NeuroImage 70 (2013) 175–188

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

When the single matters more than the group: Very high false positive rates in single case Voxel Based Morphometry

C. Scarpazza [a,b,*], G. Sartori [a], M.S. De Simone [a], A. Mechelli [b]

[a] Department of Psychology, University of Padua, Via Venezia 12, 35131 Padova, Italy
[b] Department of Psychosis Studies, Institute of Psychiatry, King's College London, De Crespigny Park, London, SE5 8AF, UK

ARTICLE INFO

Article history:
Accepted 21 December 2012
Available online 2 January 2013

Keywords:
Neuroimaging
Voxel Based Morphometry
Psychiatric disease
Neurological disease
Single case study

ABSTRACT

Voxel Based Morphometry (VBM) studies typically involve a comparison between groups of individuals; this approach however does not allow inferences to be made at the level of the individual. In recent years, an increasing number of research groups have attempted to overcome this issue by performing single case studies, which involve the comparison between a single subject and a control group. However, the interpretation of the results is problematic; for instance, any significant difference might be driven by individual variability in neuroanatomy rather than the neuropathology of the disease under investigation, or might represent a false positive due to the data being sampled from non-normally distributed populations. The aim of the present investigation was to empirically estimate the likelihood of detecting significant differences in gray matter volume in individuals free from neurological or psychiatric diagnosis. We compared a total of 200 single subjects against a group of 16 controls matched for age and gender, using two independent datasets from the Neuroimaging Informatics Tools and Resources Clearinghouse. We report that the chance of detecting a significant difference in a disease-free individual is much higher than previously expected; for instance, using a standard voxel-wise threshold of $p<0.05$ (corrected) and an extent threshold of 10 voxels, the likelihood of a single subject showing at least one significant difference is as high as 93.5% for increases and 71% for decreases. We also report that the chance of detecting significant differences was greatest in frontal and temporal cortices and lowest in subcortical regions. The chance of detecting significant differences was inversely related to the degree of smoothing applied to the data, and was higher for unmodulated than modulated data. These results were replicated in the two independent datasets. By providing an empirical estimation of the number of significant increases and decreases to be expected in each cortical and subcortical region in disease-free individuals, the present investigation could inform the interpretation of future single case VBM studies.

© 2012 Elsevier Inc. All rights reserved.

Introduction

The neuroanatomical basis of most neurological diseases is relatively well-understood and an expert neurologist can detect specific abnormalities in the brain scans of patients by visual inspection. However this is not true for psychiatric illnesses, which for a long time were considered "functional" disorders without a reliable neuroanatomical basis. Although the patients undergoing post mortem examination had suffered life-long severe mental illness, it was rare to detect any macroscopic neuropathology (Dolan, 2008). The development of structural neuroimaging techniques has provided substantial evidence that psychiatric diseases are associated with abnormalities in brain structure, and has brought about significant breakthroughs in our understanding of the neurobiology of such illnesses (Bora et al., 2009; Butler et al., 2011; Martin et al., 2010; Oquendo and Parsey, 2007; Rauch, 2000; Shirtcliff et al., 2009; Soares, 2003; Takahashi et al., 2010; Wingenfeld et al., 2010). The literature on brain volume abnormalities in psychiatric disorders is rapidly expanding, with thousands of studies published to date. These studies have revealed, for example, gray matter (GM) reductions in the prefrontal cortex, superior temporal gyrus, thalamus and amygdala in schizophrenia (Bora et al., 2011); GM reductions in the anterior cingulate and insula in bipolar disorder (Bora et al., 2010; Ellison-Wright and Bullmore, 2010); frontopolar, orbitofrontal, insular and superior temporal GM reductions in psychopathy (De Olivera-Souz et al., 2008); GM reductions of the anterior cingulate cortex (ACC), middle and inferior frontal gyrus, hippocampus and thalamus in depression (Du et al., 2012); and reduced GM volume in the hippocampus, parahippocampal gyrus, ACC, bilateral insula and calcarine cortices in posttraumatic stress disorder (Chen et al., 2006; Felmingham et al., 2009; Zhang et al., 2011). Alterations have also been found in studies of personality disorders, including GM volume loss in the amygdala in borderline disorder (Rusch et al., 2003; Soloff et al., 2008); larger GM volume in the posterior cingulate cortex and precuneus in schizotypy (Modinos et al., 2010); and white

* Corresponding author at: Department of Psychology, University of Padua, Via Venezia 12, 35131 Padova, Italy.
E-mail address: cristina.scarpazza@gmail.com (C. Scarpazza).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.12.045

Table 1
Single case VBM studies on patients with neurological and psychiatric diseases.

	Subject	Disease	Controls	p-value	Correction	GM increase or decrease	Areas
Migliaccio et al., 2012	7 subj (6♀)	PCA	29	0.05	FWE	Decrease	Bilateral ventral occipital and temporal (3/7 patients); R supramarginal gyrus (2/7 patients); bilateral inferior parietal lobule (7/7 patients)
Beeson et al., 2011	67 ♂	PPA	35	0.01	FDR	Decrease	Bilateral posterior perisylvian regions; bilateral mild lateral temporal lobes
Klingner et al., 2012	23 ♂	Bell palsy	1	0.05	–	Decrease	L M1, bilateral SMA, L insula, bilateral cerebellum
Migliaccio et al., 2011	58 ♀	PCA	15	0.05	FWE	Decrease	Bilateral superior occipital gyrus, cuneus, bilateral Inferior occipital gyrus. Bilateral lingual gyrus, bilateral superior parietal lobule, Bilateral thalamus
Sehm et al., 2011	24 ♂	Focal retrograde amnesia	20	0.01	FWE	Decrease	L temporo polar cortex, R Lingual cortex
Valdes-Sosa et al., 2011	73 ♂	Prosopo-agnosia	10	0.01	FDR	Decrease	L ventral and occipital temporal cortex, L fusiform gyrus
Bianchini et al., 2010	22 ♂	Topographical disorientation	12	–	–	–	–
Eriguchi et al., 2010	31 ♂	Citrullinemia epilepsy	111	0.05	FDR	decrease	L hippocampus
Freudenmann et al., 2010	72 ♀	Vascular encephalopathy	7	0.05	FWE	Decrease increase	R postcentral, R IPL, Bilaterial putamen, L cingulated cortex
Maguire et al., 2010	70 ♂	SD	10	–	–	decrease	L striatum, LSTG, Bilateral hippocampus, L IFG,R temporal pole, R cerebellum, R ITG
Nanri et al., 2011	84 ♀	Autoimmune ataxia + Basedow disease	–	–	–	decrease	R Cerebellar cortex
Riddoch et al., 2010	74 ♀ 70 ♂ 65 ♂	Stroke	140	–	–	decrease	R IPL, R IFG, R angular, R supramarginal gyrus, R parietooccipital regions, L parietotemporal regions, R IFG, R STG, R supramarginal, R angular gyrus
Rigoni et al., 2010	24 ♀	Psychopathy	6	0.05 0.001	FWE –	decrease	L MFG,L SFG, L Lateral temporal cortex, L superior occipital cortex, L SFG, L MFG
Muhlau et al., 2009	22 subj (12 ♀)	HD	133 for each subj	0.05	FDR	decrease	Bilateral head of caudate, R insula, R temporal pole
Narvid et al., 2009	66 ♂	FTD + HME	28	0.001	–	decrease	R ventral and medial frontal cortex, R insular cortex
Tramoni et al., 2009	34 ♂	Functional amnesia	25 subj (10 ♀)	0.05	FDR	–	–
Bozzali et al., 2008	77 ♀	CBD	8	0.05 0.001	FWE –	decrease	Bilateral Pre- and post-central gyrus, R middle frontal gyrus, L middle frontal gyrus, R SFG, R IFG, R puntamen, R lenticular nucleus, R parietal lobe
Epelbaum et al., 2008	46 ♂	Pure alexia, epilepsy	17	0.001	–	decrease	Bilateral frontal and ACC, L parietal gyrus
Feldmann et al., 2008	54 ♂	PCA	20	0.05	FWE	decrease	R occipital mesial and inferior, R fusiform, R ACC, R SFG R precuneus, L occipital inferior gyrus
Adlam et al., 2006	6 subj	PPA	12	0.05	FDR	decrease	L ventral temporal lobe (3/6 subjects), L dorso-ventral temporal lobe (2/6 subjects), R temporal pole (1/6 subjects), R rostro-ventral temporal lobe (1/6 subjects)
Brazdil et al., 2006	25 ♀	Epilepsy	–	–	–	increase	Anterior rim of the L central sulcus
Cipolotti et al., 2006	74 ♂	Hippocampal amnesia	15	0.05 0.001	FWE -	decrease	- Bilateral Head of body of hippocampus and enthorinal cortex

Colliot et al., 2006	27 subj (16 ♀)	FCD	39	–	–	Increase Increase Decrease	regions with FCD lesions (21/27 patients), frontal and temporal extralesional regions (16/27 patients), frontal and temporal extralesional regions (8/27 patients)
Joubert et al., 2006	61 ♀ 59 ♀ 73 ♂	Right FTD	28	0.01	FWE	Decrease	R inferior temporal gyrus, R hippocampus, R middle temporal gyrus, R caudate, L STG
Suzuki et al., 2005	43 ♂ 32 ♂ 38 ♂ 44 ♀ 36 ♀	Schizophrenia	20 (10 ♀)	0.05	FWE	Decrease	L IFG and L frontal pole (4 patients only)
Zahn et al., 2005	10 subj (6♀) 5 subj (1♀)	AD PPA	10 (5♀)	0.001	–	Decrease	Bilateral anterior lateral temporal (8/10 AD and 5/5 PPA); L medial temporal (9 AD and 1 PPA); R posterior parietal (9 AD and 5 PPA); R posterior cingulate and R precuneus (9 AD and 0 PPA).
Gorno-Tempini et al., 2004a	56 ♀	PPA + CBD	16	0.05 0.001	FWE –	Decrease	L IFG, L anterior insula, L SMA, Bilateral caudate
Gorno-Tempini et al., 2004b	67 ♀	Right SD	35	0.05 0.001	FWE –	Decrease	R anterior temporal lobe, R amygdala/anterior hippocampus, R collateral sulcus, R fusiform gyrus, R posterior insula/STG L amygdala/anterior hippocampus L collateral sulcus/fusiform gyrus
Thompson et al., 2004	63 ♂ 64 ♂	SD FTD	17	0.001	–	Decrease	L infero lateral temporal lobe, L ACC,L insula, R temporal pole
Salmond et al., 2003	14subj (1♀)	Autism	18 (12♀)	0.001	–	Decrease	Bilateral Hippocampus (in 7 subjects), amygdala, OFC (in 13 subjects), STG, cerebellum (in 11 subjects)
Rosen et al., 2002	54 ♀ 81 ♂ 65 ♂	PPA AD SD	20 for each subj	0.001	–	Decrease	L DLPFC, L IFG, Bilateral temporo parietal cortex, bilateral IPL, L anterior temporal
Gitelman et al., 2001	33 ♀ 38 ♂ 41♂ 56 ♀ 71 ♀	Herpes sinplex	10 for each subj	0.05	FWE	Decrease	Bilateral amygdala, hippocampus, enthorinal cortex, insula, gyrus rectus, nucleus accumbens
Mummery et al., 2000	6 subj (5♀)	SD	14	0.05	FWE	Decrease	Bilateral Temporal pole, L Middle temporal gyrus, Bilateral Amygdala, Bilateral Fusiform gyrus
Woermann et al., 1999	20 subj (12♀)	JME	30 subj (16♀)	0.05	FWE	Increase decrease	Bilateral temporo posterior lobe (2/20 subjects); bilateral frontopolar areas (3/20 subjects)

♂ = male, ♀ = female. L = left, R = right. PPA: primary progressive aphasia; PCA: posterior cortical atrophy; SD: semantic dementia, HD: Huntington disease, FTD: fronto temporal dementia; bvFTD: behavioral variant frontotemporal dementia; CBD: corticobasal degeneration; AD: Alzheimer's disease; HME: hereditary multiple exostoses; JME: juvenile myoclonic epilepsy; FCD: focal cortical dysplasia. IPL: inferior parietal lobule; SMA: supplementar motor area; M1: primary motor area; STG: superior temporal gyrus, ITG: inferior temporal gyrus; IFG: inferior frontal gyrus; SFG: superior frontal gyrus; ACC: anterior cingulate cortex; OFC: orbito frontal cortex; DLPFC: dorsolateral prefrontal cortex. FWE: family-wise error; FDR: False Discovery Rate; TIV: total intracranial volume.

matter (WM) reductions in the genu of the corpus callosum, the uncinate fasciculus, the corona radiate, the internal capsulae and the inferior fronto-occipital fasciculus in antisocial personality behavior (Sundram et al., 2011).

All the above studies but one (Sundram et al., 2011) were performed using Voxel Based Morphometry (VBM), a whole brain technique for characterizing regional volume and tissue concentration differences in structural magnetic resonance imaging (MRI) (Ashburner and Friston, 2000, 2001; Good et al., 2001; Mechelli et al., 2005). These studies typically compared a group of patients against a group of controls and reported neuroanatomical differences at group level. The results of these studies have had limited translational impact in clinical and forensic practice, where one needs to make inferences at the level of the individual. In recent years however, an increasing number of research groups have attempted to overcome this issue by performing single case studies in which an individual was compared against a control group. Table 1 shows existing neuroimaging studies of neurological and psychiatric disorders that have used such VBM single case approach. However, the interpretation of the results is problematic due to a number of outstanding methodological issues. The main concern is the possibility that any statistically significant difference between a single subject and a control group might reflect normal individual variability in neuroanatomy rather than the neuropathology of the specific disease under investigation. A second concern is that the use of two-sample t-tests requires the data to be sampled from normally distributed populations; therefore the validity of the above studies was based on the assumption that the patient's value constituted the mean value of a hypothetical population with a variance equal to that of the control group (see for Muhlau et al., 2009 for review). This issue was evaluated by Salmond et al. (2002), who compared the false-positive rates in a series of VBM analyses at different degrees of smoothness. The authors demonstrated that the number of false positives decreased with smoothing and therefore suggested that VBM single case analysis could be performed as long as a sufficient smoothing kernel was applied (Salmond et al., 2002). More recently, Viviani et al. (2007) examined the impact of non-normality on the likelihood of Type I error rates in single case VBM studies. Using both simulated and empirical data, the authors reported that smoothing was only partially effective in reducing the impact of deviation from normality; however it was possible to use voxel-by-voxel logit transformation of the raw signal value to minimize the impact of non-normality (Viviani et al., 2007).

The aim of the present investigation was to examine the reliability of single case structural neuroimaging studies of neurological and psychiatric patients by empirically estimating the likelihood of detecting significant differences in gray matter volume in individuals free from neurological or psychiatric diagnosis. Using Voxel Based Morphometry, we compared a total of 200 single subjects against a group of 16 controls matched for age and gender. This allowed us to empirically estimate, in the brain of a disease-free individual, the number of detected 10-voxel differences in each cortical and subcortical region for different corrected voxel-wise significance thresholds. Based on previous analyses of neuroanatomical variability in the healthy population (Fornito et al., 2008; Good et al., 2001; Pruessner et al., 2002; Salmond et al., 2002; Spasojević et al., 2011) we hypothesized a large number of significant differences in single subjects relative to their control group when using a standard statistical threshold of $p<0.05$ (corrected for multiple comparisons across the whole brain). However, these differences would be minimized with the use of more conservative statistical thresholds such as $p<0.01$ (corrected) and $p<0.001$ (corrected). In addition, based on previous ontogenetic and phylogenetic findings on brain evolution (Gogtay et al., 2004; Hill et al., 2010; Huttenlocher, 1990; Sherwood et al., 2008; Stewart and Disotell, 1998; Watson et al., 2006), we expected that differences would not be equally widespread across the whole brain. For instance there is increasing evidence that quantitative differences in the neocortex are associated with individual differences in cognition and behavior in humans (Carreiras et al., 2009; Casey et al., 2000, 2005; Draganski et al., 2004; Fleming et al., 2010; Scholz et al., 2009). Moreover, within the neo-cortex, we expected to find more differences in frontal and temporal lobes, which *control* language (Miller and Cohen, 2001) and high-level cognitive and social functions which differ widely among individuals (Adolphs, 2009; Asplund et al., 2010; Bocková et al., 2007; Fleming et al., 2010), rather than in occipital and parietal cortices, which *control* more *ancestral* functions, such as vision, *praxis* and attention. Finally, we expected the number of significant differences in single subjects relative to their control group to vary as a function of the degree of smoothing applied to the data, which has been shown to influence false positive rate in previous studies (Salmond et al., 2003; Viviani et al., 2007).

Methods

Subjects

We used data from the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC) which are available at http://fcon_1000.projects.nitrc.org/fcpClassic/FcpTable.html (Biswal et al., 2010; Buccigrossi et al., 2008). We chose the Cambridge (Massachusetts, USA) and Beijing (China) datasets because of their large sample size ($n=198$) and their matched age range (18–28). We randomly selected 100 subjects (50 males) from each dataset; all selected participants were right-handed and aged from 18 to 24 (mean: 20.63; standard deviation: ± 1.8). All selected participants had never received a neurological or psychiatric diagnosis. All the subjects in the Beijing dataset were of Chinese origin, whereas ethnicity information was not available for the Cambridge dataset.

MRI data acquisition

All participants underwent the acquisition of a structural MRI scan using a 3 T MRI system. A T1-Weighted sagittal three-dimensional magnetization-prepared rapid gradient echo (MPRAGE) sequence was acquired, covering the entire brain. For the acquisition of the Cambridge dataset, the following parameters were used: 144 slices, voxel resolution 1.2, 1.2,1.2; matrix 192×192. For the acquisition of the Beijing dataset, the following parameters were used: 128 slices, voxel resolution 1.0,1.0,1.3; matrix 181×175.

Data analysis

Preprocessing

After checking for scanner artifacts and gross anatomical abnormalities for each subject, we reoriented the original images along the anterior-posterior commissure (AC-PC) line and set the AC as the origin of the spatial coordinates to assist the normalization algorithm. The new segmentation procedure implemented in SPM8 (http://www.fil.ion.ucl.ac.uk/spm), running under Matlab 7.1 (Math Works, Natick, MA, USA) was used to segment all the images into gray matter (GM) and white matter (WM), i.e. maps of probability values representing the probability of a voxel containing a specific tissue type. A fast diffeomorphic image registration algorithm (DARTEL; Ashburner, 2007), available as a toolbox in SPM8, was used to warp the GM partitions into a new study-specific reference space representing an average of all the subjects included in the analysis. As an initial step, a set of study-specific templates and the corresponding deformation fields, required to warp the data from each subject to the new reference space, were created using the GM partitions (Ashburner and Friston, 2009). Each subject-specific deformation field was used to warp the corresponding GM partition into the new reference space with the aim of maximizing accuracy and sensitivity (Yassa and Stark, 2009); the warped GM partitions were affine transformed into the MNI space and smoothed with a 8-mm full-width at half-maximum (FWHM) Gaussian kernel typically used

in VBM studies (Eriguchi et al., 2010; Feldmann et al., 2008; Freudenmann et al., 2010; Migliaccio et al., 2011; Sehm et al., 2011; Valdes-Sosa et al., 2011; Zahn et al., 2005). An additional 'modulation' step was used to scale the GM probability values by the Jacobian determinants of the deformations in order to ensure that the total amount of gray matter in each voxel was conserved after the registration (Ashburner and Friston, 2000; Good et al., 2001; Mechelli et al., 2005). After this preprocessing, we obtained smoothed, modulated, normalized data that were used for the statistical analysis. In order to examine how the degree of smoothing affected the results, we also smoothed the data using 4-mm and 12-mm FWHM Gaussian kernels consistent with previous studies (Salmond et al., 2002; Viviani et al., 2007). In addition, in order to examine the impact of modulation on the results, we repeated the pre-processing without using the modulation step.

Statistical analysis

Each single subject scan was compared to 16 gender- and age-matched (±1 year) controls. This sample size was chosen for two main reasons: firstly, a typical neuroimaging study of regional differences includes 12–16 subjects per experimental group (Friston et al., 1999); secondly, a recent analysis of effect size in classical inference has suggested that, in order to optimize the sensitivity to large effects while minimizing the risk of detecting trivial effects, the optimum sample size for a study is 16 (Friston, 2012). Age was entered into the design matrix as a covariate of no interest to minimize any impact of this variable on the findings. To exclude voxels outside the brain, we used an implicit mask to remove all the voxels whose intensity fell below the 20% of the mean image intensity. To identify regionally specific changes that were not confounded by global differences in gray matter volume, we used the proportional scaling option. A two-sample t-test was performed to identify significant increases and decreases in each subject relative to their control group. Statistical inferences were made at voxel-level using three different thresholds with family-wise error (FWE) correction for multiple comparisons across the whole brain: $p<0.05$, $p<0.01$ and $p<0.001$. In addition, we only report clusters which comprised of 10 or more voxels. The number of significant increases and decreases in one dataset was compared against the number of significant increases and decreases in the other dataset using independent sample t-tests and a statistical threshold of $p<0.05$. Likewise, the number of significant increases was compared against the number of significant decreases within each dataset using paired sample t-tests and a statistical threshold of $p<0.05$.

Brain areas individuation

From the SPM output, i.e. the list of MNI coordinates of the areas showing significant increases or decreases, we derived the corresponding areas using the Automated Anatomical Labeling (AAL) atlas as implemented in Pickatlas software (http://fmri.wfubmc.edu/software/PickAtlas).

Results

Number of significant clusters in each subject

We compared each subject with 16 gender- and age-matched controls and then computed the average number of significant clusters using three different statistical thresholds ($p<0.05$ FWE corrected; $p<0.01$ FWE corrected; $p<0.001$ FWE corrected). Table 2 reports the average number of clusters in each subject relative to their control group, for increases and increases separately, as a function of the statistical threshold.

p<0.05 FWE corrected

Using a statistical threshold of $p<0.05$ FWE corrected, single subjects showed an average of 7.46 (±4.54) significant increases in gray matter volume relative to their control group. A direct comparison indicated that the number of increases did not differ between the Beijing (7.16±4.67) and Cambridge (7.77±4.42) datasets (two independent sample t-test, t value = 0.948, df = 198, p-value = 0.344). In addition, single subjects showed an average of 2.69 (±2.97) significant decreases in gray matter volume relative to their control group. Again, the number of decreases did not differ between the Beijing (2.68±3.14) and Cambridge (2.71±2.80) datasets (two independent sample t-test, t value = 0.071, df = 198, p-value = 0.943). In contrast the number of increases was significantly greater than the number of decreases (Paired sample t-test, t value = 30.10, df = 199, p-value < 0.001).

p<0.01 FWE corrected

Using a statistical threshold of $p<0.01$ FWE corrected, single subjects showed an average of 3.49 (±2.87) significant increases relative to the control group. A direct comparison indicated that this number did not differ between the Beijing (3.25±2.73) and Cambridge (3.73±3) datasets (two independent t-test, t value = 1.18, df = 198, p-value = 0.239). In addition, single subjects showed an average of 0.87 (±1.43) significant decreases relative to the control group. Again, this number did not differ between the Beijing (0.89±1.51) and Cambridge (0.85±1.37) datasets (two independent sample t-test, t value = −0.196, df = 198, p-value = 0.845). In contrast the number of increases was significantly greater than the number of decreases (Paired sample t-test, t value = 12.309, df = 199, p-value = $p<0.001$).

p<0.001 FWE corrected

Using a statistical threshold of $p<0.001$ FWE corrected, single subjects showed an average of 0.95 (±1.27) significant increases in gray matter volume relative to their control group. A direct comparison indicated that this number did not differ between the Beijing (0.82±1.2) and Cambridge (1.08±1.32) datasets (two independent sample t-test, t value = 1.45, df = 198, p-value = 0.149). In addition, single subjects showed an average of 0.08 (±0.34) significant decreases in gray matter volume relative to their control group. Again, this

Table 2
Number of significant differences in each subject.

	p<0.05 FWE		P<0.01 FWE		P<0.001 FWE	
	Increases	Decreases	Increases	Decreases	Increases	Decreases
Beijing	7.16 (4.67)	2.68 (3.14)	3.25 (2.73)	0.89 (1.51)	0.82 (1.20)	0.08 (0.3)
Cambridge	7.77 (4.42)	2.71 (2.80)	3.73 (3)	0.85 (1.37)	1.08 (1.32)	0.09 (0.37)
Mean	7.46 (4.54)	2.69 (2.97)	3.49 (2.87)	0.87 (1.43)	0.95 (1.27)	0.08 (0.34)
p*	0.344	0.943	0.239	0.845	0.149	0.838

Mean and standard deviation (in brackets) are reported for each dataset (Cambridge, Beijin), for each statistical threshold and for each direction. FWE: family-wise error correction; Increases: average number of significant increases (with an extent threshold of 10 voxels) in each subject relative to their control group; decreases: average number of significant decreases (with an extent threshold of 10 voxels) in each subject relative to their control group. p* = two independent sample t-test.

number did not differ between the Beijing (0.08 ± 0.3) and Cambridge (0.09 ± 0.37) datasets (two independent sample t-test, t value = 0.205, df = 198, p-value = 0.838); In contrast the number of increases was significantly greater than the number of decreases (Paired sample t-test, t value = 9.28, df = 199, p-value < 0.001).

Likelihood of detecting differences in a single subject

After establishing the average number of regions showing significant differences, we computed the likelihood that a single subject would show a given number of significant increases or decreases when compared against its control group. The results are summarized in Table 3.

As expected, the likelihood of a single subject showing significant differences to their control group depended on the statistical threshold used. For example, the likelihood of detecting at least one *increase* was 93.5% at $p<0.05$ (corrected), 81% at $p<0.01$ (corrected) and 49% at $p<0.001$ (corrected), whereas the likelihood of detecting at least one *decrease* was 71% at $p<0.05$ (corrected), 36.5% at $p<0.01$ (corrected) and 5.5% at $p<0.001$ (corrected). In addition, the likelihood that a single subject would show a given number of significant differences relative to their control group was inversely related to the value of that number. For instance, with a statistical threshold of $p<0.05$ (corrected), the likelihood of detecting at least 1, 5, and 10 *increases* was 93.5%, 68% and 39.5% respectively, whereas the likelihood of detecting at least 1, 5 and 10 *decreases* was 71%, 24%, and 3.5% respectively.

Table 3
Likelihood of detecting n differences in a specific subject.

	Cambridge		Beijing		Mean likelihood	
	Increases	Decreases	Increases	Decreases	Increases	Decreases
P<0.05 FWE corrected						
1 region	94	75	93	67	93.5%	71%
2 regions	91	60	87	49	89%	54.5%
3 regions	86	48	79	37	82.5%	42.5%
4 regions	81	35	72	29	76.5%	32%
5 regions	74	25	62	23	68%	24%
6 regions	70	18	56	18	73%	18%
7 regions	64	9	49	16	56.5%	12.5%
8 regions	58	5	46	13	52%	9%
9 regions	50	3	41	9	45.5%	6%
10 regions	43	2	36	5	39.5%	3.5%
11 regions	33	1	32	1	32.5%	1%
12 regions	23	–	23	–	23%	–
13 regions	20	–	16	–	18%	–
14 regions	11	–	9	–	10%	–
15 regions	7	–	7	–	7%	–
16 regions	3	–	4	–	3.5%	–
P<0.01 FWE corrected						
1 region	84	36	78	37	81%	36.5%
2 regions	73	20	68	22	70.5%	21%
3 regions	63	5	55	14	59%	9.5%
4 regions	50	3	39	9	44.5%	6%
5 regions	41	1	32	5	36.5%	3%
6 regions	28	1	26	2	27%	1.5%
7 regions	18	–	14	–	16%	–
8 regions	14	–	7	–	10.5%	–
9 regions	6	–	4	–	5%	–
≥10 regions	5	–	1	–	3%	–
P<0.001 FWE corrected						
1 region	56	6	42	5	49%	5.5%
2 regions	31	3	23	1	27%	2%
3 regions	13	–	12	–	12.5%	–
4 regions	7	–	3	–	5%	–
5 regions	4	–	1	–	2.5%	–
6 regions	1	–	1	–	1%	–

The number of subjects showing up to one, two, three, etc. statistically significant differences (with an extent threshold of 10 voxels) in a single subject compared to their control group are reported for each dataset (Cambridge, Beijin), for each statistical threshold and for each direction. The likelihood of detecting up to one, two, three, etc. significant differences in a single subject compared to their control group is reported in percentage. FWE: family-wise error correction.

Number of significant differences in each region

After establishing the average number of differences between a single subject and their control group and the likelihood of detecting a given number of significant differences in a single subject, we examined the number of clusters (within each specific region. The results are summarized in Table 4 and represented graphically in Figs. 1 & 2. It can be seen that the number of clusters within each specific region tended to be similar for the Beijing and Cambridge datasets, irrespective of the statistical threshold used. As expected, the number of significant differences in a region was inversely related to the statistical threshold used; a one-way ANOVA confirmed that both the number of increases ($F = 65.23$, $df = 341$, p-value < 0.001) and decreases ($F = 107.33$, $df = 342$, p-value < 0.001) varied significantly as a function of the statistical threshold. Figs. 1 & 2 also show that there was a large degree of variability in the total number of significant differences from one region to another. More specifically, the areas showing the greatest number of increases (exceeding the third quartile of the distribution) appeared to be localized in the neocortex and included the precentral gyrus, superior and middle frontal gyrus, superior, middle and inferior temporal gyrus, postcentral gyrus, parietal superior gyrus and precuneus, whereas there were fewer differences in subcortical regions such as the parahippocampus, hippocampus, amygdala, caudate, putamen and pallidum. Likewise, the areas showing the greatest number of decreases tended to be localized in the neocortex and included the superior, middle and medial frontal gyrus, pars opercularis, insula, anterior and middle cingulum, superior and middle temporal gyrus, postcentral gyrus, parietal superior gyrus and precuneus.

Each region tended to show a greater number of increases than decreases; for instance, the average number of increases per region was 6.5 ± 5.6 whereas the average number of decreases was 2.3 ± 2 in $p<0.05$ corrected, 3 ± 2.8 and 0.7 ± 0.7 in $p<0.01$ corrected and 0.8 ± 0.9 and 0.06 ± 0.1 in $p<0.001$ corrected. It is worth nothing that there are also regions which did not show any significant difference in any of the two hundred comparisons performed in total (see Table 4), and that the number of these regions is greater in the case of decreases ($n = 15$) than increases ($n = 6$). More specifically, regions which did not show any increases included the right middle orbital gyrus, right posterior cingulum, left amygdala, left cerebellum 3 L and 10 L and right cerebellum 7b, whereas regions which did not show any decreases included the bilateral middle orbital frontal gyrus, bilateral amygdala, right heschl, right superior occipital gyrus, bilateral inferior occipital gyrus, left supramarginal gyrus, right paracentral lobule, bilateral pallidum, bilateral cerebellum 3 and right cerebellum 8.

Likelihood of detecting local maxima in a specific region

After establishing the number of significant differences within each specific region, we computed the likelihood that a specific region would show significant local maxima when a single subject is contrasted against their control group. The results are summarized in Table 4 and represented graphically in Fig. 3.

As expected, the likelihood that a specific region would show a given number of significant differences depended on the statistical threshold used. The regions with the highest likelihood of showing significant differences were located in neocortical regions, particularly in the bilateral middle frontal gyrus, middle temporal gyrus and inferior temporal gyrus for both decreases and increases. In the vast majority of regions, the likelihood of increases was greater than that of decreases; for instance, the average likelihood of detecting significant increases in a given region was 6.5% ± 5.6% whereas the average

Table 4
Number of significant differences and likelihood of detecting significant differences in each region.

	P<0.05 FWE corrected				P<0.01 FWE corrected				P<0.001 FWE corrected			
	Increases		Decreases		Increases		Decreases		Increases		Decreases	
	n	%	N	%	n	%	n	%	n	%	N	%
Precentral L	18–15	16.5%	3–4	3.5%	9–4	6.5%	2–2	2%	3–0	1.5%	0–0	–
Precentral R	12–10	11%	5–2	3.5%	5–4	4.5%	3–1	2%	2–1	1.5%	1–0	0.5%
Frontal Sup L	13–14	13.5%	5–6	5.5%	8–10	9%	1–1	1%	3–3	3%	0–0	–
Frontal Sup R	13–15	14%	7–5	6%	6–7	6.5%	0–2	1%	5–1	3%	0–0	–
Frontal Sup Orb L	4–4	4%	1–3	2%	2–3	2.5%	0–2	1%	1–1	1%	0–0	–
Frontal Sup Orb R	4–2	3%	2–3	2.5%	2–2	2%	1–1	1%	2–0	1%	1–1	1%
Frontal Mid L	25–19	22%	7–3	5%	12–8	10%	2–1	1.5%	1–1	1%	0–0	–
Frontal Mid R	24–22	23%	10–6	8%	8–8	8%	1–2	1.5%	1–1	1%	0–0	–
Frontal Mid Orb L	5–4	4.5%	3–1	2%	5–2	3.5%	1–0	0.5%	0–0	–	0–0	–
Frontal Mid Orb R	11–7	9%	1–3	2%	6–4	5%	1–0	0.5%	0–2	1%	0–0	–
Frontal Inf Oper L	6–1	3.5%	4–6	5%	2–0	1%	0–3	1.5%	0–0	–	0–0	–
Frontal Inf Oper R	8–4	6%	4–5	4.5%	5–2	3.5%	0–1	0.5%	2–1	1.5%	0–0	–
Frontal Inf Tri L	7–12	9.5%	1–5	3%	5–3	4%	0–0	–	1–0	0.5%	0–0	–
Frontal Inf Tri R	11–7	9%	1–3	2%	6–4	5%	0–2	1%	1–0	0.5%	0–0	–
Frontal Inf Orb L	12–8	10%	1–2	1.5%	7–1	4%	0–0	–	1–0	0.5%	0–0	–
Frontal Inf Orb R	7–5	6%	3–1	2%	3–1	2%	1–0	0.5%	2–0	1%	0–0	–
Rolandic Oper L	3–8	5.5%	1–2	1.5%	2–3	2.5%	0–2	1%	1–0	0.5%	0–0	–
Rolandic Oper R	4–6	5%	2–5	3.5%	4–4	4%	1–0	0.5%	0–2	1%	0–0	–
Supp Mot Area L	9–8	8.5%	3–1	2%	4–4	4%	1–0	0.5%	0–0	–	0–0	–
Supp Mot Area R	10–9	9.5%	4–4	4%	3–6	4.5%	1–2	1.5%	0–1	0.5%	0–0	–
Olfactory L	0–1	0.5%	0–1	0.5%	0–0	–	0–0	–	0–0	–	0–0	–
Olfactory R	4–3	3.5%	0–2	1%	1–2	1.5%	0–1	0.5%	1–0	0.5%	0–0	–
Frontal Sup Medial L	14–9	11.5%	3–9	6%	7–3	5%	2–5	3.5%	1–1	1%	0–1	0.5%
Frontal Sup Medial R	6–8	7%	0–1	0.5%	4–4	4%	0–0	–	2–2	2%	0–0	–
Rectus L	2–7	4.5%	1–2	1.5%	0–4	2%	0–0	–	1–1	1%	0–0	–
Rectus R	3–2	2.5%	0–1	0.5%	1–0	0.5%	0–1	0.5%	1–0	0.5%	0–0	–
Insula L	11–5	8%	3–7	5%	3–0	1.5%	2–1	1.5%	2–0	1%	0–0	–
Insula R	13–2	7.5%	5–7	6%	6–2	4%	1–1	1%	1–0	0.5%	1–0	0.5%
Cingulum Ant L	7–2	4.5%	4–4	4%	1–1	1%	1–1	1%	0–0	–	0–0	–
Cingulum Ant R	2–3	2.5%	3–4	3.5%	1–2	1.5%	2–4	3%	0–2	1%	0–0	–
Cingulum Mid L	8–4	6%	3–2	2.5%	4–0	2%	0–1	0.5%	1–0	0.5%	0–0	–
Cingulum Mid R	10–8	9%	5–7	6%	6–6	6%	2–1	1.5%	0–0	–	0–0	–
Cingulum Post L	2–0	1%	1–0	0.5%	0–0	–	0–0	–	0–0	–	0–0	–
Cingulum Post R	0–0	–	3–1	2%	0–0	–	1–0	0.5%	0–0	–	0–0	–
Hippocampus L	6–4	5%	3–3	3%	5–2	3.5%	1–1	1%	1–0	0.5%	0–0	–
Hippocampus R	3–3	3%	2–2	2%	0–2	1%	1–2	1.5%	0–1	0.5%	0–0	–
Parahippocampus L	4–7	5.5%	2–3	2.5%	2–2	2%	1–1	1%	1–1	1%	1–1	1%
Parahippocampus R	7–12	9.5%	3–3	3%	3–8	5.5%	1–1	1%	1–5	3%	0–1	0.5%
Amygdala L	0–0	–	0–0	–	0–0	–	0–0	–	0–0	–	0–0	–
Amygdala R	1–1	1%	0–1	0.5%	0–0	–	0–1	0.5%	0–0	–	0–0	–
Heschl L	0–1	0.5%	2–0	1%	0–0	–	2–0	1%	0–0	–	1–0	0.5%
Heschl R	1–0	0.5%	0–0	–	0–0	–	0–0	–	0–0	–	0–0	–
Temporal Sup L	13–21	17%	3–2	2.5%	5–8	6.5%	1–0	0.5%	3–1	2%	0–0	–
Temporal Sup R	18–15	16.5%	2–4	3%	10–8	9%	0–1	0.5%	3–2	2.5%	0–1	0.5%
Temporal Pole Sup L	3–6	4.5%	1–0	0.5%	2–2	2%	0–0	–	0–0	–	0–1	0.5%
Temporal Pole Sup R	6–3	4.5%	5–3	4%	4–2	3%	2–1	1.5%	2–0	1%	0–0	–
Temporal Mid L	24–27	25.5%	10–8	9%	5–12	8.5%	3–2	2.5%	0–3	1.5%	0–0	–
Temporal Mid R	27–14	20.5%	7–8	7.5%	10–9	9.5%	1–3	2%	2–4	3%	0–0	–
Temporal Pole Mid L	4–4	4%	2–2	2%	3–0	1.5%	1–2	1.5%	1–0	0.5%	0–0	–
Temporal Pole Mid R	9–3	6%	0–3	1.5%	4–1	2.5%	0–2	1%	2–0	1%	0–0	–
Temporal Inf L	21–20	20.5%	8–6	7%	9–7	8%	1–2	1.5%	4–2	3%	0–0	–
Temporal Inf R	26–15	20.5%	4–10	7%	12–9	10.5%	1–5	3%	2–3	2.5%	0–0	–
Occipital Sup L	3–6	4.5%	2–3	2.5%	1–2	1.5%	0–2	1%	0–1	0.5%	0–0	–
Occipital Sup R	5–6	5.5%	0–0	–	3–5	4%	0–0	–	1–0	0.5%	0–0	–
Occipital Mid L	18–10	14%	8–2	5%	8–5	6.5%	2–1	1.5%	4–0	2%	0–0	–
Occipital Mid R	7–14	10.5%	3–3	3%	3–5	4%	1–1	1%	1–0	0.5%	0–0	–
Occipital Inf L	7–8	7.5%	0–0	–	4–4	4%	0–0	–	3–1	2%	1–0	0.5%
Occipital Inf R	7–8	7.5%	0–0	–	3–3	3%	1–0	0.5%	1–0	0.5%	0–0	–
Calcarine L	8–5	6.5%	2–0	1%	3–3	3%	1–0	0.5%	3–3	3%	0–0	–
Calcarine R	5–8	6.5%	1–0	0.5%	4–4	4%	0–0	–	1–4	2.5%	0–0	–
Cuneus L	2–4	3%	0–3	1.5%	1–2	1.5%	1–0	0.5%	1–0	0.5%	0–0	–
Cuneus R	7–8	7.5%	7–0	3.5%	2–4	3%	0–0	–	2–1	1.5%	0–0	–
Lingual L	8–9	8.5%	4–2	3%	2–5	3.5%	0–1	0.5%	0–1	0.5%	0–0	–
Lingual R	15–11	13%	2–3	2.5%	9–7	8%	2–0	1%	0–2	2%	0–0	–
Fusiform L	13–14	13.5%	3–4	3.5%	5–5	5%	1–1	1%	3–3	3%	0–0	–
Fusiform R	10–11	10.5%	6–4	5%	5–5	5%	3–4	3.5%	3–2	2.5%	0–0	–
Postcentral L	18–10	14%	6–2	4%	12–5	8.5%	2–1	1.5%	3–2	2.5%	0–0	–
Postcentral R	10–11	10.5%	2–4	3%	9–7	8%	0–2	1%	1–3	2%	0–0	–
Parietal Sup L	15–9	12%	1–2	1.5%	10–7	8.5%	1–0	0.5%	2–2	2%	0–0	–
Parietal Sup R	10–3	6.5%	2–3	2.5%	8–0	4%	2–1	1.5%	0–0	–	0–0	–
Parietal Inf L	7–16	11.5%	6–5	5.5%	2–6	4%	2–0	1%	1–1	1%	0–0	–

(continued on next page)

Table 4 (*continued*)

	P<0.05 FWE corrected				P<0.01 FWE corrected				P<0.001 FWE corrected			
	Increases		Decreases		Increases		Decreases		Increases		Decreases	
	n	%	N	%	n	%	n	%	n	%	N	%
Parietal Inf R	11–3	7%	4–3	3.5%	7–1	4%	0–1	0.5%	0–1	0.5%	0–0	–
Supramarginal L	10–9	9.5%	0–0	–	4–4	4%	0–0	–	2–1	1.5%	0–0	–
Supramarginal R	6–9	7.5%	3–4	3.5%	5–2	3.5%	1–1	1%	0–0	–	0–0	–
Angular L	7–7	7%	2–3	2.5%	5–3	4%	1–1	1%	0–0	–	1–0	0.5%
Angular R	5–8	6.5%	3–3	3%	2–2	2%	0–1	0.5%	0–0	–	0–1	0.5%
Precuneus L	19–13	16%	7–6	6.5%	12–12	12%	1–1	1%	5–0	2.5%	0–0	–
Precuneus R	10–14	12%	6–4	5%	6–6	6%	1–1	1%	2–1	1.5%	0–0	–
Paracentral Lobule L	5–4	4.5%	2–0	1%	1–2	1.5%	0–0	–	1–0	0.5%	0–0	–
Paracentral Lobule R	5–0	2.5%	0–0	–	0–0	–	0–0	–	0–0	–	0–0	–
Caudate L	0–2	1%	4–1	3%	0–0	–	1–0	0.5%	0–0	–	1–0	0.5%
Caudate R	3–2	2.5%	0–1	0.5%	0–0	–	0–0	–	0–0	–	1–0	0.5%
Putamen L	3–3	3%	1–2	1.5%	2–1	1.5%	1–0	0.5%	0–0	–	0–0	–
Putamen R	1–3	2%	3–4	3.5%	0–2	1%	1–1	1%	0–1	0.5%	0–0	–
Pallidum L	0–1	0.5%	0–0	–	0–1	0.5%	0–0	–	0–0	–	0–0	–
Pallidum R	0–1	0.5%	0–0	–	0–1	0.5%	0–0	–	0–0	–	0–0	–
Thalamus L	1–1	1%	0–2	1%	0–0	–	0–0	–	0–0	–	0–0	–
Thalamus R	2–3	2.5%	0–1	0.5%	1–2	1.5%	0–0	–	0–1	0.5%	0–0	–
Cerebelum Crus 1 L	9–9	9%	3–1	2%	6–5	5.5%	0–0	–	2–1	1.5%	0–0	–
Cerebelum Crus 1 R	2–7	4.5%	6–2	4%	2–5	3.5%	1–1	1%	0–0	–	0–0	–
Cerebelum Crus 2 L	2–7	4.5%	0–2	1%	2–1	1.5%	1–0	0.5%	2–0	1%	0–0	–
Cerebelum Crus 2 R	4–3	3.5%	1–2	1%	1–2	1.5%	0–1	0.5%	0–0	–	0–1	0.5%
Cerebelum 3 L	0–0	–	0–0	–	0–0	–	0–0	–	0–0	–	0–0	–
Cerebelum 3 R	0–1	0.5%	0–0	–	0–1	0.5%	0–0	–	0–0	–	0–0	–
Cerebelum 4–5 L	4–2	3%	1–3	2%	3–0	1.5%	0–0	–	0–0	–	0–0	–
Cerebelum 4–5 R	2–3	2.5%	2–1	1.5%	0–0	–	1–0	0.5%	0–0	–	0–0	–
Cerebelum 6 L	3–5	4%	1–1	1%	0–2	1%	1–0	0.5%	0–2	1%	0–0	–
Cerebelum 6 R	3–4	3.5%	1–0	0.5%	0–2	1%	0–0	–	0–1	0.5%	0–0	–
Cerebelum 7b L	2–1	1.5%	3–0	1.5%	0–1	0.5%	0–0	–	0–1	0.5%	0–0	–
Cerebelum 7b R	0–0	–	1–1	1%	0–0	–	0–1	0.5%	0–0	–	0–0	–
Cerebelum 8 L	4–6	5%	2–1	1.5%	2–1	1.5%	0–0	–	1–1	1%	0–0	–
Cerebelum 8 R	7–8	7.5%	0–0	–	1–3	2%	0–0	–	0–1	0.5%	0–0	–
Cerebelum 9 L	4–8	6%	3–0	1.5%	2–1	1.5%	0–0	–	1–0	0.5%	0–0	–
Cerebelum 9 R	1–2	1.5%	1–0	0.5%	0–0	–	1–0	0.5%	0–0	–	0–0	–
Cerebelum 10 L	0–0	–	0–0	–	0–0	–	0–0	–	0–0	–	0–0	–
Cerebelum 10 R	1–1	1%	1–0	0.5%	1–1	1%	0–0	–	1–0	0.5%	0–0	–
Vermis	7–5	6%	0–3	1.5%	2–0	1%	0–2	1%	2–0	1%	0–0	–

n denotes the raw number of statistically significant differences (with an extent threshold of 10 voxels) in each region for each dataset (Cambridge, Beijin), for each statistical threshold and for each direction. The first number refers to results obtained with the Cambridge dataset, the second number refers to the results obtained with the Beijing dataset. % denotes the likelihood of detecting significant differences in a specific region; FWE: family-wise error correction; R: right; L: left.

likelihood of detecting a significant decrease was 2.3% ± 2% in $p<0.05$ corrected, 3% ± 2.8% in $p<0.01$ corrected and 0.06% ± 0.1% in $p<0.001$ corrected.

Impact of smoothing

The above results were obtained using GM images smoothed with an 8-mm FWHM Gaussian kernel. Consistent with previous studies (Salmond et al., 2002; Viviani et al., 2007), we repeated the analyses using GM images smoothed with 4-mm and 12-mm FWHM Gaussian kernels. A full description and graphical representation of these additional results are provided in the Supplementary material. In order to assess whether smoothing had a significant impact on the number of significant differences, a repeated measure analysis of variance (ANOVA) was performed. Group (two levels: Cambrigde and Beijing) was used as a between-subject factor, and Smoothing (three levels: 4 mm, 8 mm and 12 mm), Threshold (three levels: $p=0.05$, $p=0.01$ and $p=0.001$ FWE corrected) and Direction (two levels: increase and decrease) were entered as a within-subject factors. This analysis revealed a significant main effect of Smoothing ($F[2;396]=7.03$; $p=0.0009$), Threshold ($F[2;396]=437,33$; $p<0.0001$) and Direction ($F[1;198]=699.75$; $p<0.0001$), but not of Group ($F[1;198]=0.53$; $p=0.58$). Neuman–Keuls post-hoc comparisons indicated that the number of significant differences in single subjects relative to their control group was smaller for 8-mm than 4-mm smoothing ($p=0.03$) and for 12-mm than 4-mm smoothing ($p=0.0005$) but did not differ between 8-mm and 12-mm smoothing ($p=0.1$). The effect of smoothing on the number of significant differences did not differ between the Beijing and Cambridge datasets ($p>0.05$) and furthermore there was no significant Smoothing × Group interaction ($F[2;396]=0.53$; $p=0.58$).

In addition, there was a significant Smoothing × Direction interaction ($F[2;396]=152.16$; $p<0.0001$). Newman–Keuls post-hoc comparisons showed that number of significant increases in single subjects compared to their control group became smaller with higher smoothing (p-value<0.001 for the comparison between 4 mm and 8 mm smoothing; p-value<0.001 for the comparison between 4 mm and 12 mm smoothing; p-value<0.001 for the comparison between 8 mm and 12 mm smoothing), whereas the number of significant decreases in single subjects compared to their control group showed the reverse effect (p-value<0.001 for the comparison between 4 mm and 8 mm smoothing; p-value<0.001 for the comparison between 4 mm and 12 mm smoothing; p-value<0.001 for the comparison between 8 mm and 12 mm smoothing). The effect of smoothing on the number of increases and decreases did not differ between the Beijing and Cambridge datasets ($p>0.05$ for each comparison) and furthermore there was no significant Smoothing × Direction × Group interaction ($F[3;396]=0.13$; $p=0.87$).

Impact of modulation

In order to examine the impact of modulation on the results, we repeated the analyses using unmodulated GM images smoothed

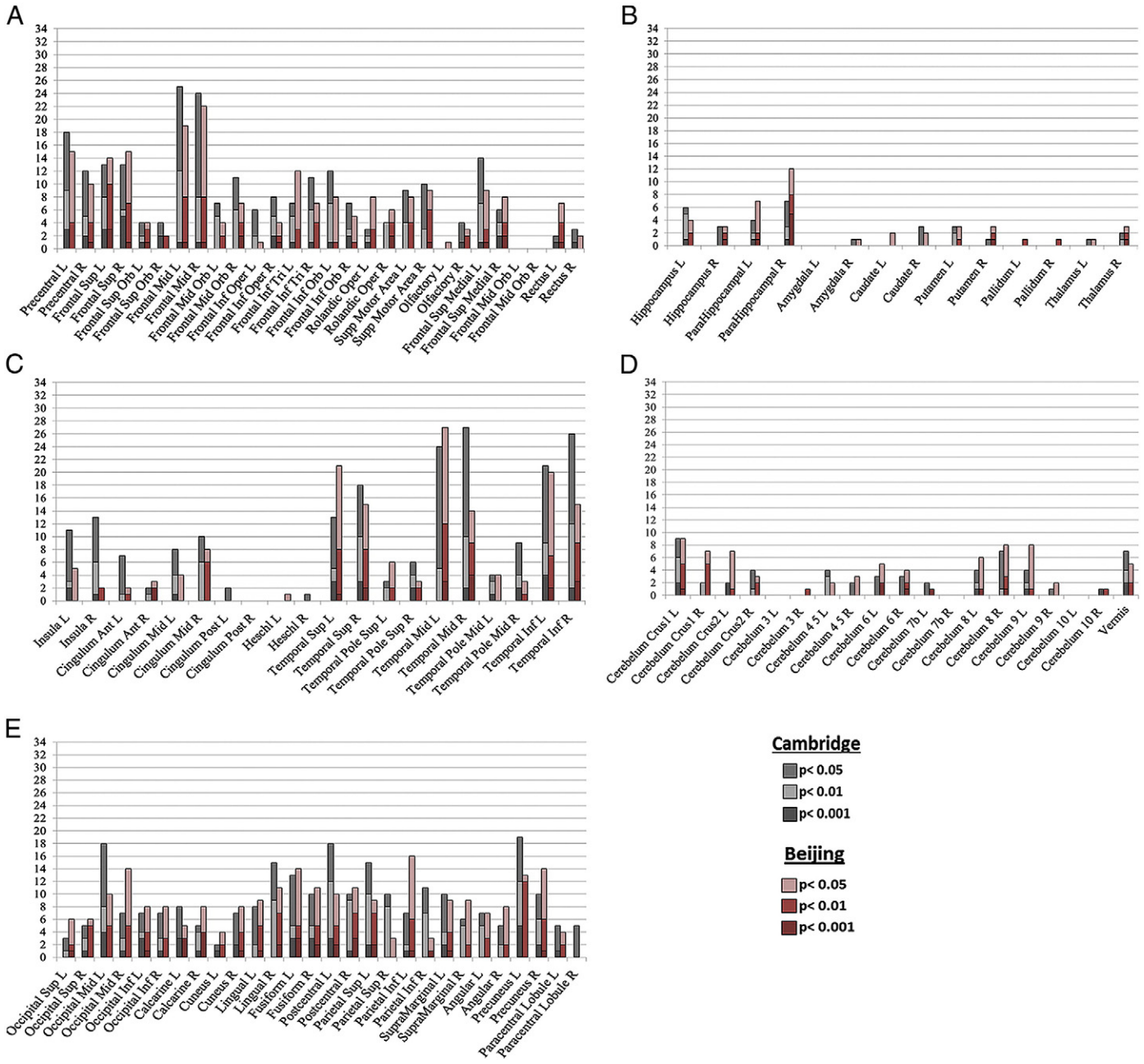

Fig. 1. Total number of significant increases for each region and statistical threshold in the single subject VBM analyses. A = frontal lobe; B = subcortical structures; C = temporal lobe; D = cerebelum; E = parieto–occipital lobe.

with a 8-mm FWHM Gaussian kernel. A full description and graphical representation of these additional results are provided in the Supplementary material. In order to assess whether modulation had a significant impact on the number of significant differences, a repeated measure analysis of variance (ANOVA) was performed. Group (two levels: Cambrigde and Beijing) was used as a between-subject factor, and Modulation (two levels: modulated vs. unmodulated images), Threshold (three levels: $p=0.05$, $p=0.01$ and $p=0.001$ FWE corrected) and Direction (two levels: increase and decrease) were entered as a within-subject factors. This analysis revealed a significant main effect of Modulation ($F[1;198]=6.18$; $p=0.01$; 3.13 ± 3.63 vs 2.59 ± 3.60 significant differences for unmodulated and modulated data respectively), Threshold ($F[2,396]=772,12$; $p<0.0001$; 5.34 ± 4.47 vs 2.53 ± 2.72 vs. 0.71 ± 1.14 significant differences for $p=0.05$, $p=0.01$ and $p=0.001$ FWE corrected respectively) and Direction ($F[1;198]=94.71$; $p<0.0001$; 3.38 ± 3.89 vs. 2.33 ± 3.26 significant differences for increase and decrease respectively), but not of Group ($F[1;198]=0.39$; $p=0.52$). The Modulation × Group interaction was not significant ($F[1;198]=0.25$; $p=0.61$) indicating that the effect of modulation on the number of significant differences did not differ between the Beijing and Cambridge dataset. However, there was a significant Modulation × Direction interaction ($F[1;198]=262.80$; $p<0.0001$). Newman–Keuls post-hoc comparisons showed that there was a significant difference between number of increases and decreases in single subjects compared to their control group for both modulated and unmodulated data. However, the number of significant increases was higher than the number of significant decreases for modulated

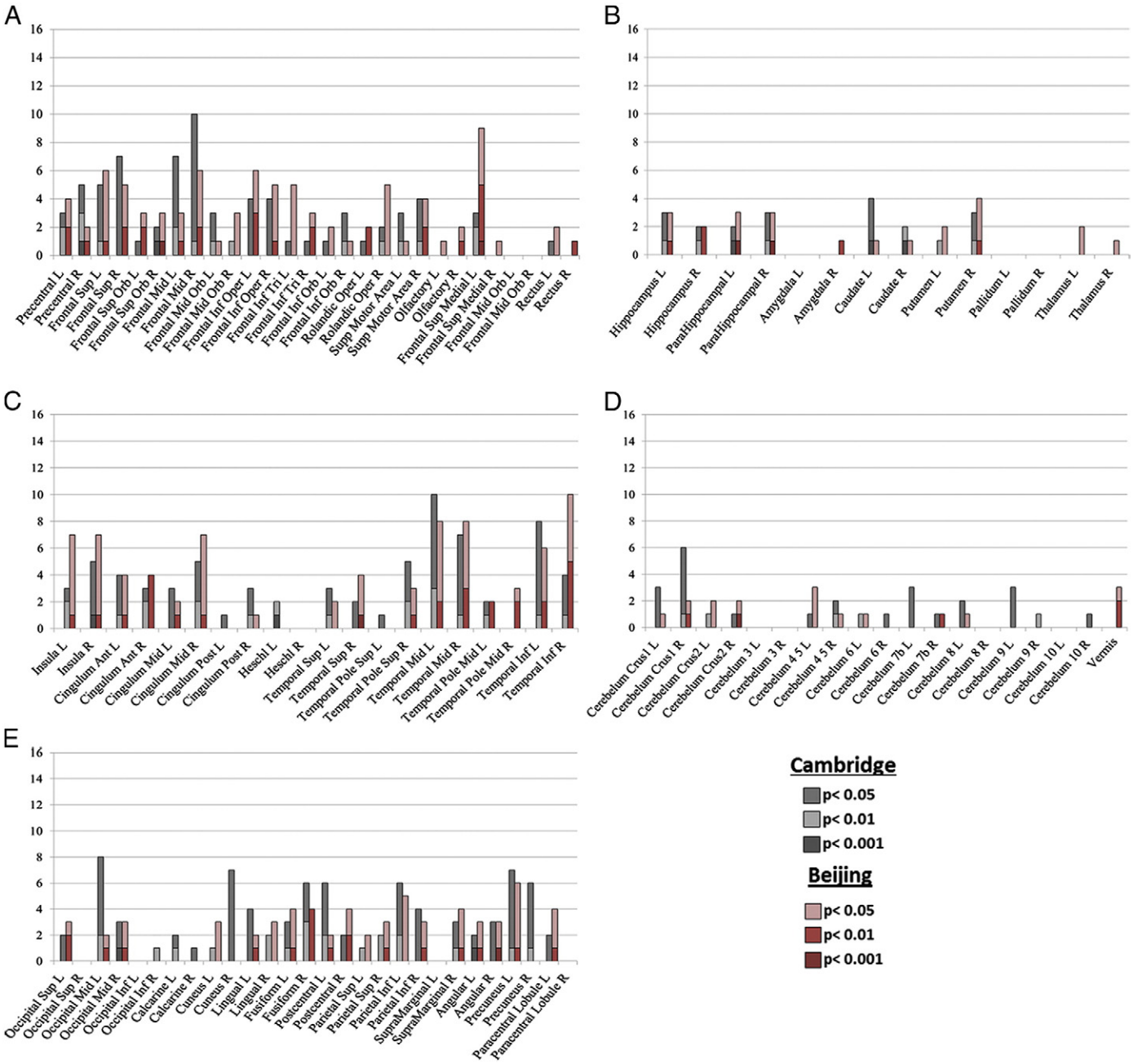

Fig. 2. Total number of significant decreases for each region and statistical threshold in the single subject VBM analyses. A = frontal lobe; B = subcortical structures; C = temporal lobe; D = cerebelum; E = parieto–occipital lobe.

data (3.96 ± 4.16; 1.21 ± 2.20 respectively, $p<0.001$), whereas the opposite was true for unmodulated data (2.80 ± 3.5; 3.46 ± 3.73 respectively $p<0.001$).

Discussion

Previous investigations had examined the impact of non-normality on the validity of VBM studies and identified smoothing and unbalancedness as critical parameters (Salmond et al., 2002; Viviani et al., 2007). Here we expanded these results by empirically estimating the likelihood of detecting significant neuroanatomical differences in the general population. Our specific aim was to report the number of detected 10-voxel differences for different corrected voxel-wise significance thresholds (rather than comparing empirical family-wise error rates to theoretical ones). This was achieved by comparing disease-free single subjects against gender- and age-matched healthy controls using different statistical thresholds. In contrast with previous investigations, we used an optimization of the standard VBM approach which is thought to improve accuracy; we preprocessed the data both with and without modulation and investigated the impact of this variable on the results; we examined significant difference in each cortical and sub-cortical region independently; we characterized increases and decreases in gray matter volume separately; and we repeated our analysis using two large independent datasets.

Consistent with our first hypothesis, we detected a number of significant differences in single subjects relative to controls using a standard statistical threshold of $p<0.05$ (corrected); for instance the likelihood of a single subject showing at least one significant difference was as high as 93.5% for increases and 71% for decreases. This aspect of our results suggests that, when comparing a single neurological or psychiatric patient against a group of controls, the chance of detecting a significant

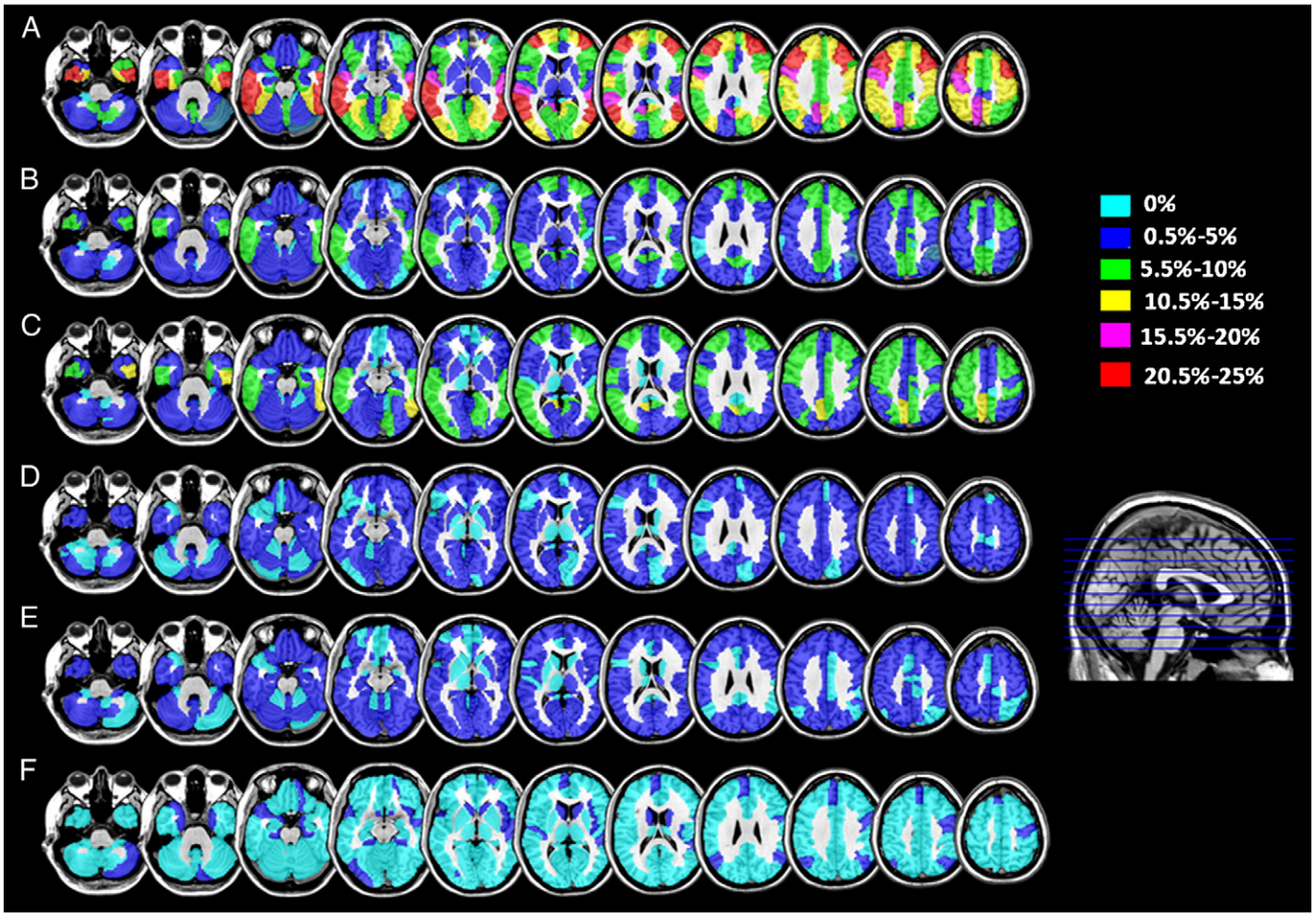

Fig. 3. Areas showing different percentages of significant differences (at least 10 voxels cluster) in the single subject VBM analyses. Increases and decreases in a single subject relative to their control group are shown separately using statistical thresholds of $p<0.05$ FWE corrected, $p<0.01$ FWE corrected and $p<0.001$ FWE corrected. (A) shows increases at $p<0.05$ corrected; (B) shows decreases at $p<0.05$ corrected; (C) shows increases at $p<0.01$ corrected; (D) shows decreases at $p<0.01$ corrected; (E) shows increases at $p<0.001$ corrected and (F) shows decreases at $p<0.001$ corrected.

difference which is not related to the disorder under investigation is much higher than previously expected. This is an important observation since previous single case VBM studies typically used a standard statistical threshold of $p<0.05$ (corrected) and interpreted any significant differences in terms of neuropathology. We also found that the number of significant differences decreased with the use of more conservative statistical thresholds; for instance the likelihood of a single subject showing at least one significant difference decreased to 49% for increases and 5.5% for decreases when using a statistical threshold of $p<0.001$ (corrected). However, changes in the statistical threshold were associated with proportionally smaller changes in the number of significant differences; for example, a 50-fold reduction in the statistical threshold (from $p<0.05$ to $p<0.001$) resulted in a 1.9-fold reduction in the likelihood of a single subject showing at least one significant difference (from 93.5% to 49%). This aspect of our results might be explained by the fact that parametric assumptions become more critical as one moves along the tail of the distribution.

Consistent with our second hypothesis, we found that significant differences were not equally distributed across the 110 areas of the AAL atlas. More specifically, both increases and decreases were more likely to be expressed in the neocortex, particularly the frontal and temporal lobes. The interpretation of significant differences in these regions when comparing a single subject against a group of controls should therefore be particularly cautious. The temporal and frontal lobes sub-serve higher cognitive functions that distinguish humans from nonhuman primates and are associated with high expansion in human postnatal development. More specifically, these regions coordinate purposeful behavior towards reaching higher-level goals (Asplund et al., 2010; Bocková et al., 2007), high levels of social information processing metacognitive introspective abilities (Adolphs, 2009; Fleming et al., 2010), language (Miller and Cohen, 2001), semantic (Olson et al., 2007) and declarative memory (Gour et al., 2011), and comprehension of other's mental state (Jimura et al., 2010). All these abilities widely differ among individuals and such differences in cognition are known to be associated with quantitative differences in the neocortex (Carreiras et al., 2009; Casey et al., 2000, 2005; Draganski et al., 2004; Fleming et al., 2010; Scholz et al., 2009). Nevertheless, the high number of significant differences detected in the present study is most unlikely to be due solely to individual variability in neuroanatomy; this is because such variability would not only allow single subjects to be more distant from their control group but also inflate the standard variance estimated from the controls, with the two effects cancelling each other out. An alternative possibility is that the higher number of significant effects in the temporal and frontal regions is the result of less accurate spatial registration in these areas. Future studies could examine this by testing for a correlation between the number of detected significant differences across different regions and the spatially varying registration accuracy. A possible solution could be the use of more advances registration algorithms which use local image structure, as opposed to intensity information, to detect complex image relationships (Mellor and Brady,

2004, 2005). Again, it is most unlikely the high number of significant differences detected in the present study is due solely to artifacts introduced during the spatial registration of the images, since such artifacts would not only make single subjects more different from their control group but also inflate the standard variance estimated from the controls. The most likely possibility is that the high number of significant differences is related to the interaction of non-normality with random field theory (Salmond et al., 2002; Viviani et al., 2007). A solution could be the use of transformation strategies to achieve conformity to normality and minimize false positive rate, as demonstrated by Salmond et al. (2002) and Viviani et al. (2007) respectively; a logit transformation however requires probability values of gray matter maps to lie between 0 and 1 and therefore is not appropriate for modulated data which often have values above 1. Another solution could be the use of non-parametric analytical methods (Nichols and Holmes, 2002) which do not require the data to be normally distributed and have been employed successfully in the context of VBM (Bendfeldt et al., 2010). For instance, the statistical non-parametric mapping toolbox SnPM (SnPM, http://www.nitrc.org/projects/snpm/) and the FSL software (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/) could be used to analyze the data using permutation theory instead of random field theory. However, the use of standard permutation methods may not allow the comparison of a single case against a control group of 16 subjects, because the relatively small number of controls would not provide enough permutations for an accurate p-value to be computed.

Following previous studies (Salmond et al., 2002; Viviani et al., 2007), we also examined the impact of smoothing on the results. Consistent with our prediction, the chance of detecting a significant difference between a single subject and their control group decreases with a higher degree of smoothing; this replicates previous reports that smoothing is at least partially effective in reducing the impact of deviation from normality (Salmond et al., 2002; Viviani et al., 2007). In addition, there was a significant interaction between smoothing and the direction of the effect, with the number of significant increases becoming smaller with higher smoothing and the number of significant decreases showing the reverse effect. This unexpected aspect of our results suggests that non-normality of the data may have been responsible for a larger proportion of increases than decreases (because smoothing reduces the proportion of voxels which violate the assumption of normality as shown in Salmond et al., 2002).

We also found that the likelihood of detecting significant increases in a single subject relative to their control group was higher than that of detecting significant decreases. This was not in line with a previous study which found higher false positive rates for decreases than increases (Salmond et al., 2002). This apparent inconsistency could be explained by the fact that we employed the modulation step, which involves multiplying the spatially normalized gray matter by its relative volume before and after spatial normalization, and therefore our results were based on the absolute volume of gray matter (Mechelli et al., 2005), whereas Salmond et al. used unmodulated data and therefore their results were based on gray matter density i.e. the proportion of gray matter to all tissue types within a region. Consistent with this explanation, when we repeated the statistical analyses using unmodulated data, we found that the likelihood of detecting significant decreases in a single subject relative to their control group was indeed higher than that of detecting significant increases. In addition, the interaction between the use of modulation and the direction of the differences was significant, suggesting that the multiplication of the spatially normalized gray matter by its relative volume before and after spatial normalization may have an impact on the non-normality of the data.

A significant strength of the present study is that we repeated our analysis using two independent datasets acquired from Caucasian and Chinese participants in Cambridge and Beijing respectively. Overall the results were consistent, with no significant differences in the number of significant differences per subject between the two datasets. While this provides some support to the generalizability of our results to other research centers, it should be noted that both datasets were acquired using a 3 T MRI system. The extent to which these results apply to different field strengths is therefore unclear from the present data.

The results of the present investigation have important implications for future single case structural neuroimaging studies. More specifically, they provide an empirical estimation of the number of significant differences to be expected when comparing a disease-free individual against a group of age- and gender-matched controls. This information could inform the interpretation of future single case VBM studies of psychiatric or neurological patients. For example, the likelihood of detecting 1 cluster showing a significant decrease in gray matter volume in a single subject relative to their control group at $p<0.05$ (corrected) is 71% within the healthy population; thus it may be misconceived to interpret such result as a neuropathological correlate of the neurological or psychiatric disorder under investigation. In contrast, the likelihood of detecting 15 clusters showing a significant decreases in gray matter volume in a single subject relative to their control group at $p<0.05$ (corrected) is estimated to be 0% in the healthy population; such result would therefore be best attributed to the disorder under investigation. The results of our investigation could be particularly useful for the assessment of mental insanity in a forensic setting. Such assessment is made complicated by the fact that psychiatric symptoms can be easily faked or exaggerated and that most defendants assessed for mental insanity do not have a previous psychiatric history. One strategy for dealing with malingering of psychiatric disorders consists in validating the reported symptoms with an anatomo-clinical correlation (Rogers, 2008); the results of the present investigation could be used to inform the use of structural neuroimaging to perform such validation in a forensic context (Rigoni et al., 2010; Sartori et al., 2011).

The present study has a number of limitations. Firstly, our results are based on the assumption that all subjects were free from neurological or psychiatric disorders. While it is true that none of the participants had received a neurological or psychiatric diagnosis, we cannot exclude the possibility that some of them had developed undetected clinical symptoms associated with neuroanatomical alterations. This however is most unlikely to account for our results, since any undiagnosed illness would most likely concern only a fraction of our sample whereas we detected significant differences in the vast majority of subjects. Secondly, we used a control group of 16 subjects and therefore the empirical estimates reported in Tables 2–4 are specific to studies using this sample size. In contrast the sample size of previous single case VBM studies has been highly variable and has ranged from as few as 6 to as many as 140 subjects (Table 1). Thirdly, the present investigation focused on single case studies and therefore could not provide any information in false positive rates in group studies. Given a previous report of higher false positive rates in unbalanced than balanced designs (Salmond et al., 2002), it would be interesting to examine how false positive rates vary as a function of sample size in future studies. Fourthly, while subjects in the Beijing dataset were ethnically homogeneous, information about ethnicity was not available for the Cambridge dataset. However, it is highly unlikely that ethnic heterogeneity can account for our results given that these did not differ between the two datasets. A final limitation is that for practical reasons we have restricted our investigation to the use of FWE correction for multiple comparisons; this type of correction was preferred because it has been adopted in the vast majority of single case VBM studies (see Table 1). However it is important to acknowledge that there are other types of correction (e.g. False Discovery Rate) which might yield different patterns of results.

In summary, the present study provides evidence that, when comparing a single neurological or psychiatric patient against a group of controls with VBM, the chance of detecting a significant

difference which is not related to the disorder under investigation is much higher than previously expected. Interpretation of the results of single case studies should be particularly cautious in the case of significant differences in temporal and frontal lobes, where false positive rates appear to be highest. Ironically, these are also the areas of the brain which are typically affected in most psychiatric and neurological disorders, further complicating the interpretation of the results of single case studies in these clinical populations. Nevertheless, the empirical estimation of the number of significant differences in single case studies provided here could be used to inform the interpretation of future single case VBM studies of psychiatric or neurological patients. A final consideration is that the present study used two freely available datasets from the NITRC database. Another recent investigation also used datasets from the NITRC database to investigate the validity of parametric analysis of functional MRI data (Eklund et al., 2012). We believe that these two examples well illustrate the potential of sharing large datasets for accelerating research about the human brain.

Acknowledgments

The authors would like to thank Dr. Zang and Dr. Buckner for providing the data through the Neuroimaging Informatics Tools and Resources Clearinghouse.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2012.12.045.

References

Adlam, A.L.R., Patterson, K., Rogers, T.T., Nestor, P.J., Salmond, C.H., Acosta-Cabronero, J., Hodges, J.R., 2006. Semantic dementia and fluent primary progressive aphasia: two sides of the same coin? Brain 3066–3080.

Adolphs, R., 2009. The social brain: neural basis of social knowledge. Annu. Rev. Psychol. 60, 693–716.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38 (1), 95–113.

Ashburner, A., Friston, K., 2000. Voxel-based morphometry — the methods. Neuroimage 11, 805–821.

Ashburner, A., Friston, K., 2001. Why voxel-based morphometry should be used. Neuroimage 14, 1238–1243.

Ashburner, J., Friston, K.J., 2009. Computing average shaped tissue probability templates. Neuroimage 45 (2), 333–341.

Asplund, C.L., Todd, J.J., Snyder, A.P., Marois, R., 2010. A central role for the lateral prefrontal cortex in goal-direct and stimulus-driven attention. Nat. Neurosci. 13, 507–514.

Beeson, P.M., King, R.M., Bonakdarpour, B., Henry, M.L., Cho, H., Rapcsak, S.Z., 2011. Positive effects of language treatment for logopenic variant of primary progressive aphasia. J. Mol. Neurosci. 45, 724–736.

Bendfeldt, K., Blumhagen, J.O., Egger, H., Loetscher, P., Denier, N., Kuster, P., Traud, S., Mueller-Lenke, N., Naegelin, Y., Gass, A., Hirsch, J., Kappos, L., Nichols, T.E., Radue, E.W., Borgwardt, S.J., 2010. Spatiotemporal distribution pattern of white matter lesion volumes and their association with regional grey matter volume reductions in relapsing–remitting multiple sclerosis. Hum. Brain Mapp. 31, 1542–1555.

Bianchini, F., Incoccia, C., Palermo, L., Piccardi, L., Zompanti, L., Sabatini, U., Peran, P., Guariglia, C., 2010. Developmental topographical disorientation in a healthy subject. Neuropsychologia 48, 1563–1573.

Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kötter, R., Li, S.J., Lin, C.P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.A., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.F., Zhang, H.Y., Castellanos, F.X., Milham, M.P., 2010. Toward discovery science of human brain function. Proc. Natl. Acad. Sci. U. S. A. 107 (10), 4734–4739.

Bocková, M., Chládek, J., Jurák, P., Halámek, J., Rektor, I., 2007. Executive functions processed in the frontal and lateral temporal cortices: intracerebral study. Clin. Neurophysiol. 118 (12), 2625–2636.

Bora, E., Yucel, M., Allen, N.B., 2009. Neurobiology of human affiliative behaviour: implications for psychiatric disorders. Curr. Opin. Psychiatry 22 (3), 320–325.

Bora, E., Fornito, A., Yucel, M., Pantelis, C., 2010. Voxelwise meta-analysis of gray matter abnormalities in bipolar disorder. Biol. Psychiatry 67, 1097–1105.

Bora, E., Fornito, A., Radua, J., Walterfang, M., Seal, M., Wood, S.J., Yücel, M., Velakoulis, D., Pantelis, C., 2011. Extensive gray matter deficits in frontal, temporal, cingulate and insular cortex and thalamus, and increased gray matter in the basal ganglia. Neuroanatomical abnormalities in schizophrenia: a multimodal voxelwise meta-analysis and meta-regression analysis. Schizophr. Res. 127 (13), 46–57.

Bozzali, M., Cercignani, M., Baglio, F., Scotti, G., Farina, E., Pugnetti, L., Ashburner, J., Nemni, R., Falini, A., 2008. Voxel-wise analysis of diffusion tensor MRI improves the confidence of diagnosis of corticobasal de generation non-invasively. Parkinsonism Relat. Disord. 14, 436–439.

Brazdil, M., Mikl, M., Chlebus, P., Pazourkova, M., Novak, Z., Chrastina, J., Prasek, J., Bolcak, K., Fojtikova, D., Horky, J., Tomcik, J., Lzicarova, E., Rektor, I., 2006. Combining advanced neuroimaging techniques in presurgical workup of non-lesional intractable epilepsy. Epileptic Disord. 8 (3), 190–194.

Buccigrossi, R., Ellisman, M., Grethe, J., Haselgrove, C., Kennedy, D.N., Martone, M., Preuss, N., Reynolds, K., Sullivan, M., Turner, J., Wagner, K., 2008. The Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC). AMIA Annu Symp Proc., p. 1000.

Butler, T., Weisholtz, D., Isenberg, N., Harding, E., Epstein, J., Stern, E., Silbersweig, D., 2012. Neuroimaging of frontal-limbic dysfunction in schizophrenia and epilepsy-related psychosis: toward a convergent neurobiology. Epilepsy Behav. 23 (2), 113–122.

Carreiras, M., Seghier, M.L., Baquero, S., Estevez, A., Lozano, A., Devlin, J.T., Price, J.C., 2009. An anatomical signature for literacy. Nature 461, 983–986.

Casey, B.J., Giedd, J.N., Thomas, K.M., 2000. Structural and functional brain development and its relation to cognitive development. Biol. Psychol. 54, 241–257.

Casey, B.J., Tottenham, N., Liston, C., Durston, S., 2005. Imaging the developing brain: what have we learned about cognitive development? Trends Cogn. Sci. 9, 104–110.

Chen, S., Xia, W., Li, L., Li, J., He, Z., Zhang, Z., Yan, L., Zhang, J., Hu, D., 2006. Gray matter density reduction in the insula in fire survivors with posttraumatic stress disorder: a voxel based morphometric study. Psychiatry Res. 146, 65–72.

Cipolotti, L., Bird, C., Good, T., Macmanus, D., Rudge, P., Shallice, T., 2006. Recollection and familiarity in dense hippocampal amnesia: a case study. Neuropsychologia 44, 489–506.

Colliot, O., Bernasconi, N., Khalili, S.B., Antel, V., Naessens, V., Bernasconi, A., 2006. Individual voxel-based analysis of gray matter in focal cortical dysplasia. Neuroimage 29, 162–171.

De Olivera-Souz, R., Hare, R.D., Bramati, I.E., Garrido, G.J., Azevedo Ignacio, F., Tovar-Mill, F., Moll, J., 2008. Psychopathy as a disorder of the moral brain: fronto-temporo-limbic grey matter reductions demonstrated by voxel-based morphometry. Neuroimage 40, 1202–1213.

Dolan, R.J., 2008. Neuroimaging of cognition: past, present and future. Neuron 60, 496–502.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., May, A., 2004. Changes in grey matter induced by training. Nature 427, 311–312.

Du, M.Y., Wu, Q.Z., Yue, Q., Li, J., Liao, Y., Kuang, W.H., Huang, X.Q., Chan, R.C., Mechelli, A., Gong, Q.Y., 2012. Voxelwise meta-analysis of gray matter reduction in major depressive disorder. Prog. Neuropsychopharmacol. Biol. Psychiatry 36 (1), 11–16.

Eklund, A., Andersson, M., Josephson, C., Johannesson, M., Knutsson, M., 2012. Does parametric fMRI analysis with SPM yield valid results? — an empirical study of 1484 rest datasets. Neuroimage 61, 565–578.

Ellison-Wright, I., Bullmore, E., 2010. Anatomy of bipolar disorder and schizophrenia: a meta-analysis. Schizophr. Res. 117 (1), 1–12.

Epelbaum, S., Pinel, P., Gaillard, R., Delmaire, C., Perrin, M., Dupont, S., Dehaene, S., Cohen, L., 2008. Pure alexia as a disconnection syndrome: new diffusion imaging evidence for an old concept. Cortex 44, 962–974.

Eriguchi, Y., Yamasue, H., Doi, N., Nishida, T., Abe, O., Yamada, H., Aoki, S., Suga, M., Inoue, H., Nonaka, H., Obata, T., Ikehira, H., Kobayashi, K., Kasai, K., 2010. A case of adult onset type II citrullinemia with comorbid epilepsy even after liver transplantation. Epilepsia 51 (12), 2484–2487.

Feldmann, A., Trauninger, A., Toth, L., Kotek, G., Kosztolanyi, P., Illes, E., Pfund, Z., Komoly, S., Nagy, F., Illes, Z., 2008. Atrophy and decreased activation of frontoparietal attention areas contribute to higher visual dysfunction in posterior cortical atrophy. Psychiatry Res.: Neuroimaging 164, 178–184.

Felmingham, K., Williams, L.M., Whitford, T., Falconer, E., Kemp, A.H., Peduto, A., Bryant, R.A., 2009. Duration of posttraumatic stress disorder predicts hippocampal grey matter loss. Neuroreport 20, 1402–1406.

Fleming, S.M., Weil, R.S., Nagy, Z., Dolan, R.J., Rees, G., 2010. Relating introspective accuracy to individual differences in brain structure. Science 329, 1541–1543.

Fornito, A., Wood, S.J., Whittle, S., Fuller, J., Adamson, C., Saling, M.M., Velakoulis, D., Pantelis, C., Yucel, M., 2008. Morphometry of the medial frontal cortex: associations with cortical thickness, surface area, volume and sulcal depth. Hum. Brain Mapp. 29, 222–236.

Freudenmann, R.W., Kolle, M., Huwe, A., Luster, M., Reske, S.N., Huber, M., Lepping, P., Wolf, R.C., Schonfeldt-Lecuona, C., 2010. Delusional infestation: neural correlates and antipsychotic therapy investigated by multimodal neuroimaging. Prog. Neuropsychopharmacol. Biol. Psychiatry 34, 1215–1222.

Friston, K.J., 2012. Ten ironic rules for non-statistical reviewers. Neuroimage 61 (4), 1300–1310.

Friston, K.J., Holmes, A.P., Worsley, K.J., 1999. How many subjects constitute a study? Neuroimage 10, 1–5.

Gitelman, D.R., Ashburner, J., Friston, K.J., Tyler, L.K., Price, C.J., 2001. Voxel-based morphometry of herpes simplex encephalitis. Neuroimage 13, 623–631.

Gogtay, N., Giedd, J.N., Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., Nugent, T.F., Herman, D.H., Clasen, L.S., Toga, A.W., Rapoport, J.L., Thompson, P.M., 2004.

Dynamic mapping of human cortical development during childhood through early adulthood. Proc. Natl. Acad. Sci. U. S. A. 101 (21), 8174–8179.
Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N.A., Friston, K.J., Frackowiak, S.J., 2001. A voxel based morphometric study of ageing in 456 normal adult human brains. Neuroimage 14, 21–36.
Gorno-Tempini, M.L., Murray, R.C., Rankin, K., Weiner, M.W., Miller, B.L., 2004a. Clinical, cognitive and anatomical evolution from nonfluent progressive aphasia to corticobasal syndrome: a case report. Neurocase 10 (6), 426–436.
Gorno-Tempini, M.L., Rankin, K., Woolley, J.D., Rosen, H.J., Phengresamy, L., Miller, B.L., 2004b. Cognitive and behavioral profile in a case of right anterior temporal lobe neurodegeneration. Cortex 40, 631–644.
Gour, N., Ranjeva, J.P., Ceccaldi, M., Confort-Gouny, S., Barbeau, E., Soulier, E., Guye, M., Didic, M., Felician, O., 2011. Basal functional connectivity within the anterior temporal network is associated with performance on declarative memory tasks. Neuroimage 58 (2), 687–697.
Hill, J., Inder, T., Neil, J., Dierker, D., Harwell, J., Van Essen, D., 2010. Similar patterns of cortical expansion during human development and evolution. PNAS 107 (29), 13135–13140.
Huttenlocher, P.R., 1990. Morphometric study of human cerebral cortex development. Neuropsychologia 28 (6), 517–527.
Jimura, K., Konishi, S., Asari, T., Miyashita, Y., 2010. Temporal pole activity during understanding other persons' mental states correlates with neuroticism trait. Brain Res. 1328, 104–112.
Joubert, S., Felician, O., Barbeau, E., Ranjeva, J.P., Christophe, M., Didic, M., Poncet, M., Ceccaldi, M., 2006. The right temporal lobe variant of frontotemporal dementia. Cognitive and neuroanatomical profile of three patients. J. Neurol. 253, 1447–1458.
Klingner, C.M., Volk, G.F., Brodoehl, S., Burmeister, H.P., Witte, O.W., Guntinas-Lichius, O., 2012. Time course of cortical plasticity after facial nerve palsy: a single case study. Neurorehabil. Neural Repair 26 (2), 197–203.
Maguire, E.A., Kumaran, D., Hassabis, D., Kopelman, M.D., 2010. Autobiographical memory in semantic dementia: a longitudinal fMRI study. Neuropsychologia 48, 123–136.
Martin, E.I., Ressler, K.J., Binder, E., Nemeroff, C.B., 2010. The neurobiology of anxiety disorders: brain imaging, genetics, and psychoneuroendocrinology. Clin. Lab. Med. 30 (4), 865–891 (Review).
Mechelli, A., Price, C.J., Friston, K.J., Ashburner, J., 2005. Voxel based morphometry of the human brain: methods and applications. Curr. Med. Imaging Rev. 1, 105–113.
Mellor, M., Brady, M., 2004. Non-rigid multimodal image registration using local phase. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2004, pp. 789–796 (Saint-Malo; France) Springer.
Mellor, M., Brady, M., 2005. Phase mutual information as a similarity measure for registration. Med. Image Anal. 9 (4), 330–343.
Migliaccio, R., Agosta, F., Toba, M.N., Samri, D., Corlier, F., de Souza, L.C., Chupin, M., Sharman, M., Gorno-Tempini, M.L., Dubois, B., Filippi, M., Bartolomeo, P., 2011. Brain networks in posterior cortical atrophy: a single case tractography study and literature review. Cortex 48 (10), 1298–1309.
Migliaccio, R., Agosta, F., Scola, E., Magnani, G., Cappa, S.F., Pagani, E., Canu, E., Comi, G., Falini, A., Gorno-Tempini, M.L., Bartolomeo, P., Filippi, M., 2012. Ventral and dorsal visual streams in posterior cortical atrophy: A DT MRI study. Neurobiol. Aging 33 (11), 2572–2584.
Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202.
Modinos, G., Mechelli, A., Ormel, J., Groenewold, N.A., Aleman, A., McGuire, P., 2010. Schizotypy and brain structure: a voxel-based morphometry study. Psychol. Med. 40, 1423–1431.
Muhlau, M., Wohlschlager, A.M., Gaser, C., Valent, M., Weindl, A., Nunnemann, S., Peinemann, A., Etgen, T., Ilg, R., 2009. Voxel based morphometry in individual patients: a pilot study in early Huntington disease. AJNR Am. J. Neuroradiol. 30, 539–543.
Mummery, C.J., Patterson, K., Price, C.J., Ashburner, J., Frackowiak, R.S.J., Hodges, J.R., 2000. A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. Ann. Neurol. 47, 36–45.
Nanri, K., Shibuya, M., Taguchi, T., Hasegawa, A., Tanaka, N., 2011. Selective loss of Purkinje cells in a patient with anti-gliadin-antibody-positive autoimmune cerebellar ataxia. Diagn. Pathol. 6, 14–18.
Narvid, J., Gorno-Tempini, M.L., Slavotinek, A., DeArmond, S.J., Cha, Y.H., Miller, B.L., Rankin, K.P., 2009. Of brain and bone: the unusual case of Dr.A. Neurocase 15 (3), 190–205.
Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15 (1), 1–25.
Olson, I.R., Plotzker, A., Ezzyat, Y., 2007. The Enigmatic temporal pole: a review of findings on social and emotional processing. Brain 130 (Pt. 7), 1718–1731 (Review).
Oquendo, M.A., Parsey, R.V., 2007. What have we learned about the neurobiology of major depression? Am. J. Psychiatry 164 (4), 540–542.
Pruessner, J.C., Köhler, S., Crane, J., Pruessner, M., Lord, C., Byrne, A., Kabani, N., Collins, D.L., Evans, A.C., 2002. Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. Cereb. Cortex 12 (12), 1342–1353.
Rauch, S.L., 2000. NeuroImaging research and the neurobiology of obsessive compulsive disorder: where do we go from here? Biol. Psychiatry 47, 168–170.
Riddoch, M.J., Bodley Scott, S., Humphreys, G.W., 2010. No direction home: extinction is affected by implicit motion. Cortex 46, 678–684.
Rigoni, D., Pellgrini, S., Mariotti, V., Cozza, A., Mechelli, A., Ferrara, S.D., Pietrini, P., Sartori, G., 2010. How neuroscience and behavioral genetics improve psychiatric assessment: report on a violent murder case. Front. Behav. Neurosci. 4, 160.
Rogers, R., 2008. Clinical Assessment of Malingering and Deception, third edition. Guilford Press.
Rosen, H.J., Kramer, J.H., Gorno-Tempini, M.L., Schuff, N., Weiner, M., Miller, B.L., 2002. Patterns of cerebral atrophy in primary progressive aphasia. Am. J. Geriatr. Psychiatry 10, 89–97.
Rusch, N., van Elst, L.T., Ludaescher, P., Wilke, M., Huppertz, H.J., Thiel, T., Schmahl, C., Bohus, M., Lieb, K., Heblinger, B., Henning, J., Ebert, D., 2003. A voxel based morphometric MRI study in female patients with borderline personality disorders. Neuroimage 20, 385–392.
Salmond, C.H., Ashburner, J., Vargha-Khadem, F., Connelly, A., Gadian, D.G., Friston, K.J., 2002. Distributional assumption in voxel-based morphometry. Neuroimage 17, 1027–1030.
Salmond, C.H., de Haan, M., Friston, K.J., Gadian, D.G., Vargha-Khadem, F., 2003. Investigating individual differences in brain abnormalities in autism. Philos. Trans. R. Soc. Lond. 358, 405–413.
Sartori, G., Pellegrini, S., Mechelli, A., 2011. Forensic neurosciences: from basic research to applications and pitfalls. Curr. Opin. Neurol. 24 (4), 371–377 (Aug, Review).
Scholz, J., Klein, M.C., Behrens, T.E.J., Johansen-Berg, H., 2009. Training induces changes in white matter architecture. Nat. Neurosci. 12, 1370–1371.
Sehm, B., Frisch, S., Thone-Otto, A., Horstmann, A., Villringer, A., Obrig, H., 2011. Focal retrograde amnesia: voxel-based morphometry findings in a case without MRI lesions. PLoS One 6 (10), e26538.
Sherwood, C.C., Subiaul, F., Zawidzki, T.W., 2008. A natural history of the human mind: tracing evolutionary changes in brain and cognition. J. Anat. 212 (4), 426–454.
Shirtcliff, E.A., Vitacco, M.J., Graf, A.R., Gostisha, A.J., Merz, J.L., Zahn-Waxler, C., 2009. Neurobiology of empathy and callousness: implications for the development of antisocial behavior. Behav. Sci. Law 27 (2), 137–171.
Soares, J.C., 2003. Contribution from brain imaging to the elucidation of pathophysiology of bipolar disorders. Int. J. Neuropsychopharmacol. 6, 171–180.
Soloff, P., Nutche, J., Goradia, D., Diwadkar, V., 2008. Structural brain abnormalities in borderline personality disorder: a voxel based morphometry study. Psychiatry Res.: Neuroimaging 164, 223–236.
Spasojević, G.D., Malobabić, S., Suščević, D., Stijak, L., Nikolić, V., Gojković, I., 2011. Morphological variability of the subcallosal area of man. Surg. Radiol. Anat. 33 (4), 313–318.
Stewart, C.B., Disotell, T.R., 1998. Primate evolution—in and out of Africa. Curr. Biol. 8, R582–R588.
Sundram, F., Deely, Q., Sarkar, S., Daly, E., Latham, R., Craig, M., Raxzek, M., Fahy, T., Picchioni, M., Barker, G.J., Murphy, D.C.M., 2011. White matter microstructural abnormalities in the frontal lobe of adults with antisocial personality disorder. Cortex 48 (2), 216–229.
Suzuki, M., Nohara, S., Hagino, H., Takahashi, T., Kawasaki, Y., Yamashita, I., Watanabe, N., Seto, H., Kurachi, M., 2005. Prefrontal abnormalities in patients with simple schizophrenia: structural and functional brain-imaging studies in five cases. Psychiatry Res.: Neuroimaging 140, 157–171.
Takahashi, T., Yücel, M., Lorenzetti, V., Tanino, R., Whittle, S., Suzuki, M., Walterfang, M., Pantelis, C., Allen, N.B., 2010. Volumetric MRI study of the insular cortex in individuals with current and past major depression. J. Affect. Disord. 121 (3), 231–238.
Thompson, S.A., Graham, K.S., Williams, G., Patterson, K., Kapur, N., Hodges, J.R., 2004. Dissociating person-specific from general semantic knowledge: roles of the left and right temporal lobes. Neuropsychologia 42, 359–370.
Tramoni, E., Aubert-Khalfa, S., Guye, M., Ranjeva, J.P., Felician, O., Ceccaldi, M., 2009. Hypo-retrieval and hyper-suppression mechanisms in functional amnesia. Neuropsychologia 47, 611–624.
Valdes-Sosa, M., Bobes, M.A., Quinones, I., Garcia, L., Valdes-Hernandez, P.A., Iturria, Y., Melie-Garcia, L., Lopera, F., Asencio, J., 2011. Covert face recognition without the fusiform-temporal pathways. Neuroimage 57, 1162–1176.
Viviani, R., Beschoner, P., Ehrhard, K., Schmitz, B., Thöne, J., 2007. Non-normality and transformations of random fields, with an application to voxel-based morphometry. Neuroimage 35 (1), 121–130.
Watson, R.E., Desesso, J.M., Hurtt, M.E., Cappon, G.D., 2006. Postnatal growth and morphological development of the brain: a species comparison. Birth Defects Res. B Dev. Reprod. Toxicol. 77 (5), 471–484.
Wingenfeld, K., Spitzer, C., Rullkötter, N., Löwe, B., Wingenfeld, K., Spitzer, C., Rullkötter, N., Löwe, B., 2010. Borderline personality disorder: hypothalamus pituitary adrenal axis and findings from neuroimaging studies. Psychoneuroendocrinology 35 (1), 154–170 (Review).
Woermann, F.G., Free, S.L., Koepp, M.L., sisodiya, S.M., Duncan, J.S., 1999. Abnormal cerebral structure in juvenile myoclonic epilepsy demonstrated with voxel-based analysis of MRI. Brain 122, 2101–2107.
Yassa, M.A., Stark, C.E., 2009. A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. Neuroimage 44 (2), 319–327.
Zahn, R., Buechert, M., Overmans, J., Talazko, J., Specht, K., Ko, C.W., Thiel, T., Kaufmann, R., Dykierel, P., Juengling, F., Hull, M., 2005. Mapping of temporal and parietal cortex in progressive non fluent aphasia and Alzheimer's disease using chemical shift imaging, voxel-based morphometry and positron emission tomography. Psychiatry Res.: Neuroimaging 140, 115–131.
Zhang, J., Tan, Q., Yin, H., Zhang, X., Huan, Y., Tang, L., Wang, H., Xu, J., Li, L., 2011. Decreased gray matter volume in the left hippocampus and bilateral calcarine cortex in coal mine flood disaster survivors with recent onset PTSD. Psychiatry Res.: Neuroimaging 192, 84–90.