



Towards person-centered neuroimaging markers for resilience and vulnerability in Bipolar Disorder



Sophia Frangou^{a,*}, Danai Dima^{a,b,c}, Jigar Jogia^d

^a Department of Psychiatry, Icahn School of Medicine at Mount Sinai, USA

^b King's College London, UK

^c City University, UK

^d Aston University, UK

ARTICLE INFO

Article history:

Accepted 31 August 2016

ABSTRACT

Improved clinical care for Bipolar Disorder (BD) relies on the identification of diagnostic markers that can reliably detect disease-related signals in clinically heterogeneous populations. At the very least, diagnostic markers should be able to differentiate patients with BD from healthy individuals and from individuals at familial risk for BD who either remain well or develop other psychopathology, most commonly Major Depressive Disorder (MDD). These issues are particularly pertinent to the development of translational applications of neuroimaging as they represent challenges for which clinical observation alone is insufficient. We therefore applied pattern classification to task-based functional magnetic resonance imaging (fMRI) data of the n-back working memory task, to test their predictive value in differentiating patients with BD (n=30) from healthy individuals (n=30) and from patients' relatives who were either diagnosed with MDD (n=30) or were free of any personal lifetime history of psychopathology (n=30). Diagnostic stability in these groups was confirmed with 4-year prospective follow-up. Task-based activation patterns from the fMRI data were analyzed with Gaussian Process Classifiers (GPC), a machine learning approach to detecting multivariate patterns in neuroimaging datasets. Consistent significant classification results were only obtained using data from the 3-back versus 0-back contrast. Using contrast, patients with BD were correctly classified compared to unrelated healthy individuals with an accuracy of 83.5%, sensitivity of 84.6% and specificity of 92.3%. Classification accuracy, sensitivity and specificity when comparing patients with BD to their relatives with MDD, were respectively 73.1%, 53.9% and 94.5%. Classification accuracy, sensitivity and specificity when comparing patients with BD to their healthy relatives were respectively 81.8%, 72.7% and 90.9%. We show that significant individual classification can be achieved using whole brain pattern analysis of task-based working memory fMRI data. The high accuracy and specificity achieved by all three classifiers suggest that multivariate pattern recognition analyses can aid clinicians in the clinical care of BD in situations of true clinical uncertainty regarding the diagnosis and prognosis.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Bipolar disorder (BD) is an affective disorder that ranks amongst the leading causes of disability worldwide across all age groups (World Health Organization, 2008). This motivates efforts to characterize valid and reliable biological markers of disease expression in order to facilitate early identification and novel treatment discovery.

Magnetic resonance imaging (MRI) has been extensively used to investigate the neural correlates of disease expression in BD.

Structural MRI (sMRI) studies have demonstrated that BD is associated with reductions in whole brain and regional gray matter volumes (Hallahan et al., 2011; Kempton et al., 2011; Fears et al., 2014). Functional MRI (fMRI) studies have provided further information in terms of changes in regional blood-oxygen-level-dependent (BOLD) signal, most commonly in the domains of affect processing and executive control, where both genetically-derived and disease-related deficits have been reported (Glahn et al., 2010; Fears et al., 2014). The common network for affect processing notably involves the amygdala (AMG), ventral striatum and putamen and the ventral prefrontal (VPFC), ventral anterior cingulate (ACC) and insular cortices (Lindquist et al., 2012). The common network supporting executive control functions includes dorsal

* Correspondence to: Department of Psychiatry, Icahn School of Medicine at Mount Sinai, 1425 Madison Avenue, Box 1230, New York, NY 10029, USA.

E-mail address: sophia.frangou@mssm.edu (S. Frangou).

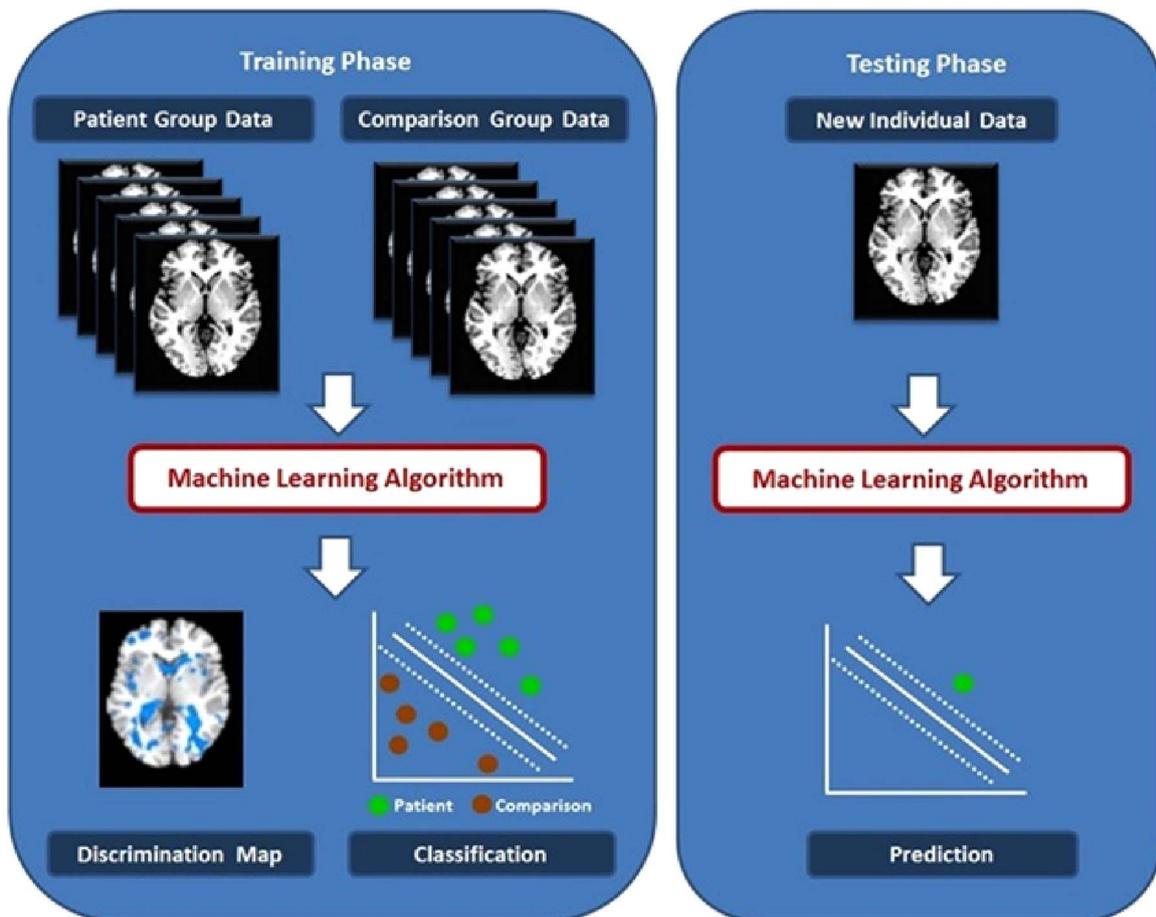


Fig. 1. Machine Learning Classification. This is a simplified illustration of the Gaussian Process Classifier functions that outlines the general principles of multivariate machine learning classification as applied to neuroimaging data from two hypothetical groups, referred to as patient and comparison group. During the training phase (left panel), the classifier is separately presented with multiple neuroimaging datasets of individuals that belong to one or the other group. The algorithm uses these data to assign a predictive weight to each voxel as more or less likely to be associated with one or the other group. The output of the classifier is a discrimination maps showing regions that have the most significant contribution to classification (left panel, bottom left) and values regarding the performance of the classifier based on their separating hyperplane (left panel, bottom right). During the test phase, a previously unseen dataset is presented to the algorithm and is classified based on its probability of belonging to either the patient or comparison group.

striatal structures as well as the dorsolateral prefrontal (DLPFC), dorsal ACC, and parietal (PAR) cortices (Niendam et al., 2012). In patients with BD, exaggerated activation during affective and executive tasks has been consistently observed in the AMG, insula, and ventral ACC coupled with reduced PFC engagement (Chen et al., 2011; Cusi et al., 2012; Delvecchio et al., 2012; Fusar-Poli et al., 2012; Jogia et al., 2012; Cremaschi et al., 2013; Dima et al., 2013). These observations have improved the characterization of the biological underpinnings of BD but have had limited clinical utility as they are based on group-level inferences that cannot be readily applied to the categorization of single individuals.

Advances in machine learning techniques, a field of artificial intelligence, represent a major development that could lead to clinical useful neuroimaging applications in psychiatry. Multivariate pattern recognition is a particular type of machine learning concerned with the discovery of regularities in data through the use of computer algorithms (Vapnik 1995). Frequently used machine learning approaches are support vector machines (SVM) and Gaussian Process Classifiers (GPC) (Fig. 1). Our group and others have shown that the application of multivariate pattern recognition analyses can reliably classify patients with BD from healthy individuals (Rocha-Rego et al., 2014; Mwangi et al., 2014; Schnack et al., 2014) and from patients with schizophrenia (Schnack et al., 2014) or Major Depressive Disorder (MDD) (Grotegerd et al., 2013; Grotegerd et al., 2014) with an accuracy of approximately 70-80%.

It can be argued however, that neuroimaging may be able to make a unique contribution in situations where clinical assessment and observation are not sufficient for diagnosis and prognosis. For clinicians, one of the greatest challenges lies in the differential diagnosis of BD from MDD. Although mania is the diagnostic hallmark of BD, in the majority of patients the disorder first presents with depressive symptoms (Forty et al., 2009). Even after disease onset, depressive symptoms dominate and contribute to morbidity and psychosocial disability (Judd et al., 2002). A substantial body of research has focused on identifying phenomenological features that could differentiate MDD from BD depression; the weight of evidence suggests that reliable and accurate differentiation at the level of the individual patient is beyond the resolution of even rigorous and detailed clinical assessment (Mitchell et al., 2008), particularly amongst those with a family history of BD (Mitchell et al., 2011). Consequently, many patients are misdiagnosed and treated as having MDD either because they present with depression at illness onset or because they generally under-report manic symptoms. This has important treatment implications as antidepressant treatment in BD may exaggerate mood instability (El-Mallakh et al., 2015). A further challenge lies in predicting the outcome of asymptomatic individuals with a family history of BD. As a group, these individuals are a higher risk than the general population for developing BD (Duffy et al., 2015; Fullerton et al., 2015).

Therefore the aim of the current study was to test whether neuroimaging can indeed assist clinicians when faced with true clinical uncertainty in situations where clinical acumen and observation are insufficient. We tackle two challenges; one focuses on the differential diagnosis of BD from MDD in the presence of family history of BD in the MDD patients. The other on the correct identification of individuals who have remained well but would be conventionally considered “at risk” based on having a first-degree relative with BD. We therefore test the hypothesis that GPCs of task-based fMRI data during the n-back working memory task can identify patterns of neural function that will prove useful in differentiating patients with BD from healthy unrelated individuals, and individuals at familial risk for BD who have either developed MDD or remained free of psychopathology. We focus on the functional neuroanatomy of the n-back task because it has been shown to be robust to variations in the paradigms used, scanner types and acquisition sequences (Owen et al., 2005; Dima et al., 2014) and has been reliably used to elicit disease-related abnormalities in patients with BD and their relatives (Fusar-Poli et al., 2012; Cremaschi et al., 2013).

Methods

The study sample comprised 120 demographically matched participants consisting of 30 patients with BD-type I, 30 of their first-degree relatives diagnosed with MDD, 30 psychiatrically healthy first-degree relatives and 30 unrelated healthy controls (Table 1). Participants were drawn from with the VIBES study sample (Frangou 2009; Kempton et al., 2009a, 2009b; Walterfang et al., 2009; Takahashi et al., 2010; Forcada et al., 2011; Lelli-Chiesa et al., 2011; Perrier et al., 2011; Pompei et al., 2011a, 2011b; Ruberto et al., 2011; Jogia et al., 2011, 2012a, 2012b; Dima et al., 2013; Rocha-Rego et al., 2014; Delvecchio et al., 2015). Participants with BD or MDD fulfilled the respective diagnostic criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders, 4th

Table 1
Demographic, clinical and task performance data.

	Unrelated healthy controls (n=30)	Healthy relatives (n=30)	MDD relatives (n=30)	BD patients (n=30)
Age (years)	33.4 (11.6)	35.3 (5.6)	32.9 (9.9)	34.7 (7.7)
Male: Female	15:15	14:16	16:14	15:15
Full scale IQ	108.4 (10.9)	110.5 (10.5)	109 (11.4)	107.1 (12.1)
Age of onset (years)	n/a	n/a	20.1 (9.1)	19.2 (10.8)
HDRS ^a	0.1 (0.5)	0.1 (0.4)	1.3 (0.9)	3.2 (1.1)
YMRS ^a	0.1 (0.8)	(0) (0)	0.1 (1.3)	1.3 (0.7)
BPRS ^a	24.4 (0.7)	24.5 (0.6)	25.4 (0.9)	27.1 (2.4)
3-back, % correct ^b	73.2 (12.4)	88.5 (14.3)	73.4 (17.2)	69.8 (16.7)
3-back, response time [sec] ^c	0.85 (0.3)	0.79 (0.3)	0.84 (0.5)	0.87 (0.6)

Except for sex, data are presented as mean (standard deviation). Bipolar disorder=BD; BPRS=Brief Psychiatric Rating Scale; IQ=Intelligence Quotient; HDRS=Hamilton depression Rating Scale; MDD=Major Depressive Disorder; YMRS=Young Mania Rating Scale; n/a=not applicable; there were no group differences in age, sex distribution, IQ and age of onset, $P > 0.7$;

^a Significant effect of group for HDRS, YMRS and BPRS $P < 0.001$; Post-hoc pairwise Bonferroni corrected comparisons showed that patients with BD more symptomatic than all other groups, $P < 0.01$.

^b Significant effect of group $F = 15.8$, $P < 0.01$; Post-hoc Bonferroni corrected pairwise comparisons showed healthy relatives outperformed all other groups, $P < 0.02$; Significant effect of group $F = 21.4$, $P < 0.01$; Post-hoc Bonferroni corrected pairwise comparisons showed healthy relatives outperformed all other groups, $P < 0.02$.

edition, revised (DSM-IV) (APA, 1994). Healthy relatives had no lifetime history of any major psychiatric disorder. Healthy unrelated controls had no family history or personal lifetime history of any major psychiatric disorder. All participants were free of any medical comorbidity and had no lifetime history of substance dependence or substance abuse in the six months leading to their brain scan. The diagnostic status of all participants was assessed using the Structured Clinical Interview for DSM-IV for Axis I diagnoses (First et al., 2002a, 2002b). Patients with BD and their relatives with MDD did not differ in the age of onset of their respective diagnosis. The diagnostic stability of patients with BD, their MDD and healthy relatives was confirmed through annual interview over a 4-year period following their brain scan. On the day of scanning, all participants were assessed using the Hamilton Depression Rating Scale (HDRS) (Hamilton, 1960), the Young Mania Rating Scale (YMRS) (Young et al., 1978), the Brief Psychiatric Rating Scale (BPRS) (Lukoff et al., 1986) and the Wechsler Adult Intelligence Scale 3rd Edition (Wechsler, 1997). To ensure patients with BD and relatives with MDD were in remission their psychopathology was assessed weekly over period of 1 month leading to their scan and at each assessment patients' scored below 7 in the HDRS and YMRS. Although the level of symptomatology was very low, patients with BD were more symptomatic than all other groups ($p < 0.001$). The BPRS, HDRS and YMRS scores were highly correlated (all $r > 0.78$, all $p < 0.001$). BD patients were medicated at the time of scanning with atypical antipsychotics ($n = 21$), antiepileptics ($n = 8$), lithium ($n = 14$), as monotherapy ($n = 18$) or combination therapy ($n = 12$). Three relatives with MDD were on selective serotonin re-uptake inhibitors. All medicated participants had remained on the same type and dose of medication for a minimum of 6 months prior to scanning.

Informed consent was obtained from all participants. The study was approved by the institutional ethics review board.

Working memory functional imaging task

The n-back task was employed in a block design incorporating alternating experimental and sensorimotor control conditions. A series of letters in yellow font were displayed on a blue screen for two seconds each. Participants were instructed to indicate by a button press whether the letter currently displayed matched the letter from the preceding n trials. In the sensorimotor control (0-back) the letter “X” was the designated target. In the experimental conditions (1, 2, 3-back) the target letter was defined as any letter that was identical to the one presented in the preceding one, two, or three trials. There were 18 epochs in all, each lasting 30 s, comprising 14 letters with a ratio of target to non-target letters ranging from 2:12 to 4:10 per epoch. The entire experiment lasted 9 min and included a total of 49 target and 203 non-target stimuli. To avoid any systematic order effects the conditions were pseudo-randomized. Performance was evaluated in terms of reaction time to target letters and accuracy (% correct responses). The task was explained to participants prior to scanning but there was no training.

Image acquisition

Gradient echo planar magnetic resonance (MR) images were acquired using a 1.5-T GE Neuro-optimized Signa MR system (General Electric, Milwaukee, WI, USA) fitted with 40 mT/m high-speed gradients. Foam padding and a forehead strap were used to limit head motion. A quadrature birdcage head coil was used for radio frequency (RF) transmission and reception. A total of 180 T2*-weighted MR brain volumes depicting BOLD contrast were acquired at each of 36 near-axial planes parallel to the intercommissural (AC-PC) plane; repetition time (TR)=3000 ms, echo

time (TE)=40 ms, slice thickness=3 mm, voxel dimensions=3.75 × 3.75 × 3.30 mm, interslice gap=0.3 mm, matrix size=64*64, flip angle=90°. Prior to each acquisition sequence, four dummy data acquisition scans were performed to allow the scanner to reach a steady state in T1 contrast. During the same session, a high-resolution T1-weighted structural image was acquired in the axial plane (inversion recovery prepared, spoiled gradient-echo sequence; TR=18 ms, TE=5.1 ms, TI=450 ms, slice thickness=1.5 mm, voxel dimensions=0.9375 × 0.9375 × 1.5 mm, matrix size 256*192, field of view=240 × 180 mm, flip angle=20°, number of excitations=1) for subsequent co-registration.

Image processing

Conventional fMRI analyses were implemented using Statistical Parametric Mapping (SPM8) (www.fil.ion.ucl.ac.uk/spm/software/spm8/). fMRI images were realigned, normalized and smoothed using an 8 mm full-width-half-maximum Gaussian kernel. To ensure data quality, task registration parameters were extracted and were used to identify participants with excessive interscan motion (defined as >4 mm translation, >4° rotation) and to conduct group comparisons. No subjects were excluded and there were no significant group differences in motion. The smoothed single-participant images were analyzed via multiple regressions using the linear convolution model, with vectors of onset representing the memory load conditions (1, 2, and 3 back) and the 0-back condition as sensorimotor control. Six movement parameters were also entered as nuisance covariates. Serial correlations were removed using an AR(1) model. A high pass filter (128 s) was applied to remove low-frequency noise. Contrast images of each memory load condition (1-, 2-, 3- back) versus 0-back were produced for each participant. At the individual subject level, a standard general linear modeling (GLM) approach was used to obtain estimates of the response size (beta) in each the memory load condition against the 0-back control condition.

Conventional fMRI analysis

We examined the effect of group (patients with BD, MDD relatives, healthy relatives and unrelated healthy controls) on BOLD signal using a standard general linear modeling (GLM) in SPM8 separately for each memory load condition (1-, 2-, 3- back) against the 0-back control. We examined each condition separately, instead of modeling activation changes with increasing load. This is because it has long been known that cortical loci, particularly in prefrontal regions, show non-linear changes in activation with increasing memory load (Callicott et al., 1999) which could increase voxel level variability. Suprathreshold clusters were identified using family wise error (FWE) voxel-wise correction of $p < 0.05$. Stereotactic coordinates were converted from MINI spatial array to that of Talairach and Tournoux (www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html) and corresponding anatomical and Brodmann area (BA) labels were identified with The Talairach Daemon Client (www.talairach.org). Measures of brain activation (weighted parameter estimates) from each subject were extracted using the MarsBaR toolbox (marsbar.sourceforge.net) from regions of interest (ROIs) defined on the basis of the whole brain analysis as 5 mm radius spheres at peak height coordinates within each suprathreshold cluster. These measures were also used to examine the role of potentially confounding variables of age, IQ, age of onset, task performance and medication. Although we report uncorrected P values, due to the large number of correlations undertaken the level of significant inference was set at a conservative P value of 0.005.

Multivariate pattern classification

Probability of group membership was determined using Gaussian Process Classifiers (GPCs) implemented in the Pattern Recognition for Neuroimaging Toolbox (PRoNT) (www.mlnl.cs.ucl.ac.uk/pronto/) using whole-brain individual beta maps/GLM coefficients for the contrasts of 1-back, 2-back and 3-back versus 0-back. Technical descriptions of GPC inference have been presented elsewhere (Schrouff et al., 2013). Briefly, the classifier is first trained to determine a predictive distribution that best distinguishes between two groups (e.g., case and controls); any parameters controlling the behavior of this distribution are computed by maximizing the logarithm of the marginal likelihood on the training data only. Then in the test phase, the classifier predicts the group membership of a previously unseen example. This is achieved by integrating over the predictive distribution for the test case and passing the output through a sigmoidal function, resulting in predictive probabilities scaled between 0 and 1 which precisely quantify the predictive uncertainty of the classifier for the test case.

We focus on the usefulness of fMRI data derived from the n-back task in differentiating patients with BD from (a) healthy individuals, (b) their relatives with MDD, and (c) from their psychiatrically healthy relatives. Each classifier was trained a leave-two-out cross-validation. For each cross-validation iteration, a matched pair of subjects from each group was excluded first and then the data were partitioned into training and test sets. For each trial, we thresholded the probabilistic predictions at 0.5 to convert the probabilistic predictions to class labels allowing the sensitivity and specificity of classification to be computed over all trials (Rasmussen and Williams, 2006). Statistical significance of each classifier was determined by permutation testing, as described previously (Marquand et al., 2010). Briefly, permutation testing was performed by repeatedly retraining the classifier after permuting the class labels (1000 permutations). A P-value for classification accuracy was computed by counting the number of permutations for which the permuted accuracy was equal or greater than the true accuracy (obtained with non-permuted labels), then dividing by 1000. In addition for each classifier, Pearson correlation analyses were carried out between GPC predictive probability and total BPRS scores, age, IQ, task performance, age of onset and medication dose.

As a secondary outcome we generated an unthresholded GPC weight map for each classifier. As GPC classifiers are multivariate, these discrimination maps do not describe focal activation effects but instead they represent the spatially distributed pattern of coefficients that quantify the contribution of each voxel to the classifier's decision function. We then estimated the positive (PPV) and negative (NPV) predictive value of each classifier.

Results

Conventional fMRI analysis

There were no group differences in the 1-back or 2-back versus 0-back contrasts. In the 3-back versus 0-back contrast a main effect of group was found ($p = 0.05$ FWE voxel level corrected) in the middle frontal gyrus (BA10) ($x = 36$ $y = 58$ $z = 0$, F score=6.64, cluster size 240), the superior temporal gyrus (BA22) ($x = 58$ $y = -44$ $z = 12$, F score=6.54, cluster size 169) and the posterior cingulate cortex (BA30) ($x = -28$ $y = -68$ $z = 8$, F score=6.97, cluster size 76). Patients with BD had reduced mean level of activation in the middle frontal gyrus compared to healthy relatives and unrelated controls ($p < 0.02$); however relatives with MDD did not differ compared to any other group and had an intermediate

level of mean activation between healthy relatives/ unrelated controls and BD patients. In the superior temporal gyrus, BD patients showed a greater mean level of activation compared to healthy controls and psychiatrically well relatives ($p < 0.05$); however, relatives with MDD did not differ compared to any group and had an intermediate level of mean activation between healthy relatives/controls and BD patients. In the posterior cingulate cortex, the mean level of activation was reduced in relatives with MDD compared to all the other groups ($p < 0.04$); all other pairwise comparisons were not significant.

We examined the role of potential confounders in several ways. We used three multivariate analyses of variance with medication status (antiepileptics vs. not on antiepileptic, lithium vs. not on lithium and antipsychotic vs. not on antipsychotics) as the independent factors and levels of activation in suprathreshold clusters in patients with BD as dependent variables. No group differences were found (all $P > 0.28$). We examined correlations between activation in suprathreshold regions in patients with BD and dose of lithium and antipsychotics on the day of scanning. None were significant (all $r < 0.21$, $P > 0.28$). Similarly, we did not observe any correlation with age of onset in patients and relatives with MDD (all $r < 0.23$, $P > 0.07$). Further analyses in the entire sample did not reveal significant correlations between the level of activation in any suprathreshold cluster and age (all $r < 0.10$, $P > 0.27$), IQ (all $r < 0.08$, $P > 0.38$), accuracy (all $r < 0.06$, $P > 0.51$) and response time (all $r < 0.16$, $P > 0.08$). None of these correlations survive correction for multiple testing.

Multivariate classification

Patients with BD versus unrelated healthy controls: Classifiers using the 1-back or 2-back versus 0-back contrast did not lead to a

statistically significant differentiation between patients with BD and healthy controls. Classification based on the 1-back versus 0-back contrast resulted in 65.5% accuracy, 61.5% sensitivity (true positives for BD) and 69.2% specificity (true negatives for unrelated controls) and was not significant ($p = 0.01$). Classification based on the 2-back versus 0-back contrast resulted in 73.1% accuracy, 76.9% sensitivity and 69.3% specificity and was also not significant ($p = 0.08$). In contrast, the classifier based on the 3-back versus 0-back contrast significantly differentiated patients from unrelated controls with an accuracy of 83.5% ($p = 0.001$). The sensitivity of the classification was 84.6% and the specificity 92.3%. The PPV (probability that individuals classified as BD patients were correctly identified) and NPV (probability that individuals classified as healthy controls were correctly identified) values were respectively 0.91 and 0.85. The unthresholded discrimination map showing the global spatial pattern by which the two groups differ based on the 3-back versus 0-back classifier is displayed in Fig. 2A; the largest clusters discriminating patients with BD from unrelated controls were located in the left inferior/middle/superior frontal gyrus and in the superior parietal lobule.

Patients with BD versus relatives with MDD: Significant results were found for classifiers based on contrast images from the 1-back ($p = 0.003$), 2-back ($p = 0.001$) and 3-back ($p = 0.001$) conditions. The classifier based on the 1-back versus 0-back contrast resulted in 76.9% accuracy, 53.9% sensitivity (true positives for patients with BD), 100% specificity (true negatives for relatives with MDD), PPV of 1 and NPV of 0.68. The classifier based on the 2-back versus 0-back contrast resulted in 73% accuracy, 61.5% sensitivity, 84.6% specificity, 0.79 PPV and 0.68 NPV. The classifier based on the 3-back versus 0-back contrast resulted in 73.1% accuracy, 53.9% sensitivity, 94.5% specificity, 0.90 PPV and 0.67 NPV. To maintain consistency with the results of the other classification

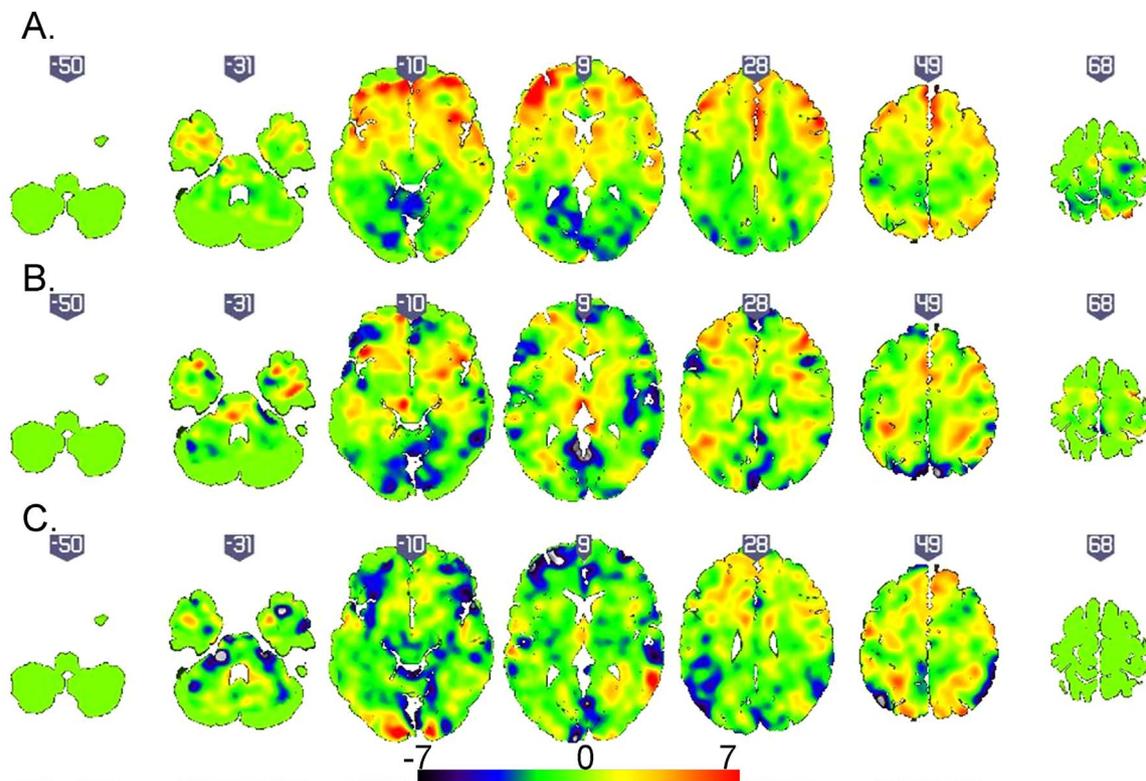


Fig. 2. Unthresholded discrimination maps for the classifier based on the 3-back > 0-back contrast: (A). Patients with Bipolar Disorder (BD) versus unrelated healthy individuals. In the corresponding discrimination map, positive coefficients (red) indicate clusters with predictive value for BD (B). Patients with BD versus relatives with Major Depressive Disorder. In the corresponding discrimination map, positive coefficients (red) indicate clusters with predictive value for BD (C). Patients with BD versus healthy relatives. In the corresponding discrimination map, positive coefficients (red) indicate clusters with predictive value for psychiatrically healthy relatives.

problems, the unthresholded discrimination map showing the global spatial pattern by which the two groups differ based on the 3-back versus 0-back classifier is displayed in Fig. 2B; the largest discriminating clusters were located in the left superior frontal gyrus, right middle frontal gyrus, bilaterally in the middle /superior frontal gyrus and the right temporal lobe.

Patients with BD versus healthy relatives: Classifiers using the 1-back or 2-back versus 0-back contrast did not differentiate between patients with BD and healthy relatives. Classification based on the 1-back versus 0-back contrast resulted in 54.5% accuracy, 36.6% sensitivity (true positives for patients with BD) and 72.3% specificity (true negatives for healthy relatives) but the results were not significant ($p=0.41$). Classification based on the 2-back versus 0-back contrast resulted in 59.1% accuracy, 36.3% sensitivity and 81.2% specificity and was also not significant ($p=0.17$). In contrast, the classifier based on the 3-back versus 0-back contrast was significant ($p=0.004$) and had 81.8% accuracy, 72.7% sensitivity, 90.9% specificity (true for healthy relative), 0.88 PPV and NPV 0.76. The unthresholded discrimination map showing the global spatial pattern by which the two groups differ based on the 3-back versus 0-back classifier is displayed in Fig. 2C; the largest discriminating clusters were located in the lingual gyrus and the cerebellum on the left.

For each classifier, Pearson correlation analyses between GPC predictive probabilities and age, IQ, task accuracy and reaction time, age of onset, BPRS total score and medication dose were not significant ($r < 0.28$; $p > 0.14$).

Discussion

We demonstrate the potential translational utility of task-based fMRI in aiding diagnosis and prognosis in BD. The approach highlighted here provides proof of concept for the development of new tools for the categorization of individuals where there is general agreement that clinical data alone are insufficient.

In this study we show that patients diagnosed with BD, show sufficient consistency in their neurofunctional patterns for them to be reliably differentiated from healthy individuals with an accuracy of 83.4%. Using structural MRI data alone, we have previously achieved classification accuracies of 69–78% in differentiating patients with BD from healthy individuals based on whole-brain gray or white matter classifiers (Rocha-Rego et al., 2014). Others who have also used structural MRI data have reported similar or lower accuracies (Schnack et al., 2014). It would therefore appear that task-based fMRI data may improve classification accuracy in BD as they may be more sensitive to disease-related pathology. Additionally, it is encouraging that the clusters that contribute to the correct classification of patients compared to controls, show biological plausibility. The conventional fMRI analysis of this same dataset, as well as results from independent samples (Cremaschi et al., 2013; Fusar-Poli et al., 2012), have consistently identified decreased prefrontal activation in patients with BD in the 3-back vs 0-back contrast compared to healthy controls. This is consistent with the clusters contributing to successful classification of patients from controls which implicate the prefrontal cortex, particularly lateral and frontopolar regions, and the dorsal parietal cortex, both key functional nodes of the working memory network (Niendam et al., 2012). Nevertheless, it could be argued that differentiating patients with established BD from healthy controls is of marginal clinical interest as real-life diagnostic assessments deal with more complex issues than separating people that have been ill for some time from those who are not ill at all (Wolfers et al., 2015). The value of testing classifiers for patients versus controls based on different imaging modalities is threefold. First, it serves to identify the type of data, or combinations of data, derived

from neuroimaging applications that would be adapted for clinical practice. Once this is achieved, one could test the best-performing classifiers on different clinical groups that are prodromal or have atypical or comorbid presentations. Second, examining the relationship between the multivariate neuroimaging signature and the clinical or demographic features of a disorder can lead to mechanistic insights regarding etiology and progression. Third, the availability of objective brain imaging tests can improve the social perception of psychiatry. At the level of individual patients, objective tests would provide re-assurance that clinical assessment and diagnostic assignment are not purely based on subjective judgment. At the societal level, it would help integrate psychiatry into mainstream medicine.

To our knowledge this is the first study to employ neuroimaging to address core issues of clinical uncertainty. We provide preliminary evidence for using neuroimaging classifiers to differentiate MDD from BD, especially in individuals who have a family history of BD. There are several clinical indicators of increased likelihood of conversion to BD in those who present with depression and have a family history of the disorder (Mitchell et al., 2008; 2011). However, none can be reliably applied to individual patients. In contrast, GPC classifiers differentiated patients with BD from relatives with MDD with an accuracy ranging from 73 to 77%. The sensitivity of these classifiers were low (53–62%). This is not surprising given the phenomenological overlap between BD and MDD and the dominance of depressive psychopathology in both disorders (Judd et al., 2002; Forty et al., 2009). What is more important however is that the classifier had very high specificity ranging from 84 to 100%. In principle this means that at the level of the individual patient clinicians would be able to exclude the possibility of BD, with a very high level of confidence, after a 10-min brain scan. This finding requires replication in different samples and settings and in more diverse clinical populations in terms of their age and duration of illness. Nevertheless, our results suggest that this is an avenue of research worth pursuing in demonstrating the translational value of neuroimaging.

Our third classifier, based on the 3-back versus 0-back contrast, differentiated patients with BD from their psychiatrically healthy relatives with an accuracy of 81.8%. This classifier identified high-risk individuals unlikely to convert to BD with 90.9% specificity. However, its sensitivity was 72.7% which suggest that some high-risk individual likely to convert may be missed. These results are very encouraging and could potentially inform early intervention services, where positive family history is a key criterion of risk and possible service inclusion (Duffy et al., 2015). There are a number of behavioral indicators of increased likelihood of conversion to syndromal BD but longitudinal studies suggest that they have low predictive value as a significant number of high-risk individuals never convert (Tijssen et al., 2010). Correctly identifying those who are in need of treatment is of paramount significance both in terms of cost-effectiveness of early intervention services but also in terms of preventing unnecessary treatment, concern and self-stigmatization in those unlikely to convert to BD. Of the clusters contributing to the correct identification of healthy relative, one was in the lingual gyrus, an occipital region involved in visual processing, and another in the cerebellum. Intriguingly, the lingual gyrus is closely related to cognitive function particularly with increasing age in BD (Fears et al., 2014) and shows dysfunctional connectivity during a range of tasks in patients but not their unaffected relatives (Dima et al., 2013). The volume of the cerebellum has been shown to be increased in healthy relatives of patients (Kempton et al., 2009a, 2009b) compared to controls while in BD patients it is known to progressively decrease (Moorhead et al., 2007). Therefore clusters contributing to the correct categorization of healthy relatives may point us to regions that are particularly important for maintaining resilience when preserved or for

disease expression when abnormal.

Patients with BD differed from the other groups (MDD relatives, healthy relatives, unrelated healthy individuals) in medication status. This raises the issue of whether classification might have been based on medication rather than diagnostic status. We cannot address this issue directly but we note that correlations between medication variables and GPC predictive probabilities were low and not statistically significant. We have previously shown that medication status did not have a significant influence on classifier performance when using brain structural data to classify patients with BD from controls (Rocha-Rego et al., 2014). Patients with BD were on a variety of psychotropic agents with different mechanisms of action and it is therefore unlikely that the neural correlates of medication effects would be consistent across all participants. This mitigates the possibility that the classifiers could have identified uniform medication-related classification rules. In addition, the very low levels of psychopathology in patients suggest that it is unlikely that classification rules were derived from neural patterns associated with symptomatic expression on the day of the scan. Further investigations are, however, needed to confirm the reproducibility of our findings in more diverse and larger samples.

In summary, this is the first study to our knowledge that has tested the contribution of neuroimaging to problems that arise in the clinical care of BD for which adequate non-imaging solutions have been elusive. Although in need to replication and refinement, our data provide clear direction for the development of translational imaging applications in psychiatry.

Conflict of interest

No conflict of interest.

Acknowledgments

Dr Sophia Frangou is partially supported by R01 MH104284-01A1 from National Institute of Mental Health (NIMH).

References

- American Psychiatric Association, 1994. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition. American Psychiatric Association, Washington, DC.
- Callicott, J.H., Mattay, V.S., Bertolino, A., Finn, K., Coppola, R., Frank, J.A., Goldberg, T. E., Weinberger, D.R., 1999. Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cereb. Cortex* 9 (1), 20–26.
- Chen, C.H., Suckling, J., Lennox, B.R., Ooi, C., Bullmore, E.T., 2011. A quantitative meta-analysis of fMRI studies in bipolar disorder. *Bipolar Disord.* 13 (1), 1–15.
- Cremaschi, L., Penzo, B., Palazzo, M., Dobreca, C., Cristoffanini, M., Dell'Osso, B., Altamura, A.C., 2013. Assessing working memory via N-back task in euthymic bipolar I disorder patients: a review of functional magnetic resonance imaging studies. *Neuropsychobiology* 68 (2), 63–70.
- Cusi, A.M., Nazarov, A., Holshausen, K., MacQueen, G.M., McKinnon, M.C., 2012. Systematic review of the neural basis of social cognition in patients with mood disorders. *J. Psychiatry Neurosci.* 37 (3), 154–169.
- Delvecchio, G., Fossati, P., Boyer, P., Brambilla, P., Falkai, P., Gruber, O., Hietala, J., Lawrie, S.M., Martinot, J.L., McIntosh, A.M., Meisenzahl, E., Frangou, S., 2012. Common and distinct neural correlates of emotional processing in bipolar disorder and major depressive disorder: a voxel-based meta-analysis of functional magnetic resonance imaging studies. *Eur. Neuropsychopharmacol.* 22 (2), 100–113.
- Dima, D., Jogia, J., Frangou, S., 2014. Dynamic causal modeling of load-dependent modulation of effective connectivity within the verbal working memory network. *Hum. Brain Mapp.* 35 (7), 3025–3035.
- Dima, D., Jogia, J., Collier, D., Vassos, E., Burdick, K.E., Frangou, S., 2013. Independent modulation of engagement and connectivity of the facial network during affect processing by CACNA1C and ANK3 risk genes for bipolar disorder. *JAMA Psychiatry* 70 (12), 1303–1311.
- Duffy, A., Jones, S., Goodday, S., Bentall, R., 2015. Candidate risks indicators for bipolar disorder: early intervention opportunities in high-risk youth. *Int. J. Neuropsychopharmacol.* <http://dx.doi.org/10.1093/ijnp/pyv071>
- El-Mallakh, R.S., Vöhringer, P.A., Ostacher, M.M., Baldassano, C.F., Holtzman, N.S., Whitham, E.A., Thommi, S.B., Goodwin, F.K., Ghaemi, S.N., 2015. Anti-depressants worsen rapid-cycling course in bipolar depression: a STEP-BD randomized clinical trial. *J. Affect. Disord.* 184, 318–321.
- Fears, S.C., Service, S.K., Kremeyer, B., Araya, C., Araya, X., Bejarano, J., Ramirez, M., Castrillón, G., Gomez-Franco, J., Lopez, M.C., Montoya, G., Montoya, P., Aldana, I., Teshiba, T.M., Abaryan, Z., Al-Sharif, N.B., Ericson, M., Jalbrzikowski, M., Luyckx, J., Navarro, L., Tishler, T.A., Altshuler, L., Bartzokis, G., Escobar, J., Glahn, D.C., Ospina-Duque, J., Risch, N., Ruiz-Linares, A., Thompson, P.M., Cantor, R.M., Lopez-Jaramillo, C., Macaya, G., Molina, J., Reus, V.I., Sabatti, C., Freimer, N.B., Bearden, C.E., 2014. Multisystem component phenotypes of bipolar disorder for genetic investigations of extended pedigrees. *JAMA Psychiatry* 71 (14), 375–387.
- First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 2002a. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Patient Edition (SCID-I/P)*. New York Biometrics Research, New York, Research Version.
- First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 2002b. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Non-Patient Edition (SCID-I/NP)*. New York Biometrics Research, New York, Research Version.
- Forcada, I., Papachristou, E., Mur, M., Christodoulou, T., Jogia, J., Reichenberg, A., Vieta, E., Frangou, S., 2011. The impact of general intellectual ability and white matter volume on the functional outcome of patients with Bipolar Disorder and their relatives. *J. Affect. Disord.* 130 (3), 413–420.
- Forty, L., Jones, L., Jones, L., Smith, D.J., Caesar, S., Fraser, C., Gordon-Smith, K., Hyde, S., Craddock, N., 2009. Polarity at illness onset in bipolar I disorder and clinical course of illness. *Bipolar Disord.* 11 (1), 82–88.
- Frangou, S., 2009. Risk and resilience in bipolar disorder: rationale and design of the Vulnerability to Bipolar Disorders Study (VIBES). *Biochem. Soc. Trans.* 37 (Pt 5), 1085–1089.
- Fullerton, J.M., Koller, D.L., Edenberg, H.J., Foroud, T., Liu, H., Glowinski, A.L., McInnis, M.G., Wilcox, H.C., Frankland, A., Roberts, G., Schofield, P.R., Mitchell, P. B., Nurnberger, J.I., Bipolar High Risk Study Group, BiGS Consortium, 2015. Assessment of first and second degree relatives of individuals with bipolar disorder shows increased genetic risk scores in both affected relatives and young at-risk individuals. *Am. J. Med.* <http://dx.doi.org/10.1002/ajmg.b.32344>
- Fusar-Poli, P., Howes, O., Bechdolf, A., Borgwardt, S., 2012. Mapping vulnerability to bipolar disorder: a systematic review and meta-analysis of neuroimaging studies. *J. Psychiatry Neurosci.* 37 (3), 170–184.
- Glahn, D.C., Almsy, L., Barguil, M., Hare, E., Peralta, J.M., Jr, Kent, Dassori, A., Contreras, J., Pacheco, A., Lanzagorta, N., Nicolini, H., Raventós, H., Escamilla, M. A., 2010. Neurocognitive endophenotypes for bipolar disorder identified in multiplex multigenerational families. *Arch. Gen. Psychiatry* 67 (2), 168–177.
- Grotegerd, D., Suslow, T., Bauer, J., Ohrmann, P., Arolt, V., Stuhmann, A., Heindel, W., Kugel, H., Dannlowski, U., 2013. Discriminating unipolar and bipolar depression by means of fMRI and pattern classification: a pilot study. *Eur. Arch. Psychiatry Clin. Neurosci.* 263 (2), 119–131.
- Grotegerd, D., Stuhmann, A., Kugel, H., Schmidt, S., Redlich, R., Zwanzger, P., Rauch, A.V., Heindel, W., Zwitterlood, P., Arolt, V., Suslow, T., Dannlowski, U., 2014. Amygdala excitability to subliminally presented emotional faces distinguishes unipolar and bipolar depression: an fMRI and pattern classification study. *Hum. Brain Mapp.* 35 (7), 2995–3007.
- 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 mega-analysis of individual adult patient data. *Biol. Psychiatry* 69 (4), 326–335.
- Hamilton, M., 1960. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* 23, 56–62.
- Jogia, J., Dima, D., Frangou, S., 2012b. Sex differences in bipolar disorder: a review of neuroimaging findings and new evidence. *Bipol Disord.* 14 (4), 461–471.
- Jogia, J., Dima, D., Kumari, V., Frangou, S., 2012a. Frontopolar cortical inefficiency may underpin reward and working memory dysfunction in bipolar disorder. *World J. Biol. Psychiatry* 13 (8), 605–615.
- Jogia, J., Ruberto, G., Lelli-Chiesa, G., Vassos, E., Maierü, M., Tatarelli, R., Girardi, P., Collier, D., Frangou, S., 2011. The impact of the CACNA1C gene polymorphism on frontolimbic function in bipolar disorder. *Mol. Psychiatry* 16 (11), 1070–1071.
- Judd, L.L., Akiskal, H.S., Schettler, P.J., Endicott, J., Maser, J., Solomon, D.A., Leon, A.C., Rice, J.A., Keller, M.B., 2002. The long-term natural history of the weekly symptomatic status of bipolar I disorder. *Arch. Gen. Psychiatry* 59 (6), 530–537.
- Kempton, M.J., Haldane, M., Jogia, J., Grasby, P.M., Collier, D., Frangou, S., 2009a. Dissociable brain structural changes associated with predisposition, resilience, and disease expression in bipolar disorder. *J. Neurosci.* 29 (35), 10863–10868.
- Kempton, M.J., Ruberto, G., Vassos, E., Tatarelli, R., Girardi, P., Collier, D., Frangou, S., 2009b. Effects of the CACNA1C risk allele for bipolar disorder on cerebral gray matter volume in healthy individuals. *Am. J. Psychiatry* 166 (12), 1413–1445.
- Kempton, M.J., Salvador, Z., Munafò, M.R., Geddes, J.R., Simmons, A., Frangou, S., Williams, S.C., 2011. Structural neuroimaging studies in major depressive disorder. Meta-analysis and comparison with bipolar disorder. *Arch. Gen. Psychiatry* 68 (7), 675–690.
- Lelli-Chiesa, G., Kempton, M.J., Jogia, J., Tatarelli, R., Girardi, P., Powell, J., Collier, D. A., Frangou, S., 2011. The impact of the Val158Met catechol-O-methyltransferase genotype on neural correlates of sad facial affect processing in patients with

- bipolar disorder and their relatives. *Psychol. Med.* 41 (4), 779–788.
- Lindquist, K.A., Wager, T.D., Kober, H., Bliss-Moreau, E., Barrett, L.F., 2012. The brain basis of emotion: a meta-analytic review. *Behav. Brain Sci.* 35 (3), 121–143.
- Lukoff, D., Liberman, R.P., Nuechterli, K.H., 1986. Symptom monitoring in the rehabilitation of schizophrenic patients. *Schizophr. Bull.* 12 (4), 578–602.
- Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., Mourao-Miranda, J., 2010. Quantitative prediction of subjective pain intensity from whole brain fMRI data using Gaussian processes. *NeuroImage* 49, 2178–2189.
- Mitchell, P.B., Goodwin, G.M., Johnson, G.F., Hirschfeld, R.M., 2008. Diagnostic guidelines for bipolar depression: a probabilistic approach. *Bipolar Disord.* 210 (1 Pt 2), 144–152.
- Mitchell, P.B., Frankland, A., Hadzi-Pavlovic, D., Roberts, G., Corry, J., Wright, A., Loo, C.K., Breakspear, M., 2011. Comparison of depressive episodes in bipolar disorder and in major depressive disorder within bipolar disorder pedigrees. *Br. J. Psychiatry* 199 (4), 303–309.
- Moorhead, T.W., McKirdy, J., Sussmann, J.E., Hall, J., Lawrie, S.M., Johnstone, E.C., McIntosh, A.M., 2007. Progressive gray matter loss in patients with bipolar disorder. *Biol. Psychiatry* 62 (8), 894–900.
- Mwangi, B., Spiker, D., Zunta-Soares, G.B., Soares, J.C., 2014. Prediction of pediatric bipolar disorder using neuroanatomical signatures of the amygdala. *Bipolar Disord.* 16 (7), 713–721.
- Niendam, T.A., Laird, A.R., Ray, K.L., Dean, Y.M., Glahn, D.C., Carter, C.S., 2012. Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognit. Affect. Behav. Neurosci.* 12 (2), 241–268.
- Owen, A.M., McMillan, K.M., Laird, A.R., Bullmore, E., 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25 (1), 46–59.
- Perrier, E., Pompei, F., Ruberto, G., Vassos, E., Collier, D., Frangou, S., 2011. Initial evidence for the role of CACNA1C on subcortical brain morphology in patients with bipolar disorder. *Eur. Psychiatry* 26 (3), 135–137.
- Pompei, F., Dima, D., Rubia, K., Kumari, V., Frangou, S., 2011a. Dissociable functional connectivity changes during the Stroop task relating to risk, resilience and disease expression in bipolar disorder. *NeuroImage* 57 (2), 576–582.
- Pompei, F., Jogia, J., Tatarelli, R., Girardi, P., Rubia, K., Kumari, V., Frangou, S., 2011b. Familial and disease specific abnormalities in the neural correlates of the stroop task in bipolar disorder. *NeuroImage* 56 (3), 1677–1684.
- Rasmussen, C., Williams, C.K.I., 2006. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge MA.
- Rocha-Rego, V., Jogia, J., Marquand, A.F., Mourao-Miranda, J., Simmons, A., Frangou, S., 2014. Examination of the predictive value of structural magnetic resonance scans in bipolar disorder: a pattern classification approach. *Psychol. Med.* 44 (3), 519–532.
- Ruberto, G., Vassos, E., Lewis, C.M., Tatarelli, R., Girardi, P., Collier, D., Frangou, S., 2011. The cognitive impact of the ANK3 risk variant for bipolar disorder: initial evidence of selectivity to signal detection during sustained attention. *PLoS One* 6 (1), e16671.
- Schnack, H.G., Nieuwenhuis, M., van Haren, N.E., Abramovic, L., Scheewe, T.W., Brouwer, R.M., Hulshoff Pol, H.E., Kahn, R.S., 2014. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *NeuroImage* 84, 299–306.
- Schrouff, J., Rosa, M.J., Rondina, J.M., Marquand, A.F., Chu, C., Ashburner, J., Phillips, C., Richiardi, J., Mourão-Miranda, J., 2013. PRoNTo: pattern recognition for neuroimaging toolbox. *Neuroinformatics* 11 (3), 319–337.
- Takahashi, T., Walterfang, M., Wood, S.J., Kempton, M.J., Jogia, J., Lorenzetti, V., Soulsby, B., Suzuki, M., Velakoulis, D., Pantelis, C., Frangou, S., 2010. Pituitary volume in patients with bipolar disorder and their first-degree relatives. *J. Affect. Disord.* 124 (3), 256–261.
- Tijssen, M.J., Van, O.S., Wittchen, H.U., Lieb, R., Beesdo, K., Wichers, M., 2010. Risk factors predicting onset and persistence of subthreshold expression of bipolar psychopathology among youth from the community. *Acta Psychiatr. Scand.* 122 (3), 255–266.
- Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.
- Walterfang, M., Wood, A.G., Barton, S., Velakoulis, D., Chen, J., Reutens, D.C., Kempton, M.J., Haldane, M., Pantelis, C., Frangou, S., 2009. Corpus callosum size and shape alterations in individuals with bipolar disorder and their first-degree relatives. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 33 (6), 1050–1057.
- Wolfers, T., Buitelaar, J.K., Beckmann, C.F., Franke, B., Marquand, A.F., 2015. From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci. Biobehav. Rev.* 57, 328–349.
- World Health Organization, 2008. *The Global Burden of Disease: 2004 Update*. World Health Organization, Geneva, Switzerland http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf.