}t1}$	7.4 (1.4)	7.9 (1.5)			7.4 (1.5)	7.3 (1.1)
Parental educ., y	12.0 (3.1)	12.3 (3.3)			10.6 (3.5)	12.0 (3.9)
Education, y[c]	12.0 (2.3)	13.9 (1.8)			13.7 (2.2)	13.6 (1.7)
Handedness (right/left or both)	19/8	19/2			28/4	19/1
	Patient (MZ)	**Co**	**Patient (DZ)**	**Co**		
Onset, age	25.4(10.0)	-	31.3(8.1)	-		
Birth order, $_{1st/2nd}$	6/9	8/4	4/6	7/4		
Lithium						
-both time points	8	0	4	0	0	0
-never	3	12	3	11	32	20
-baseline only	3	0	1	0	0	0
-follow-up only	1	0	2	0	0	0
Antipsychotics						
-both time points	2	0	3	1	0	0
-never	8	12	5	9	32	20
-baseline only	1	0	0	0	0	0
-follow-up only	4	0	2	1	0	0
Antidepressants						
-both time points	3	0	1	0	0	0
-never	8	10	7	11	32	20
-baseline only	3	0	1	0	0	0
-follow-up only	1	2	1	0	0	0
$YMRS_{base}$[d]	0.7 (0.9)	0.5 (0.8)	0.6 (0.8)	0.3 (0.9)		
IDS_{base}[e]	6.0 (7.3)	3.2 (4.4)	10.4 (12.3)	2.6 (4.0)		
$YMRS_{follow\text{-}up}$[f]	1.5 (1.6)	-	4.1 (4.0)	-		
$HDRS_{follow\text{-}up}$[g]	2.8 (3.5)	2.1 (2.7)	5.0 (5.7)	1.1 (1.9)		
$GAF_{follow\text{-}up}$[h]	71.4 (15.3)	81.3 (9.8)	67.4 (15.7)	84.5 (16.3)		
Psychotic sympt.	5	1	7	1		
Substance $abuse_{follow\text{-}up}$[i]	3	2	1	-	1	2

Abbreviations: MZ, Monozygotic; DZ, Dizygotic; n= number of individuals.

[a]10 MZ discordant pairs, 2 MZ concordant pairs, 1 MZ patient and 2 MZ co-twins; 6 DZ discordant pairs, 2 DZ concordant pairs and 5 DZ co-twins.

[b]13 MZ healthy control pairs and 6 MZ healthy controls from incomplete pairs; 8 DZ healthy control pairs, 4 DZ healthy controls from incomplete pairs.

[c]Significant effect of group [$F(3,96)=4.4$, $p=0.006$]. Post-hoc test: MZ patient pairs had significantly fewer years of education than DZ patient pairs [$p=0.015$] and MZ control pairs [$p=0.018$].

[d]At baseline, YMRS score was not assessed for 1 MZ patient and 2 MZ co-twins.

[e]At baseline, IDS score was not assessed for 1 MZ patient and 2 MZ co-twins.

[f]At follow-up, YMRS score was not assessed for 3 MZ patients and none of the co-twins. For 3 MZ patients YMRS score was not assessed on the same day of the MRI scan but approximately 2 months after.

[g]At follow-up, HDRS score was not assessed for 4 MZ patients, 2 MZ co-twins, 2 DZ patients and 1 DZ co-twin. For 3 MZ patients HDRS score was not assessed on the same day of the MRI scan but approximately 2 months after.

[h]At follow-up, GAF score was not assessed for 2 MZ patients and 1 DZ patient.

[i]6 months prior to follow-up measurement, 3 MZ patients met diagnostic criteria for alcohol abuse and/or dependency (and one of them also for abuse of cannabis, sedatives and morphine), 1 DZ patient had a cocaine dependency, 2 MZ co-twins of patients met criteria for alcohol abuse, and 1 MZ and 2 DZ control twins met diagnostic criteria for alcohol abuse.

Table 2 Mean uncorrected subcortical volumes for BD patients, co-twins of patients and healthy controls, and between-group comparisons with data uncorrected and corrected for use of lithium and antipsychotics.

Region (volume, in ml)	Mean (SD) uncorrected volume			Statistics							
	Patients[a]	Co-twins	Healthy controls[b]	Patients vs HC[c]				Co-twins vs HC			
				F	*df*	*p*	Δ	*F*	*df*	*p*	Δ
ICV[d]	1511.26 (193.4)	1547.17 (188.99)	1524.01 (134.55)	0.15	*1,189*	0.699		2.39	*1,164*	0.124	
Thalamus[e]	15.02 (1.82)	15.06 (1.55)	15.17 (1.37)	0.09	*1,188*	0.765		0.58	*1,163*	0.449	
				12.56	***1,188***	**0.000***	**pt<ctrl**				
				1.33	*1,188*	0.251					
Caudate Nucleus[f]	6.63 (0.74)	6.91 (0.72)	6.80 (0.79)	0.54	*1,181*	0.465		0.83	*1,161*	0.364	
				2.40	*1,181*	0.123					
				0.12	*1,181*	0.726					
Putamen[g]	9.42 (1.30)	9.62 (1.40)	9.82 (1.18)	2.86	*1,186*	0.092		1.11	*1,161*	0.293	
				19.91	***1,186***	**0.000***	**pt<ctrl**				
				2.57	1,186	0.111					
Pallidum[h]	3.02 (0.38)	3.07 (0.34)	3.12 (0.30)	3.64	*1,185*	0.058		1.58	*1,161*	0.210	
				0.11	*1,185*	0.741					
				2.96	*1,185*	0.087					
Hippocampus[i]	8.32 (0.95)	8.43 (0.97)	8.52 (0.74)	1.57	*1,179*	0.212		0.98	*1,158*	0.323	
				3.17	*1,179*	0.077					
				0.96	*1,179*	0.329					
Amygdala[j]	3.03 (0.42)	2.99 (0.33)	3.07 (0.35)	0.00	*1,189*	0.952		1.42	*1,163*	0.235	
				2.78	*1,189*	0.097					
				0.02	*1,189*	0.904					
Nucleus Accumbens[k]	0.94 (0.15)	0.96 (0.15)	0.98 (0.15)	1.20	*1,188*	0.275		1.27	*1,164*	0.261	
				7.38	***1,188***	**0.007**	**pt<ctrl**				
				1.08	*1,188*	0.299					

Note: Table depicts uncorrected subcortical volumes at baseline (in milliliter). Univariate analysis of variance was performed with group (patients versus healthy controls and co-twins versus healthy controls) as independent variable and subcortical brain volume of each region as dependent variable, after the effects of age, gender and intracranial volume had been regressed out (and the difference in means between Li− and Li+ patients and between Ap− and Ap+ patients had been added to the values of the Li+ or Ap+ group, resulting in volume estimates when no lithium or antipsychotics had been used).

Li +/− =Lithium use yes/no; Ap+/− =Antipsychotic use yes/no.

*Significant at α=0.007 (Bonferroni threshold), *p* values in bold face are significant at α=0.05.

[a]Including concordant pairs.

[b]Including both twins from complete pairs.

[c]In the analysis between patients and controls, the values of the subcortical brain volumes of the patients had been either uncorrected for medication use (first row), corrected for lithium use (second row) or corrected for antipsychotic use (third row).

[d]*n*=62 patients, 37 co-twins, 129 healthy controls.

[e]*n*=62 patients, 37 co-twins, 128 healthy controls (1 subject removed).

[f]*n*=57 patients, 37 co-twins and 126 healthy controls (8 subjects removed).

[g]*n*=62 patients, 37 co-twins and 126 healthy controls (3 subjects removed).

[h]*n*=61 patients, 37 co-twins and 126 healthy controls (4 subjects removed).

[i]*n*=58 patients, 37 co-twins and 123 healthy controls (10 subjects removed).

[j]*n*=62 patients, 36 co-twins and 129 healthy controls (1 subject removed).

[k]*n*=61 patients, 37 co-twins and 129 healthy controls (1 subject removed).

Netherlands). Both scanners were simultaneously used in a large multicenter collaboration, attesting to their cross-scanner reliability (Hibar et al., 2015). Furthermore, in a test set of 6 subjects that were scanned on both scanners on the same day, we calculated intraclass correlations (ICC) between subcortical measures. Here, ICC's were found to be very high (>0.97), although the small size of the test sample and high variability between the subjects should be considered. The imaging parameters were identical across scanners and measurements: T1-weighted 3D fast field echo scans with 160-180 contiguous coronal slices (echo time=4.6 ms, repetition time=30 ms, flip angle=30°, $1 \times 1 \times 1.2$ mm^3 voxels) (van der Schot et al., 2009; Brans et al., 2010). At both assessments, concordant and discordant MZ and DZ patient twins and healthy twins were randomly assigned to MRI slots, eliminating possible between-group biases due to scanner drifts. Processing of brain images and subcortical volumetric segmentation was performed with the FreeSurfer structural imaging pipeline (http://surfer.nmr.mgh.harvard.edu/). Anatomic volumes of the bilateral thalamus, caudate nucleus, putamen, pallidum, hippocampus, amygdala and nucleus accumbens were delineated using information on image intensity, probabilistic atlas location and spatial relationships between subcortical structures (Fischl et al., 2002, 2004). Data

were processed with the cross-sectional (v5.1.0) and longitudinal (v5.3.0) FreeSurfer pipelines. Subcortical brain volumes were extracted from the FreeSurfer output. Based on visual inspection of subcortical segmentations at baseline and follow-up, brain measures were excluded from the analysis if clear segmentation errors were observed. Left and right volumes were added, render ing total volumes of subcortical brain regions. Change in subcortical brain volume was calculated by subtracting the volumes at baseline from the volumes at follow-up (these volumes were obtained with the longitudinal processing stream of FreeSurfer). The resulting individual measures of volume change for each region were then divided by the number of years between measurements for each subject individually, thus yielding a measure of annual subcortical brain volume change.

2.3. Statistical analysis

2.3.1. Quality checking

After quality checking of subcortical brain segments, baseline and change measures were explored for statistical outliers. A number of subjects had outlying values for individual brain measures. If removal of these subjects from analysis did not influence results but did ensure normal distribution of the data, we chose to present results with outliers excluded. However, if outlier removal affected the nature of the association between BD and subcortical volume (change), we presented both the data with and without the outliers removed (see Tables 2 and 3 for the number of subjects that were analyzed for each brain measure).

Furthermore, we assessed whether intracranial volume (ICV) influenced subcortical brain volumes and found that ICV influenced the baseline but not the change measures and was therefore added as a covariate in the baseline analysis only.

2.3.2. Influence of medication use on subcortical brain measures

As the influence of lithium on the brain in BD has been noted repeatedly (Hafeman et al., 2012), we assessed whether lithium use influenced the subcortical brain measures. At baseline, univariate analysis of variance in BD patients with group ('lithium users' and 'lithium non-users') as between-subject variable, and subcortical brain volume as the dependent variable (after the effects of age, gender and intracranial volume had been regressed out) revealed that lithium users had significantly larger volumes of the thalamus ($p=0.012$) and putamen ($p=0.047$) than patients who did not use lithium. Therefore, at baseline, we chose to correct for lithium use by adding the difference in means between non-users and users to the values of subjects from the latter group, obtained after regression on age, gender and intracranial volume. This normalized the volumes of the lithium users to the level of the non-users (van der Schot et al., 2009). For completeness, we also assessed the differences in subcortical brain volumes between patients who used antipsychotics or antidepressants and patients who did not. Although patients who used antipsychotics had a smaller thalamus ($p=0.036$) compared to patients who did not use antipsychotics, correcting for their use did not contribute to volumetric differences between patients and controls (see Supplementary table S1 and Table 2). Based on this and the fact that the influence of antipsychotics and antidepressants on the brain in BD appears to be limited (Hafeman et al., 2012), we chose not to correct for their use.

For the change analysis, correction for medication use was not applied, given the small sample sizes of the respective medication using groups ('medication use at both measurements', 'started using medication during interval', 'quit using medication during interval' and 'never used medication'), yielding low statistical power to reliably detect group differences.

2.3.3. Univariate analysis of variance

At baseline, univariate analysis of variance was performed with group ('BD patients' versus 'healthy controls' and 'co-twins of patients' versus 'healthy controls') as between-subject variables and subcortical brain volume of each region as the dependent variable, after the effects of age, gender and intracranial volume had been regressed out (and the difference in means between Li− and Li+ patients had been added to the values of the latter group, resulting in volume estimates when no lithium been used). For the change measures, we assessed with a linear model whether within-group subcortical volume change in each region as the dependent variable differed significantly from zero, correcting for age at baseline and gender. Subsequently, univariate analysis of variance was performed with group ('BD patients' versus 'healthy controls' and 'co-twins of patients' versus 'healthy controls') as between-subject variable and annual subcortical volume change in each region as the dependent variable, after the effects of age at baseline and gender had been regressed out. Preliminary analysis and rendering of standardized residuals suitable for genetic model fitting was performed with the Statistical Package for the Social Sciences (IBM SPSS, release 21.0.0.0).

2.3.4. Genetic model fitting

To estimate relative genetic and environmental contributions to the association between subcortical brain volume (change) and liability to develop BD, a bivariate liability threshold model was chosen and implemented in structural equation modeling software OpenMx (Kenny et al., 2009), running under the statistical programming environment R (R Development Core Team, 2008). Here, a bivariate Cholesky decomposition was fitted to the standardized residuals of our brain measures (corrected for age, gender and ICV [ICV correction was applied in baseline measures only], as well as lithium use in patients) to estimate additive genetic (A), and unique environmental (E) variance components of brain volume (changes) and phenotypic, genetic and environmental overlap between BD and the brain change measures. As there is no evidence of shared environment (C) influencing BD (McGuffin et al., 2003), this factor was only estimated for the brain measures and not for BD. Disease status was dichotomous and assumed to represent an underlying continuous liability with a mean (SD) of 0 (1). Patients will have a higher values on the liability scale, thereby crossing a certain threshold (patient status=1). All other individuals will have lower liability scores and not cross the critical threshold (patient status=0; discordant co-twin of patient or control twin pairs). As we included approximately equal numbers of concordant, discordant, and healthy twin pairs, the critical threshold and heritability (the relative contribution of genetic variance to total variance) for the underlying liability for BD could not be estimated from this sample. Prevalence and heritability of BD were thus fixed to population values. Prevalence was set to 1% (Regeer et al., 2004) and heritability was set to 85% (McGuffin et al., 2003). Varying the prevalence and heritability of BD (e.g. 2% and 75% respectively) did not change the results. To apply the threshold model to the brain measures, the obtained standardized residuals of the subcortical brain measures were rendered into five ordinal categories identical for all subjects - thereby equating them across groups - and put in the model. Thresholding was based on normality plots, with the boundaries of the 'outer' two categories set at −1.5 SD and 1.5 SD respectively, with the other three categories falling in between, being 1 SD wide.

The phenotypic correlation (r_{ph}), an index of association between phenotypes (e.g. liability to develop BD and brain volume (change)), was based on calculations of within-twin/between-trait correlations. Heritability (h^2) and influence of shared and unique environment (c^2, e^2) as well as disentanglement of the observed correlation between liability for BD and subcortical measures into genetic and environmental components was based on polychoric cross-twin/within-trait and cross-twin/cross-trait correlations within MZ and DZ groups (Neale and Miller, 1997). The heritability of brain measures was determined within the bivariate model. A larger correlation between

Table 3 Mean uncorrected annual subcortical volume change for BD patients, co-twins of patients and healthy controls.

Region (volume change, in ml/year × 1000)	Mean (SD) volume change						Statistics							
	Patients[1]	p^2	Co-twins	p^2	Healthy controls[3]	p^2	Patients versus HC				Co-twins versus HC			
							F	*df*	*p*	Δ	*F*	*df*	*p*	Δ
Thalamus[4]	−29.34 (50.01)	**0.013**	−31.71 (46.50)	**0.005***	−37.72 (36.54)	**0.000***	0.57	1, *72*	0.455		0.73	1, *70*	0.394	
Caudate Nucleus[5]	−50.80 (33.72)	**0.000***	−38.86 (30.99)	**0.000***	−47.62 (29.79)	**0.000***	0.04	1, *70*	0.834		1.49	1, *71*	0.226	
Putamen[6]	−16.07 (21.85)	**0.002***	−12.85 (23.89)	**0.017**	−18.32 (24.33)	**0.000***	0.19	1, *72*	0.661		0.85	1, *70*	0.360	
Pallidum[7]	7.96 (15.85)	**0.029**	9.11 (12.63)	**0.003***	7.50 (11.41)	**0.000***	0.06	1, *73*	0.808		0.16	1, *72*	0.687	
Hippocampus[8]	43.46 (29.75)	**0.000***	42.81 (26.23)	**0.000***	41.15 (23.02)	**0.000***	0.23	1, *74*	0.634		0.05	1, *72*	0.827	
Amygdala[9]	4.64 (16.76)	0.128	3.12 (12.06)	0.232	5.33 (13.97)	**0.010**	0.02	1, *74*	0.881		0.46	1, *72*	0.501	
Nucleus Accumbens[10]	0.86 (7.63)	0.582	1.98 (9.30)	0.204	−2.58 (7.51)	**0.020**	3.16	1, *74*	0.079		**5.47**	**1,72**	**0.022**	**co > ctrl**
Total brain volume[11]	−1477.47 (2465.39)	**0.004***	−1595.34 (2022.33)	**0.002***	−1922.40 (1754.02)	**0.000***	1	1, *73*	0.321		0.60	1, *71*	0.443	
No outliers removed														
Thalamus	−19.19 (59.43)	0.130	−30.71 (64.44)	**0.011**	−40.66 (41.93)	**0.000***	2.83	1, *75*	0.097		1.36	1, *73*	0.248	

*Significant at α=0.007 (Bonferroni threshold), *p* values in bold face are significant at α=0.05.
[1]Including concordant pairs.
[10]*n*=25 patients, 23 co-twins and 51 healthy controls (1 subject removed).
[11]*n*=24 patients, 22 co-twins and 51 healthy controls (3 subjects removed).
[2]Within-group annual change for each region, corrected for age at baseline and gender. *P*-value indicates whether within-group mean volume change differs significantly from zero.
[3]Including both twins from complete pairs.
[4]*n*=23 patients, 21 co-twins, 51 healthy controls (5 subjects removed).
[5]*n*=21 patients, 22 co-twins and 51 healthy controls (6 subjects removed).
[6]*n*=23 patients, 21 co-twins and 51 healthy controls (5 subjects removed).
[7]*n*=23 patients, 22 co-twins and 52 healthy controls (3 subjects removed).
[8]*n*=25 patients, 23 co-twins and 51 healthy controls (1 subject removed).
[9]*n*=25 patients, 23 co-twins and 51 healthy controls (1 subject removed).

traits in MZ twins than in DZ twins suggests higher genetic contribution due to MZ twins being genetically identical whereas DZ twins only share on average 50% of their segregating genes. If there is no difference between MZ and DZ correlations then a larger influence of shared environmental factors is more likely (Boomsma et al., 2002). The genetic (r_g) and (shared and unique) environmental (r_e) correlations indicate the degree of overlap in genes or (shared or unique) environment influencing phenotypes. The phenotypic correlation can be written as the sum of the genetic correlation weighted by the square root of the heritabilities of the two traits (r_g*h_{BD}*h_{brain}) and the environmental correlations weighted by the square root of environmental variance associated with the two traits (r_c*c_{BD}*c_{brain}, r_e*e_{BD}*e_{brain}). These quantities are written as $r_{ph\text{-}g}$, $r_{ph\text{-}c}$ and $r_{ph\text{-}e}$ (Toulopoulou et al., 2007). The significance of variance components was tested by fitting different nested models to the data and by comparing their goodness of fit using Akaike's Information Criterion (AIC). A saturated model in which means, variances and correlations are estimated freely served as a baseline model to which more restrictive models were compared. Compared to the saturated model, the AE-model had the best fit in all ROIs and was therefore applied indiscriminately. Significance of parameter estimates and correlations was determined based on 95% confidence intervals (CI) (Neale and Miller, 1997).

2.3.5. Correction for multiple testing

In all tables and figures, significance after Bonferroni correction for multiple testing is indicated. The critical threshold for significance was calculated by dividing the alpha of 0.05 by the number of brain regions assessed, which was 7, yielding an significance threshold of $p=0.05/7=0.007$.

3. Results

3.1. Demographic and clinical characteristics

Please refer to Tables 1a and 1b for demographic information relevant to the baseline and change analyses, respectively. At baseline, DZ patient pairs were significantly older than MZ patient pairs ($p=0.016$) and MZ control pairs ($p=0.045$). Furthermore, at baseline, MZ control pairs had significantly more years of education compared to MZ patient pairs ($p=0.001$).

3.2. Baseline and change analysis of subcortical brain volumes in bipolar disorder

3.2.1. Univariate analysis of variance of baseline and change measures of subcortical volume

Tables 2 and 3 show the raw subcortical brain volumes and volume changes of BD patients, co-twins of patients and healthy controls.

At baseline, univariate analysis of variance revealed smaller volumes of the thalamus ($F[1,188]=12.56$, $p=0.000$), putamen ($F[1,186]=19.91$, $p=0.000$) and nucleus accumbens ($F[1,188]=7.38$, $p=0.007$) in BD patients as compared to healthy controls, when correction for lithium use was applied. Regarding subcortical volume change, BD patients, co-twins and healthy controls showed significant within-group changes in the majority of brain regions, with volume loss in the thalamus, caudate nucleus, putamen and nucleus accumbens (controls only), and volume increase in the pallidum, hippocampus and amygdala (controls only). Subcortical volume change was not significantly different between groups, except for a more pronounced increase in volume of the nucleus accumbens over time in co-twins compared to healthy controls ($F[1, 72]=5.47$, $p=0.022$).

3.2.2. Association between bipolar disorder and subcortical brain volumes at baseline

Supplementary tables S2 and S3, Table 4 and Figure 1 show the genetic model estimates for the baseline subcortical volumes, and phenotypic, genetic and unique environmental associations with BD. BD was phenotypically and genetically associated with smaller volumes of the thalamus ($r_{ph}=-0.20$, $r_g=-0.21$, $r_{ph\text{-}g}=-0.16$), putamen ($r_{ph}=-0.28$, $r_g=-0.29$, $r_{ph\text{-}g}=-0.24$) and nucleus accumbens ($r_{ph}=-0.22$, $r_g=-0.25$, $r_{ph\text{-}g}=-0.19$), when correction for lithium use was applied. This indicates that genes contributing to BD also contribute to smaller volumes of these regions. Furthermore, subcortical volumes at baseline showed high heritability in general (range h^2: 61% [Pallidum] to 85% [Hippocampus]).

3.2.3. Association between bipolar disorder and subcortical brain volume changes

Please refer to Supplementary tables S3 and S4 and Supplementary Figure S1 for the genetic model estimates for the subcortical volume change measures, and phenotypic, genetic and unique environmental associations with BD. There was a phenotypic and genetic association between BD and volume change of the nucleus accumbens ($r_{ph:}=0.2$, $r_g=1$, $r_{ph\text{-}g}=0.3$). This effect was in large part due to the difference in volume change between co-twins of patients and healthy controls, whereas patients did not differ significantly from healthy controls but did show a trend level increase in the nucleus accumbens ($p=0.079$, see Table 3).

In general, subcortical volume change was predominantly influenced by unique environment, as the heritability of subcortical brain change was very small (range h^2: 0% [hippocampus] to 11%[nucleus accumbens]), except for the caudate nucleus that showed moderate heritability (h^2:43%) when outliers were removed (with no outliers removed h^2 was 0%).

3.2.4. Association of clinical measures with baseline and change measures of subcortical volume in patients

In those regions that were phenotypically correlated to BD (i.e. thalamus, putamen and nucleus accumbens at baseline and change in the nucleus accumbens over time) we found no association with number of hospitalizations, lifetime experience of psychotic symptoms, YMRS score, IDS score (available for the first measurement only), HDRS score (available for the second measurement only) or Global Assessment of Functioning score (GAF, available for the second measurement only) in patients.

4. Discussion

In this longitudinal twin study investigating subcortical brain volume in BD, we concentrated specifically on the contribution of genes and environment to the association between BD and baseline and change measures of subcortical brain volumes.

Table 4 **a**. Genetic/environmental influences (AE-model, with 95% confidence intervals) on cross-sectional measures of subcortical brain volume, and phenotypic, genetic and environmental correlations with bipolar disorder, *uncorrected* for lithium use in patients.

Region (volume)	h^2 %	e^2	r_{ph}	r_g	r_e	$r_{ph\text{-}g}$	$r_{ph\text{-}e}$
Thalamus	60 (39 to 75)	40 (25 to 61)	0.02 (−0.10 to 0.15)	−0.07 (−0.29 to 0.13)	0.31 (−0.11 to 0.66)	−0.05 (−0.20 to 0.09)	0.08 (−0.03 to 0.17)
Caudate Nucleus	84 (70 to 91)	16 (9 to 30)	−0.08 (−0.21 to 0.06)	−0.05 (−0.22 to 0.13)	−0.25 (−0.66 to 0.22)	−0.04 (−0.19 to 0.11)	−0.04 (−0.11 to 0.03)
Putamen	74 (58 to 84)	26 (16 to 42)	−0.12 (−0.25 to 0.01)	**−0.20** (−0.37 to −0.02)	0.17 (−0.20 to 0.52)	**−0.16** (−0.29 to −0.01)	0.03 (−0.04 to 0.11)
Pallidum	64 (43 to 78)	36 (22 to 57)	−0.11 (−0.24 to 0.02)	−0.12 (−0.31 to 0.08)	−0.10 (−0.48 to 0.29)	−0.09 (−0.23 to 0.06)	−0.02 (−0.12 to 0.07)
Hippocampus	84 (71 to 92)	16 (8 to 29)	−0.08 (−0.21 to 0.05)	−0.13 (−0.30 to 0.04)	0.17 (−0.29 to 0.58)	−0.11 (−0.25 to 0.03)	0.03 (−0.05 to 0.10)
Amygdala	69 (52 to 81)	31 (19 to 48)	0 (−0.13 to 0.13)	−0.03 (−0.22 to 0.16)	0.08 (−0.33 to 0.46)	−0.02 (−0.17 to 0.13)	0.02 (−0.07 to 0.10)
Nucleus Accumbens	64 (45 to 77)	36 (23 to 55)	−0.12 (−0.25 to 0.01)	−0.18 (−0.37 to 0.02)	0.05 (−0.34 to 0.45)	−0.13 (−0.27 to 0.01)	0.01 (−0.08 to 0.10)

b. Genetic/environmental influences (AE-model, with 95% confidence intervals) on cross-sectional measures of subcortical brain volume, and phenotypic, genetic and environmental correlations with bipolar disorder, *corrected* for lithium use in patients.

Region (volume)	h^2 %	e^2	r_{ph}	r_g	r_e	$r_{ph\text{-}g}$	$r_{ph\text{-}e}$
Thalamus	64 (45 to 78)	36 (22 to 55)	**−0.20*** (−0.32 to −0.07)	**−0.21** (−0.41 to −0.02)	−0.20 (−0.59 to 0.23)	**−0.16** (−0.30 to −0.01)	−0.05 (−0.14 to 0.05)
Caudate Nucleus	83 (69 to 91)	17 (9 to 31)	−0.13 (−0.27 to 0)	−0.07 (−0.25 to 0.10)	−0.45 (−0.84 to 0.02)	−0.06 (−0.21 to 0.09)	−0.07 (−0.14 to 0)
Putamen	80 (67 to 89)	20 (11 to 33)	**−0.28*** (−0.40 to −0.15)	**−0.29*** (−0.45 to −0.12)	−0.23 (−0.57 to 0.16)	**−0.24*** (−0.37 to −0.10)	−0.04 (−0.11 to 0.03)
Pallidum	61 (40 to 76)	39 (24 to 60)	−0.02 (−0.15 to 0.11)	−0.04 (−0.24 to 0.16)	0.04 (−0.33 to 0.41)	−0.03 (−0.17 to 0.11)	0.01 (−0.08 to 0.10)
Hippocampus	85 (72 to 93)	15 (7 to 28)	−0.11 (−0.24 to 0.02)	−0.14 (−0.31 to 0.03)	0.07 (−0.39 to 0.51)	−0.12 (−0.26 to 0.02)	0.01 (−0.06 to 0.08)
Amygdala	71 (55 to 82)	29 (18 to 45)	−0.11 (−0.23 to 0.03)	−0.10 (−0.29 to 0.09)	−0.13 (−0.52 to 0.28)	−0.08 (−0.22 to 0.07)	−0.03 (−0.11 to 0.06)
Nucleus Accumbens	66 (47 to 79)	34 (21 to 53)	**−0.22*** (−0.34 to −0.08)	**−0.25** (−0.45 to −0.06)	−0.11 (−0.50 to 0.31)	**−0.19** (−0.33 to −0.04)	−0.03 (−0.12 to 0.07)

*Significant at α=0.007 (Bonferroni threshold), estimates in bold face are significant at α=0.05.

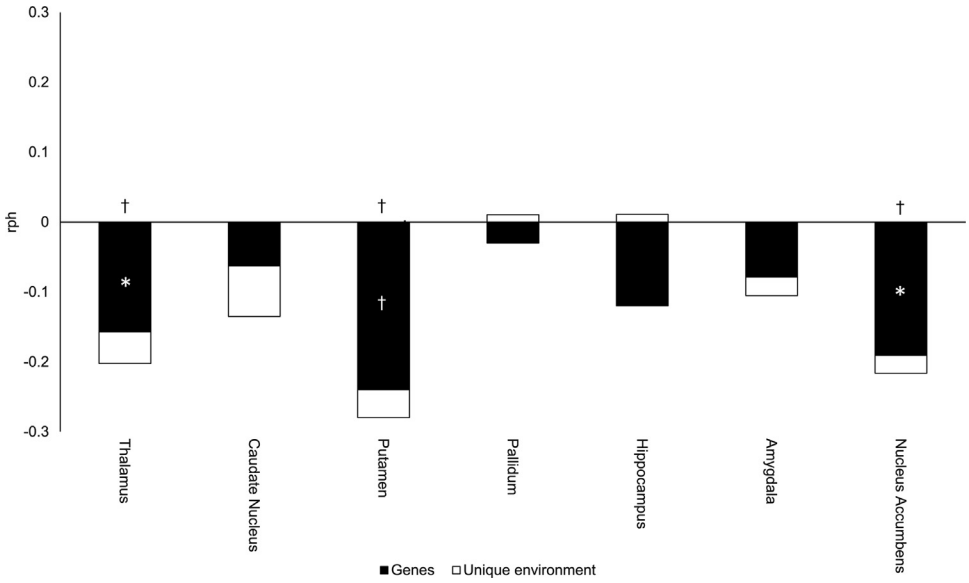


Figure 1 Genetic and unique environmental contributions to the phenotypic correlation between BD and subcortical brain volumes. Level of significance is indicated for the total correlation (outer symbols), and for genetic and unique environmental contributions (inner symbols) separately. †Significant at p= 0.007, * Significant at p= 0.05.

The main finding is that, at baseline, BD was associated with smaller volumes of the thalamus, putamen and nucleus accumbens when we corrected for lithium use. This finding stresses the importance of taking into the account the influence of lithium on the brain as its use may mask brain abnormalities associated with BD. Volumes of these regions were strongly influenced by genes associated with BD. Our findings in twins confirm results from a previous study with BD patients and their first-degree relatives, that found an association between the genetic liability to BD and smaller volume of the ventral striatum, including the anterior putamen (McDonald et al., 2004). Similarly, less grey matter density has been demonstrated in the anterior thalamus in BD patients and unaffected relatives, which the authors interpreted to reflect an association with genetic liability to psychosis in general, as schizophrenia patients and relatives showed the same thalamic grey matter deficit (McIntosh et al., 2004). However, findings across cross-sectional case-control studies are inconsistent. For example, Hibar et al. (2013) and Rimol et al. (2010) did show smaller volume of the thalamus but most studies assessing thalamic volume do not report abnormalities in this structure in BD (Emsell and McDonald, 2009; Hallahan et al., 2011; Womer et al., 2014). Here, the difficulty of isolating thalamic nuclei with current available methods may contribute to the inconsistency of findings (Blond et al., 2012). However, differences across studies could possibly also be attributed to the presence of different disease mechanisms or the inclusion of heterogeneous clinical samples in studies where correction for medication use may or may not have been applied (Emsell and McDonald, 2009; Savitz and Drevets, 2009; Hajek et al., 2012). Furthermore, smaller volumes of the left (Almeida et al., 2009) and right (Haller et al., 2011) putamen have been found previously, as have larger total (DelBello et al., 2004) and right side (Hallahan et al., 2011) volumes in this region, as well as no differences between BD patients and controls in putamen volume (Womer et al., 2014). Our finding of smaller volume of the nucleus accumbens has also been demonstrated by others (Dickstein et al., 2005; Rimol et al., 2010; Haller et al., 2011) but so has tissue preservation in this structure (Womer et al., 2014).

As the thalamus and striatum are particularly important nodes in the cortico-striato-thalamic loops involved in emotion regulatory processes (Emsell and McDonald, 2009; Marchand and Yurgelun-Todd, 2010; Blond et al., 2012; Strakowski et al., 2012), structural deficiency in these and associated regions could underlie BD (Emsell and McDonald, 2009; Blond et al., 2012), which necessitates closer examination. Moreover, based on this and several other studies, abnormal morphology in emotion processing areas may share some genetic variance with BD (Noga et al., 2001; McDonald et al., 2004; McIntosh et al., 2004), although McIntosh et al. (2006) noted an absence of a genetic association between BD and measures of the amygdala-hippocampal complex or thalamus. Therefore, future studies could benefit from evaluating to what extent volume deficiencies in subcortical brain regions are genetically and environmentally mediated, in order to determine by which mechanisms affective dysregulation develops. Here, genome-wide association studies (GWAS) could assist in identifying gene pools associated with subcortical structural abnormalities and BD. For example, specific genetic variants influencing putamen volume were identified recently (Hibar et al., 2015), although these variants do not appear to confer risk for BD (Mühleisen et al., 2014).

In our study, BD was not associated with subcortical volume change over time in any of the examined regions (including those that showed an association with BD at baseline), except for the nucleus accumbens where a significant association was found. However, this association was no longer significant after Bonferroni correction. These findings suggest BD may not be a progressive brain disease. However, due to the relatively small sample size of the

group that was assessed longitudinally, this finding should be interpreted with caution. Perhaps with a larger sample a Bonferroni significant increase in volume of the nucleus accumbens would have been found. Previous longitudinal case-control studies showed volumetric preservation, increases and decreases over time in the hippocampus (see review by Lim et al. (2013), as well as thalamic and caudate nucleus increases over time in BD (Lisy et al., 2011). Furthermore, it has been suggested that volumetric abnormalities in the amygdala arise during neurodevelopment but remain stable during adulthood in BD (Lim et al., 2013). However, based on our results, we cannot fully support this hypothesis as we found no difference in amygdala volume between BD patients and healthy controls at baseline.

At baseline, subcortical brain volumes showed high heritability, which is in line with studies by Bohlken et al. (2014) and den Braber et al. (2013) that reported similar heritability estimates in the same brain regions. In contrast, heritability of subcortical brain volume change was generally low. In part this may reflect measurement error inherent to low statistical power, using a relatively small sample for the change analyses, but could to some extent suggest subcortical brain change to be particularly influenced by factors unique to the individual rather than genes.

A number of limitations are relevant to our study. First, two different MRI scanners were used for the two measurements, therefore we cannot completely rule out the presence of scanner effects on our measures. However, we ensured scanner field strength (1.5 T), imaging parameters and (pre)processing algorithms to have been equal for all subjects across measurements. In addition, all baseline scans were obtained on one scanner while all follow-up scans were obtained on the other scanner. So, within-twin effects are controlled for. Second, females were overrepresented in patient and control pairs. Therefore, subcortical brain measures were corrected for gender. Third, using a relatively small sample size, particularly in the change analyses, reduces the statistical power to estimate genetic and environmental sources of variance in univariate and bivariate designs (Visscher, 2004; Posthuma and Boomsma, 2000). However, this is particularly relevant when heritability is expected to be low, which was not the case for our baseline measures. Moreover, in a few regions the associations between BD and subcortical volume were highly significant, surviving Bonferroni correction. Therefore, we are confident to have had sufficient power for baseline analysis of subcortical volume in BD. Fourth, although we corrected for lithium use in the baseline measures, we did not for the change measures, due to the small sample sizes of the respective lithium using groups.

In summary, BD was associated with smaller volumes of the thalamus, putamen and nucleus accumbens at baseline, with genes contributing to the disease influencing the volumes of these regions. In contrast, BD was not associated with subcortical volume change over time after Bonferroni correction, indicating no differences in change between patients, co-twins and healthy controls. Further evaluation of genetic and environmental contributions to structural brain abnormalities in BD is recommended, in particular regarding subcortical volumes assumed to be involved in emotion processing, to ascertain by which mechanisms affective dysregulation develops.

Role of funding source

Funding for this study was provided by the Netherlands Organization for Scientific Research (NWO), Grant 9120818. The NWO had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Contributors

FB contributed to the design of the study, the acquisition and preprocessing of data, performed the analyses, interpreted the data, drafted the manuscript, and made the tables and figures.
RB contributed to data analysis and interpretation of results.
SK contributed to the acquisition of data.
HS contributed to the interpretation of results.
AvdS contributed to the design of the study and the acquisition of data.
RV contributed to the design of the study and the acquisition of data.
MH supervised clinical assessment of study subjects.
DB contributed to the acquisition of data.
HHP contributed to the design of the study and supervision of the project.
WN contributed to the design of the study and supervision of the project.
RK contributed to the design of the study and supervision of the project.
NvH contributed to the design of the study, interpretation of the data, and supervised the project.
All authors critically reviewed the manuscript.

Conflict of interest

All authors declare that they have no conflicts of interest.

Acknowledgments

We thank the Netherlands Organization for Scientific Research (NWO) for supporting this study (Grant 9120818). We thank Raissa Derckx for her assistance with the quality checking of MRI data.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.euroneuro.2015.09.023.

References

Aldhafeeri, F.M., Mackenzie, I., Kay, T., Alghamdi, J., Sluming, V., 2012. Regional brain responses to pleasant and unpleasant IAPS pictures: different networks. Neurosci. Lett. 512, 94–98.
Almeida, J.R., Akkal, D., Hassel, S., Travis, M.J., Banihashemi, L., Kerr, N., Kupfer, D.J., Phillips, M.L., 2009. Reduced gray matter volume in ventral prefrontal cortex but not amygdala in bipolar disorder: significant effects of gender and trait anxiety. Psychiatry Res. 171, 54–68.
Baaré, W.F., Hulshoff Pol, H.E., Boomsma, D.I., Posthuma, D., de Geus, E.J., Schnack, H.G., van Haren, N.E., van Oel, C.J., Kahn, R.S., 2001. Quantitative genetic modeling of variation in human brain morphology. Cereb. Cortex 11, 816–824.

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression. Arch. Gen. Psychiatry 4, 561–571.

Blond, B.N., Fredericks, C.A., Blumberg, H.P., 2012. Functional neuroanatomy of bipolar disorder: structure, function, and connectivity in an amygdala-anterior paralimbic neural system. Bipolar Disord., 14.

Bohlken, M.M., Brouwer, R.M., Mandl, R.C.W., van Haren, N.E.M., Brans, R.G.H., van Baal, G.C.M., de Geus, E.J.C., Boomsma, D. I., Kahn, R.S., Hulshoff Pol, H.E., 2014. Genes contributing to subcortical volumes and intellectual ability implicate the thalamus. Hum. Brain Mapp. 35, 2632–2642.

Boomsma, D.I., 1998. Twin registers in Europe: an overview. Twin Res. 1, 34–51.

Boomsma, D., Busjahn, A., Peltonen, L., 2002. Classical twin studies and beyond. Nat. Rev. Genet. 3, 872–882.

Brans, R.G., van Haren, N.E., van Baal, G.C., Schnack, H.G., Kahn, R.S., Hulshoff Pol, H.E., 2008. Heritability of changes in brain volume over time in twin pairs discordant for schizophrenia. Arch. Gen. Psychiatry 65, 1259–1268.

Brans, R.G., Kahn, R.S., Schnack, H.G., van Baal, G.C., Posthuma, D., van Haren, N.E., Lepage, C., Lerch, J.P., Collins, D.L., Evans, A.C., Boomsma, D.I., Hulshoff Pol, H.E., 2010. Brain plasticity and intellectual ability are influenced by shared genes. J. Neurosci. 30, 5519–5524.

den Braber, A., Bohlken, M.M., Brouwer, R.M., van't Ent, D., Kanai, R., Kahn, R.S., de Geus, E.J.C., Hulshoff Pol, H.E., Boomsma, D.I., 2013. Heritability of subcortical brain measures: a perspective for future genome-wide association studies. Neuroimage 83, 98–102.

DelBello, M.P., Zimmerman, M.E., Mills, N.P., Getz, G.E., Strakowski, S.M., 2004. Magnetic resonance imaging analysis of amygdala and other subcortical brain regions in adolescents with bipolar disorder. Bipolar Disord. 6, 43–52.

Dickstein, D.P., Milham, M.P., Nugent, A.C., Drevets, W.C., Charney, D.S., Pine, D.S., Leibenluft, E., 2005. Frontotemporal alterations in pediatric bipolar disorder: results of a voxel-based morphometry study. Arch. Gen. Psychiatry 62, 734–741.

Emsell, L., McDonald, C., 2009. The structural neuroimaging of bipolar disorder. Int. Rev. Psychiatry 21, 297–313.

First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W., Benjamin, L.S., 1997. Structured Clinical Interview for DSM-IV Axis II Personality Disorders, (SCID-II). American Psychiatric Press, Washington, DC.

First, M., Spitzer, R., Gibbon, M., Williams, J., 1997. User's guide for the Structured Clinical Interview for DSM-IV Axis I Disorders - Clinical Version (SCID-CV). American Psychiatric Press, Washington, DC.

Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22.

Goodwin, F.K., Jamison, K.R., Ghaemi, S.N., 2007. Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression, 2nd ed. Oxford University Press, New York.

Hafeman, D.M., Chang, K.D., Garrett, A.S., Sanders, E.M., Phillips, M.L., 2012. Effects of medication on neuroimaging findings in bipolar disorder: an updated review. Bipolar Disord. 14, 375–410.

Hajek, T., Kopecek, M., Höschl, C., Alda, M., 2012. Smaller hippocampal volume in patients with bipolar disorder are masked by exposure to lithium: a meta-analysis. J. Psychiatry Neurosci. 37, 333–343.

Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56–62.

Hallahan, B., Newell, J., Soares, J.C., Brambilla, P., Strakowski, S. M., Fleck, D.E., Kieseppä, T., Altshuler, L.L., Fornito, A., Malhi, G.S., McIntosh, A.M., Yurgelun-Todd, D.A., Labar, K.S., Sharma, V., MacQueen, G.M., Murray, R.M., McDonald, C., 2011. Structural magnetic resonance imaging in bipolar disorder: an international collaborative meaga-analysis of individual adult patient data. Biol. Psychiatry 69, 326–335.

Haller, S., Xekardaki, A., Delaloye, C., Canuto, A., Lövblad, K.O., Gold, G., Giannakopoulos, P., 2011. Combined analysis of grey matter voxel-based morphometry and white matter tract-based spatial statistics in late-life bipolar disorder. J. Psychiatry Neurosci. 36, 391–401.

Hibar, D.P., van Erp, T.G.M., Turner, J.A., Haukvik, U.K., Agartz, I., et al., 2013. Meta-analysis of structural brain differences in bipolar disorder: the ENIGMA-Bipolar Disorder Project. OHBM, Seattle, WA.

Hibar, D.P., Stein, J.L., Renteria, M.E., Arias-Vasquez, A., Desrivières, S., Jahanshad, N., Toro, R., Wittfeld, K., et al., 2015. Common genetic variants influence human subcortical brain structures. Nature 520, 224–229.

Kenny, S., Andric, M., Boker, S.M., Neale, M.C., Wilde, M., Small, S. L., 2009. Parallel workflows for data-driven structural equation modeling in functional neuroimaging. Front. Neuroinform. 3, 34.

Lim, C.S., Baldessarini, R.J., Vieta, E., Yucel, M., Bora, E., Sim, K., 2013. Longitudinal neuroimaging and neuropsychological changes in bipolar disorder patients: review of the evidence. Neurosci. Biobehav. Rev. 37, 418–435.

Lisy, M.E., Jarvis, K.B., Delbello, M.P., Mills, N.P., Weber, W.A., Fleck, D., Strakowski, S.M., Adler, C.M., 2011. Progressive neurostructural changes in adolescent and adult patients with bipolar disorder. Bipolar Disord. 13, 396–405.

Marchand, W.R., Yurgelun-Todd, D., 2010. Striatal structure and function in mood disorders: a comprehensive review. Bipolar Disord. 12, 764–785.

McDonald, C., Bullmore, E.T., Sham, P.C., Chitnis, X., Wickham, H., Bramon, E., Murray, R.M., 2004. Association of genetic risks for schizophrenia and bipolar disorder with specific and generic brain structural endophenotypes. Arch. Gen. Psychiatry 61, 974–984.

McGuffin, P., Rijsdijk, F., Andrew, M., Sham, P., Katz, R., Cardno, A., 2003. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. Arch. Gen. Psychiatry 60, 497–502.

McIntosh, A.M., Job, D.E., Moorhead, W.J., Harrison, L.K., Forrester, K., Lawrie, S.M., Johnstone, E.C., 2004. Voxel-based morphometry of patients with schizophrenia or bipolar disorder and their unaffected relatives. Biol. Psychiatry 56, 544–552.

McIntosh, A.M., Job, D.E., Moorhead, W.J., Harrison, L.K., Whalley, H.C., Johnstone, E.C., Lawrie, S.M., 2006. Genetic liability to schizophrenia or bipolar disorder and its relationship to brain structure. Am. J. Med. Genet. Part. B (Neuropsychiatric Genetic.) 141B, 76–83.

Mühleisen, T.W., Leber, M., Schulze, T.G., Strohmaier, J., Degenhardt, F., Treutlein, J., et al., 2014. Genome-wide association study reveals two new risk loci for bipolar disorder. Nat. Commun. 5, 3339. http://dx.doi.org/10.1038/ncomms4339.

Neale, M.C., Miller, M.B., 1997. The use of likelihood-based confidence intervals in genetic models. Behav. Genet. 27, 113–120.

Noga, J.T., Vladar, K., Torrey, E.F., 2001. A volumetric magnetic resonance imaging study of monozygotic twins discordant for bipolar disorder. Psychiatry Res. 106, 25–34.

Nurnberger jr., J.I., Blehar, M.C., Kaufmann, C.A., York-Cooler, C., Simpson, S.G., Harkavy-Friedman, J., Severe, J.B., Malaspina, D., Reich, T., 1994. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genet. Initiative Arch. Gen. Psychiatry 51, 849–859.

Pfohl, B., Blum, N., Zimmerman, M., 1997. Structured Interview for DSM-IV Personality SIDP-IV. American Psychiatric Press, Washington DC.

Phillips, M.L., Drevets, W.C., Rauch, S.L., Lane, R., 2003. Neurobiology of emotion perception I: the neural basis of normal emotion perception. Biol. Psychiatry 54, 504–514.

Phillips, M.L., Ladouceur, C.D., Drevets, W.C., 2008. A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. Mol. Psychiatry 13 (829), 829–833 857..

Phillips, M.L., Swartz, H.A., 2014. A critical appraisal of neuroimaging studies of bipolar disorder: toward a new conceptualization of underlying neural circuitry and a road map for future research. Am. J. Psychiatry 171, 829–843.

Posthuma, D., Boomsma, D.I., 2000. A note on the statistical power in extended twin designs. Behav. Genet. 30, 147–158.

R Development Core Team, 2008. R: a language and environment for statistical computing. Available from: ⟨http://www.r-project.org⟩.

Regeer, E.J., ten Have, M., Rosso, M.L., Hakkaart-van Roijen, R.L., Vollebergh, W., Nolen, W.A., 2004. Prevalence of bipolar disorder in the general population: a Reappraisal Study of the Netherlands Mental Health Survey and Incidence Study. Acta Psychiatr. Scand. 110, 374–382.

Rimol, L.M., Hartberg, C.B., Nesvag, R., Fennema-Notestine, C., Hagler jr., D.J., Pung, C.J., Jennings, R.G., Haukvik, U.K., Lange, E., Nakstad, P.H., Melle, I., Andreassen, O.A., Dale, A.M., Agartz, I., 2010. Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder. Biol. Psychiatry, 68; , 41–50.

Savitz, J., Drevets, W.C., 2009. Bipolar and major depressive disorder: neuroimaging the developmental-degenerative divide. Neurosci. Biobehav. Rev. 33, 699–771.

Strakowski, S.M., Adler, C.M., Almeida, J., Altshuler, L.L., Blumberg, H.P., Chang, K.D., DelBello, M.P., Frangou, S., McIntosh, A., Phillips, M.L., Sussman, J.E., Townsend, J.D., 2012. The functional neuroanatomy of bipolar disorder: a consensus model. Bipolar Disord. 14, 313–325.

Toulopoulou, T., Picchioni, M., Rijsdijk, F., Hua-Hall, M., Ettinger, U., Sham, P., Murray, R., 2007. Substantial genetic overlap between neurocognition and schizophrenia: genetic modeling in twin samples. Arch. Gen. Psychiatry 64, 1348–1355.

van der Schot, A.C., Vonk, R., Brans, R.G., van Haren, N.E.M., Koolschijn, P.C., Nuboer, V., Schnack, H.G., van Baal, G.C., Boomsma, D.I., Nolen, W.A., Hulshoff Pol, H.E., Kahn, R.S., 2009. Influence of genes and environment on brain volumes in twin pairs concordant and discordant for bipolar disorder. Arch. Gen. Psychiatry 66, 142–151.

Visscher, P.M., 2004. Power of the classical twin design revisited. Twin Res. 7, 505–512.

Womer, F.Y., Wang, L., Alpert, K.I., Smith, M.J., Csernansky, J.G., Barch, D.M., Mamah, D., 2014. Basal ganglia and thalamic morphology in schizophrenia and bipolar disorder. Psychiatry Res.: Neuroimaging 223, 75–83.

Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. Br. J. Psychiatry 133, 429–435.