# Towards a blood-based diagnostic panel for bipolar disorder

CrossMark

Frieder Haenisch [a,1], Jason D. Cooper [a,1], Andreas Reif [b], Sarah Kittel-Schneider [b], Johann Steiner [c], F. Markus Leweke [d], Matthias Rothermundt [e], Nico J.M. van Beveren [f], Benedicto Crespo-Facorro [g], David W. Niebuhr [h], David N. Cowan [h], Natalya S. Weber [h], Robert H. Yolken [i], Brenda W.J.H. Penninx [j], Sabine Bahn [a,*]

[a] Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, United Kingdom
[b] Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, Goethe University Hospital Frankfurt, Germany
[c] Department of Psychiatry, Psychotherapy and Psychosomatic Medicine, University Magdeburg, Magdeburg, Germany
[d] Central Institute of Mental Health, University of Heidelberg, Medical Faculty Mannheim, Mannheim, Germany
[e] Department of Psychiatry, University of Münster, Germany
[f] Department of Psychiatry, Erasmus University Medical Centre, Rotterdam, The Netherlands
[g] CIBERSAM, University Hospital Marqués de Valdecilla, University of Cantabria – IDIVAL, Department of Psychiatry, Santander, Spain
[h] Walter Reed Army Institute of Research, Silver Spring, MD, USA
[i] Johns Hopkins University School of Medicine, Baltimore, MD, USA
[j] Department of Psychiatry, EMGO Institute, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands

## ARTICLE INFO

## ABSTRACT

*Background:* Bipolar disorder (BD) is a costly, devastating and life shortening mental disorder that is often misdiagnosed, especially on initial presentation. Misdiagnosis frequently results in ineffective treatment. We investigated the utility of a biomarker panel as a diagnostic test for BD.

*Methods and findings:* We performed a meta-analysis of eight case-control studies to define a diagnostic biomarker panel for BD. After validating the panel on established BD patients, we applied it to undiagnosed BD patients. We analysed 249 BD, 122 pre-diagnostic BD, 75 pre-diagnostic schizophrenia and 90 first onset major depression disorder (MDD) patients and 371 controls. The biomarker panel was identified using ten-fold cross-validation with lasso regression applied to the 87 analytes available across the meta-analysis studies.

We identified 20 protein analytes with excellent predictive performance [area under the curve (AUC) $\geqslant$ 0.90]. Importantly, the panel had a good predictive performance (AUC 0.84) to differentiate 12 misdiagnosed BD patients from 90 first onset MDD patients, and a fair to good predictive performance (AUC 0.79) to differentiate between 110 pre-diagnostic BD patients and 184 controls. We also demonstrated the disease specificity of the panel.

*Conclusions:* An early and accurate diagnosis has the potential to delay or even prevent the onset of BD. This study demonstrates the potential utility of a biomarker panel as a diagnostic test for BD.

© 2015 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/).

## 1. Introduction

Bipolar disorder (BD) is a devastating mental disorder characterised by remitting and relapsing episodes of depression and (hypo)mania, which can also include psychotic symptoms. Disease onset commonly occurs in late adolescence or early adulthood, affecting men and women equally. BD has a lifetime prevalence of 1.0% for bipolar I disorder and 1.1% for bipolar II disorder (Merikangas et al., 2007).

Diagnosis of BD is based upon operationalized criteria with the aim to identify BD mood symptoms and patterns. The initial presentation of BD overlaps with either the depressive features of major depressive disorder (MDD) or the manic psychotic features of schizophrenia (SCZ). As most individuals seek psychiatric treatment for depressive symptoms at the onset of the disorder, the condition is frequently misdiagnosed because the subsequent (hypo)manic episode cannot be anticipated (Colom et al., 2006; Vieta et al., 2009). Ghaemi et al. estimated that the average delay for BD patients to be correctly diagnosed was 7.5 years (Ghaemi et al., 1999). An MDD misdiagnosis of BD patients is commonly associated with inappropriate antidepressant treatment that can precipitate (hypo)manic symptoms, worsening the outcome for the BD patient.

Extensive research into neuroimaging based biomarkers and genetic risk factors, has not resulted in a diagnostic test for routine clinical use. However, gene expression studies in monocytes from BD patients have identified 22 discriminating inflammatory genes in a whole genome analysis (Padmos et al., 2008). In addition, a proof of concept whole-genome gene expression study demonstrated the value of blood biomarkers for predicting BD disease state (Le-Niculescu et al., 2009). The most promising finding comes from a case–control study with 'never-medicated' BD patients, where a 10-gene model predicted the patient group with 89% sensitivity and 75% specificity ($p < 0.001$) (Clelland et al., 2013). In this study, we adopted a proteomics based approach to evaluate the potential of a diagnostic biomarker blood test for BD. After defining the biomarker panel and the validation in established BD patients, we applied the test to pre-diagnostic BD patients and controls, as well as first onset MDD patients, including patients who later developed (hypo)manic symptoms. We also tested the specificity of the panel.

## 2. Methods

This study consists of discovery, validation and application stages. In the discovery stage, we defined a mood-state-independent diagnostic biomarker panel for BD in a meta-analysis of eight independent case–control studies from five different clinical centres. The eight studies include a total of 158 established BD patients and 143 controls. In the validation stage, we attempted to validate the predictive performance of the diagnostic biomarker panel in a case–control study consisting of a further 66 established BD patients and 44 controls. Finally, in the application stage, we applied and evaluated the predictive performance of the diagnostic biomarker panel in undiagnosed BD patients and tested the disease specificity of the panel.

### 2.1. Study participants

#### 2.1.1. Discovery stage

Patients were recruited in four clinical centres in Germany (Cologne, Magdeburg, Münster and Würzburg) and one in the Netherlands (Rotterdam). The recruitment inclusion and exclusion criteria were similar for all eight case–control studies. The criteria required male and female participants to be within the age range of 18–60 years, have a body mass index (BMI) between 18 and 40 kg/m (Colom et al., 2006), and test negative for recreational drug screening at the time of sampling (Table 1). BD was diagnosed according to criteria of the International Classification of Diseases – 10 (ICD-10) by a trained psychiatrist in a clinical setting. Severity of symptoms was assessed using standard questionnaire based rating scales [Hamilton Depression Rating Scale (HAMD), Young Mania Rating Scale (YMRS) and Montgomery–Åsberg Depression Rating Scale (MADRS)]. At the clinical centre in Würzburg, BD (Table 1) was diagnosed by two trained psychiatrists and the diagnosis was confirmed by the Operational Criteria Checklist for Affective and Psychotic Illness (OPCRIT). In all clinical centres, both bipolar I and bipolar II disorder patients were recruited. The BD patients were in one of the following mood states at the time of sample collection: depressed, mixed affective, hypomanic, manic or euthymic (Table 1). Age and sex matched controls from similar geographical areas, with a similar socioeconomic background were recruited with a maximum delay of four weeks. The exclusion criteria included a record of mental illness (only applied to controls), diagnosis of coronary heart disease or cardiac insufficiency, autoimmune disorders, acute or chronic infections or treatment with immunosuppressive or immunomodulating drugs or antibiotics, other neuropsychiatric disorders or chronic terminal diseases

affecting the brain, like cancer or hepatic and renal insufficiency, alcohol or drug addiction. No clinical assessment of controls was conducted. Patients and controls were fasting for at least two hours prior to blood sample collection. The study procedures and protocols received approval from the respective local ethical committees and informed written consent was obtained from all participants.

#### 2.1.2. Validation stage

We tested the predictive performance of the diagnostic biomarker panel identified in the discovery stage in a further case–control study from Würzburg in Germany (Table 1). Clinical assessments, exclusion and inclusion criteria were as described for the discovery studies.

#### 2.1.3. Application stage

We evaluated the predictive performance of the diagnostic biomarker panel in three nested case–control studies, two from the US Department of Defense Serum Repository (DoDSR) and one from the Netherlands Study of Depression and Anxiety (NESDA). For the two nested case–control studies from the military, sera were obtained from the US Department of Defense Serum Repository (DoDSR), which contains over 55 million serum specimens remaining from mandatory HIV test samples of military personnel (Perdue et al., 2015; Rubertone and Brundage, 2002). Data and sera retrieval were performed by the Armed Forces Health Surveillance Center (AFHSC) and coordinated by the Military New-Onset Psychosis Project (MNOPP) investigators at the Walter Reed Army Institute of Research (WRAIR) (Niebuhr et al., 2008). The medical and demographic data were provided by the Defense Medical Surveillance System (DMSS), AFHSC, United States Department of Defense (US DoD), Silver Spring, Maryland (data range from 1989 to 2006; released in 2007 and 2008) and serum specimens were retrieved from the DoDSR, AFHSC, US DoD (Silver Spring, MD, USA; specimens range from 1988 to 2006, released in 2007 and 2008). Sera were then transferred to the Johns Hopkins School of Medicine (Baltimore, MD, USA) prior to testing. Samples were then selected from 185 individuals who presented with psychiatric symptoms within 30 days after the blood collection and who later received a DSM-IV diagnosis of either BD or SCZ (MNOPP). Control subjects were selected from the DMSS records of active duty military service population with no inpatient or outpatient psychiatric disorder diagnoses. All data were previously collected for other purposes, and analyses were conducted on de-identified data. As only de-identified data were utilised, informed consent waivers were granted by the WRAIR Institutional Review Board (for additional information please see Supplementary Material 1). The third nested case–control study was drawn from NESDA, which is an ongoing longitudinal cohort study including 2,981 participants aged 18 through to 65 years (Penninx et al., 2008). Patients from the NESDA were recruited from three clinical sites in the Netherlands (Amsterdam, Groningen and Leiden) (Penninx et al., 2008). A four-hour baseline assessment included written questionnaires, interviews, a medical examination, a cognitive computer task and collection of blood and saliva samples. Furthermore extensive information about key (mental) health outcomes was gathered. Serum from a subset of 1701 participants was profiled. We analysed 368 controls and 102 recent onset MDD patients, 12 of which were diagnosed with BD within two years of the baseline interview (Table 1).

Informed and written consent was given by all participants, and the study protocols, analysis of samples and test methods were approved by the local Institutional Ethics Review Boards and were in compliance with the Standards for Reporting of Diagnostic Accuracy.

**Table 1**

Summary of demographic characteristics of the eight case–control studies used in the meta-analysis (discovery stage), the validation stage and the application stage. Analyte data was acquired between August 2005 and December 2013 (7 years and 4 months). BD – bipolar patients, Controls – control subjects, NR – Not recorded, NA – information not available / not applicable, *missing data points, **cannabis in the past month, d – depressive, e – euthymic, m – manic.

| | Cohort | Rotterdam | | Würzburg | | | | | | Magdeburg | | | | Cologne | | Münster | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Study | 1 | | 2 | | 3 | | 4 | | 5 | | 6 | | 7 | | 8 | |
| | Group | Controls | BD | Controls | BD | Controls | BD | Controls | BD | Controls | BD | Controls | BD | Controls | BD | Controls | BD |
| Discovery stage | Sample size | 40 | 10 | 13 | 28 | 6 | 15 | 18 | 60 | 15 | 7 | 13 | 5 | 12 | 15 | 26 | 18 |
| | Age [SD] | 26.8 [4.1] | 28.3 [11.1] | 42.6 [12.3] | 47.3 [12.8] | 44,5 [15.9] | 46,4 [13.5] | 45,7 [10.4] | 47 [10.9] | 40,8 [7.7] | 40,4 [7.7] | 29,2 [5.8] | 28,8 [6.0] | 41,7 [7.6] | 44,1 [11.1] | 44,6 [9.5] | 47,3 [14.2] |
| | Sex (m/f) | 33/7 | 7/3 | 6/7 | 11/17 | 3/3 | 6/7 | 7/11 | 29/31 | 5/10 | 2/5 | 8/5 | 3/2 | 6/6 | 5/10 | 10/16 | 10//8 |
| | BD subtype (1/2) | / | NR | / | 15/13 | / | 5/10 | / | 30/30 | / | NR | / | NR | / | NA | / | NR |
| | Mood state (d/e/m/mixed/NO) | / | NR | / | 19/0/4/5/0 | / | 1/3/6/5/0 | / | 27/31/1/1/0 | / | NR | / | NR | / | NA | / | NR |
| | Somatic medication | NA | NA | 5/13 | 14/28 | 0/6 | 11/15 | 4/18 | 37/60 | NR | NR | NR | NR | 4/12 | 2/15 | | NR |
| | Psychiatric medication | 0/40 | 6/10 | 0/13 | 28/28 | 0/6 | 15/15 | 0/18 | 57/60 | 0/15 | 0/7 | 0/13 | NR | 0/12 | 15/15 | 0/26 | NR |
| | Mood stabilizer | 0/40 | NA | 0/13 | 11/28 | 0/6 | 7/15 | 0/18 | 55/60 | 0/15 | 0/7 | 0/13 | NR | 0/12 | 8/15 | 0/26 | NR |
| | Anti-depressant | 0/40 | NA | 0/13 | 16/28 | 0/6 | 2/15 | 0/18 | 37/60 | 0/15 | 0/7 | 0/13 | NR | 0/12 | 3/15 | 0/26 | NR |
| | Antipsychiotic | 0/40 | 6/10 | 0/13 | 24/28 | 0/6 | 8/15 | 0/18 | 16/60 | 0/15 | 0/7 | 0/13 | NR | 0/12 | 2/15 | 0/26 | NR |
| | BMI [SD] | NA | NA | 26.0 [4.7]* | 27.5 [6.4]* | NA | NA | 23.6 [3.2]* | 27.8 [5.5]* | 25.5 [3.7] | 26.8 [7.2] | 23.3 [2.1] | 23.1 [3.3] | NA | NA | NR | NR |
| | Smoker | 10/40 | 10/10 | 4/13 | 11/28 | 1/6 | 4/15 | 2/18 | 21/60 | 9/15 | 5/7 | 3/13 | 1/5 | 2/15 | 11/15 | NR | NR |
| | Canabis life time | NA | 6/10 | NA | NA | NA | NA | NA | NA | 1/15 | 1/7 | 0/13 | 0/5 | 1/12 | 2/15 | NR | NR |
| | Molecular profiling | May 2009 | | Nov. 2010 | | Aug. 2011 | | Jun 2012 | | Sep. 2010 | | Mar 2011 | | Mar. 2011 | | Mar. 2011 | |

| | | Study | Würzburg | | | Study | NESDA | | | | USA Military | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Disease group | Controls | BD | | Disease group | Controls | First depressive episode patients | | | Controls | Pre-diagnostic | |
| | | | | | | | | First-onset MDD | Un-diagnosed BD | | | BD | SCZ |
| Validation stage | Sample size | | 44 | 66 | Application stage | Sample size | 368 | 90 | 12 | 184 | 110 | 75 |
| | Age [SD] | | 29.2 [8.9] | 41.9 [13.4] | | Age [SD] | 39.6 [14.9] | 38.5 [13.4] | 35.4 [10.2] | 22.4 [3.6] | 21.3 [4.2] | 24.3 [4.5] |
| | Sex (m/f) | | 23/20 | 24/39 | | Sex (m/f) | 148/220 | 33/57 | 5/7 | 136/48 | 70/40 | 67/8 |
| | BD subtype (1/2/other) | | / | 35/23/8 | | BD subtype (1/2) | / | / | / | / | / | / |
| | Mood state (d/e/m/mixed/other) | | / | 28/8/13/13/4 | | Mood state (d/e/m/mixed) | / | 90/0/0/0/0 | 12/0/0/0/0 | / | / | / |
| | Somatic medication | | 11/44 | 22/66 | | | 0/368 | 22/90 | 2/12 | NR | NR | NR |
| | Psychiatric medication | | 0/44 | 45/66 | | | 0/368 | 39/90 | 2/12 | 0/184 | 0/110 | 0/75 |
| | Mood stabilizer | | 0/44 | 41/66 | | | 0/368 | 0/90 | 0/12 | 0/184 | 0/110 | 0/75 |
| | Anti-depressant | | 0/44 | 16/66 | | | 0/368 | 39/90 | 2/12 | 0/184 | 0/110 | 0/75 |
| | Antipsychiotic | | 0/44 | 23/66 | | | 0/368 | 39/90 | 0/12 | 0/184 | 0/110 | 0/75 |
| | BMI [SD] | | 23.8 [4.1]* | 27.6 [5.1]* | | | 24.9 [4.6] | 25.6 [4.8] | 24.7 [2.7] | NR | NR | NR |
| | Smoker | | 10/44 | 23/66 | | | 96/368 | 31/90 | 8/12 | NR | NR | NR |
| | Canabis life time | | NA | NA | | | 21/368 | 3/90** | 1/12** | NR | NR | NR |
| | Molecular profiling | | Dec. 2013 | | | | Apr. 2013 | | | Jun. 2009 | | |

## 2.2. Sample preparation

Strict standard operating protocols were adhered to for serum sample preparation and used by all clinical centres, except for the U.S. Military (Perdue et al., 2015; Rubertone and Brundage, 2002) and NESDA (Penninx et al., 2008) who have established their own protocols. The protocol in brief: a blood sample was taken within two days of the clinical assessment. Serum was collected from acutely ill fasting patients and controls using Vacutainer (Becton–Dickinson, Franklin Lakes, NJ, USA). Blood clotting time was two hours at room temperature prior to centrifugation for 15 min at $1.100 \times g$ (except for studies 1 and 5–8, which employed centrifugation for 5 min at $4000 \times g$; Table 1). Samples were stored in low binding Eppendorf tubes (Hamburg, Germany) at $-80\,°C$. All sample shipments took place on dry ice.

## 2.3. Multiplex immunoassay analysis

Serum samples were randomized and processed blind to disease status using the multiplex immunoassay platform at Myriad Rules Based Medicine (Myriad RBM; Austin, Texas, USA). The Human DiscoveryMAP™ assay platform was used to measure different serum concentrations of proteins, peptides and small molecules (collectively referred to as 'analytes') (Bertenshaw et al., 2008). The clinical centres all provided anonymised patient data. The number of analytes measured differed between the studies depending on when the study samples were profiled (total range: 147 to 257 analytes). The commonly measured analytes are reported in Supplementary Table 1.

## 2.4. Statistical analysis

We pre-processed the analyte data from each study to exclude analytes with greater than 20% missing values and sample outliers were identified using the first two principal components and to impute missing data (as described previously (Schwarz et al., 2011a); Supplementary Table 2). We $log_{10}$-transformed the data to stabilize the variance.

All statistical analyses were performed in R (R Core Team, 2014). For the eight studies combined in the meta-analysis, we used ComBat, as implemented in the sva package (Leek et al., 2012), to adjust for the effects caused by running the study samples at different times (i.e. to make the subjects within and between the studies comparable). ComBat is an empirical Bayes method of adjusting for additive, multiplicative, and exponential batch effects developed for analysing microarray data (Johnson et al., 2007). In addition, we used ComBat to adjust the validation and application studies for batch effects caused by running samples within a study on different plates. We did not apply ComBat to the USA Military studies. 115 analytes and two covariates (age and sex) were common across the studies (Supplementary Table 1). We excluded a further 28 analytes with significant BD association heterogeneity across the studies (Supplementary Table 3). The 87 remaining analytes and two covariates form $2^{89} = 6.2 \times 10^{26}$ possible candidate models. We searched the model space using ten-fold cross-validation with lasso regression as implemented in the R package glmnet (Friedman et al., 2010). Lasso is a penalised method for restricting the residual sum of squares (deviance) and constraining the sum of the absolute values of the coefficients: $\sum_i |\beta| \leqslant t$, where $t$ is the 'tuning' parameter. As $t \to \infty$, $t$ has no effect and the solutions are the least squares estimates for the full model. For smaller $t$ values, solutions are shrunken versions of the least squares estimates with many coefficients decreased to the null value. $t$ was defined using ten-fold cross-validation, as the value of $t$ minimising the $t$-penalised residual sum of squares,

which is equivalent to maximising the $t$-penalised log likelihood (James et al., 2013). Although the coefficient estimates are biased to be small, a lasso estimator can have smaller error than a standard maximum likelihood estimator when applied to new data.

As the analytes were selected based on minimising the $t$-penalised residual sum of squares, $p$-values for each of the selected analytes are not relevant and not reported. We measured the predictive performance of the diagnostic biomarker panel using sensitivity and specificity and area under the receiver operating characteristic (ROC) curves (AUC: 0.9–1 = excellent; 0.8–0.9 = good; 0.7–0.8 = fair; 0.6–0.7 = poor; 0.5–0.6 = fail). Optimal trade-offs between sensitivity and specificity were determined by maximising Youden's index ($J$; where $J$ = sensitivity + specificity $-$ 1).

We attempted to validate the diagnostic biomarker panel (Table 2) in the independent case–control study from Würzburg (Table 1). We fit a logistic regression model to the Würzburg analyte data corresponding to the biomarker panel and predicted BD status. A similar approach was also adopted in the application stage. To test the predictive performance of the biomarker panel in first onset MDD and pre-diagnostic patients, we applied a fitted model derived from a BD patient–control study that was profiled and pre-processed at the same time as the MDD or SCZ patients. For example, in the analysis of the USA Military studies, we fitted the logistic model to the pre-diagnostic BD patient–control data, and then applied the fitted model to the pre-diagnostic SCZ patient–control data.

## 3. Results

### 3.1. Discovery stage

Eighty-seven analytes and two covariates (age and sex) were available across the eight case–control studies (183 BD patients and 149 controls) and included in the variable selection analysis to define the diagnostic biomarker panel. We identified a biomarker panel of 20 analytes with an excellent predictive performance (AUC = 0.90; Fig. 1A; boxplots of the $log_{10}$ transformed analyte levels are provided in Supplementary Figs. 1). When the selected analytes were grouped into functional pathways, 11 analytes were found to play a role in the inflammatory cascade (Table 2). The majority of these analytes (7 out of 11) have a pro-inflammatory function and the remainder are anti-inflammatory. A second group of seven analytes can be clustered into lipid transport-related proteins and proteins with metalloendopeptidase activity.

### 3.2. Validation stage

The diagnostic performance of the biomarker panel was tested in an independent case–control study (Table 1). Importantly, as in the meta-analysis, this study consisted of established BD patients in different mood states. The predictive performance of the biomarker panel was excellent (AUC = 0.92; Fig. 1A) despite the panel being limited to 16 of the 20 analytes selected in the discovery stage (Supplementary Table 4). Note that the analyte model coefficients were estimated based upon the validation study data and not upon the meta-analysis.

### 3.3. Application stage

In this stage, we tested the diagnostic performance of the biomarker panel on undiagnosed BD patients and tested the disease specificity of the panel. It is important to note that the patient samples used in the application stage were collected at an earlier time point in the disease course compared to the BD patients from the

**Table 2**
The 20 analytes selected in the discovery stage to form the diagnostic biomarker panel.

| Molecular function | Analyte | Lasso coefficient |
|---|---|---|
| Pro-inflammatory | CD40 Ligand | −0.11 |
| | EN-RAGE | 0.06 |
| | Growth-Regulated alpha protein | −1.81 |
| | Macrophage Inflammatory Protein-1 beta (MIP-1 beta) | 0.35 |
| | Receptor for advanced glycosylation end products (RAGE) | −0.30 |
| | Serum Amyloid P-Component | 1.41 |
| | Tumour Necrosis Factor Receptor-Like 2 | 3.43 |
| Anti-inflammatory | Carcinoembryonic Antigen (CEA) | 0.46 |
| | CD5 | −1.69 |
| | Interleukin-1 receptor antagonist | 0.36 |
| | Interleukin-10 | 0.78 |
| Lipid transport | Apolipoprotein A1 | −1.35 |
| | Apolipoprotein A2 | −0.37 |
| | Lipoprotein (a) | 0.20 |
| Metalloendopeptidase activity | Angiotensin-Converting Enzyme (ACE) | −0.49 |
| | Matrix Metalloproteinase-3 | −1.87 |
| | Matrix Metalloproteinase-7 | 2.50 |
| | Matrix Metalloproteinase-9, total | 0.33 |
| Cysteine protease inhibitor | Cystatin C | 0.53 |
| Growth factor | Hepatocyte Growth Factor | 0.23 |

discovery and validation stages. Given the overlapping clinical presentations of MDD and BD (Fig. 1; middle section), and the negative implications of misdiagnosis, the need for a diagnostic test to distinguish between the two conditions is clear. To this end, we tested the predictive performance of the panel on 102 first onset MDD patients including 12 patients who were subsequently diagnosed with BD selected from NESDA. All of the 12 misdiagnosed BD patients experienced a (hypo)manic episode within two years of the baseline blood sample collection. We obtained a good predictive performance (AUC = 0.84; Fig. 1B; Supplementary Table 4). Two of the 20 panel analytes were not available.

As extensive clinical variables were available in NESDA, we made these additional variables available for selection (Supplementary Table 5). The addition of the Inventory of Depressive Symptomatology (IDS) score led to an increase in the test performance to excellent (AUC = 0.90; Supplementary Fig. 2). Importantly, IDS scores alone only achieved a fair predictive performance (AUC = 0.78). We note that the Beck Anxiety Inventory (BAI) score led to a similar increase in the predictive performance. In addition, to test the differential diagnostic utility of the biomarker panel, we applied the fitted model from the analysis of first onset MDD and misdiagnosed BD patients to differentiate between 90 first onset MDD patients from NESDA, who did not develop BD, and 368 controls. Importantly, the predictive performance of the biomarker panel for first onset MDD patients was poor (AUC = 0.64; Fig. 1C).

We then tested the predictive performance of the biomarker panel in pre-diagnostic BD patients (Fig. 1). We applied the diagnostic panel to 110 pre-diagnostic BD patients and 184 controls from the USA Military, where patients presented with initial psychiatric symptoms within 30 days of the blood test and later obtained a diagnosis of BD. The predictive performance was fair, but bordering the AUC threshold for a good predictive performance (AUC = 0.79; Fig. 1D). One of the 20 analytes forming the biomarker panel was not available (Supplementary Table 4). As we also had 75 pre-diagnostic SCZ patient samples from the USA Military, which underwent protein profiling and were pre-processed at the same time, we tested whether the biomarker panel could differentiate between the pre-diagnostic BD and SCZ patient groups. The predictive performance was excellent (AUC = 0.91; Fig. 1E). To test the specificity of the biomarker panel, we applied the fitted model from the analysis of pre-diagnostic BD patients and controls to differentiate between pre-diagnostic SCZ patients and controls. Importantly, the predictive performance was almost equivalent

to "tossing a coin" (AUC = 0.52; classification description 'fail'; Fig. 1D).

## 4. Discussion

To date, on-going research has not provided the basis for a diagnostic test for BD with clinical utility. In the present study, one of the largest diagnostic biomarker studies for BD, we demonstrated for the first time the potential of a biomarker panel to provide a blood-based diagnostic test for BD.

Given the insidious onset and that most BD patients initially present with depressive symptoms (Colom et al., 2006), the most appropriate time for a routine diagnostic test for BD will be when individuals present with a first depressive episode. When we applied this panel to differentiate between first onset MDD patients and patients who later develop (hypo)manic symptoms, we obtained a good predictive performance (AUC = 0.84). In addition, incorporation of the IDS rating score, resulted in a further increase in the predictive performance to excellent levels (AUC = 0.90; Supplementary Fig. 2).

We also tested the biomarker panel in pre-diagnostic BD patients and controls (AUC = 0.79), and determined the differential diagnostic utility of the panel by demonstrating that the BD panel cannot identify patients from other psychiatric disorders when those are compared to healthy controls. The panel was applied to first onset MDD patients and controls (AUC = 0.64) and pre-diagnostic SCZ patients and controls (AUC = 0.52). Furthermore, and as expected, we were able to demonstrate that the panel could differentiate between pre-diagnostic BD patients and pre-diagnostic SCZ patients (AUC = 0.91).

Despite the encouraging diagnostic utility of the biomarker panel in pre-diagnostic BD patients, a routine diagnostic test to detect pre-diagnostic patients will not be practical as the prevalence of individuals at risk to develop BD will be very low amongst the population tested.

We have also checked for common analytes in published MDD (Papakostas et al., 2013; Bilello et al., 2015) and SCZ (Schwarz et al., 2011b; Chan et al., 2015) diagnostic biomarker panels. The MDD panel from Papakostas et al. describes a set of nine biomarkers, of which seven were also measured in the present study (Supplementary Table 1). None of the seven analytes were included in the BD panel. A more recent study compared 687 current and 482 remitted MDD patients with 420 controls
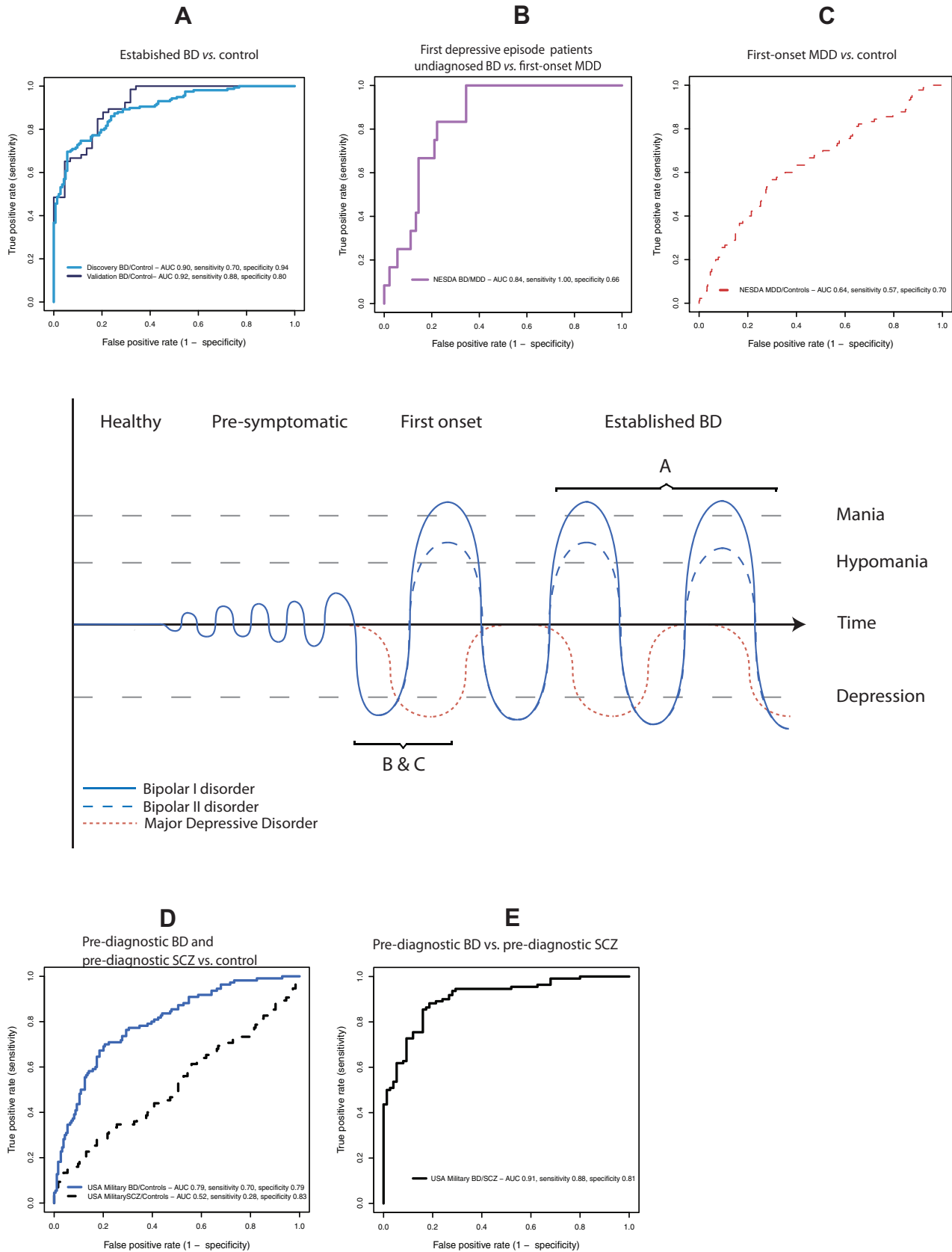
**Fig. 1.** A summary of the typical disease progression of BD and ROC curves derived from the discovery, validation and application stages. Middle section: Typical course of BD starting from no overt symptoms, an initial depressive episode and followed by a chronic remitting–relapsing disease course is depicted in blue; the dotted red line depicts a chronic MDD disease course. Note that a sub-group of BD patients (BD2) only experience (hypo)manic episodes, these patients are not represented in the figure. The labels A, B and C indicate corresponding disease stages represented in the ROC curves with the same lettering. Top and bottom sections, ROC curves and predictive performance: (A) The discovery and validation studies – established BD patients and controls; (B) First onset "MDD patients" including patients who later developed BD (NESDA); (C) First onset MDD patients and controls (NESDA); (D) Pre-diagnostic BD, pre-diagnostic SCZ and controls (USA military); (E) Pre-diagnostic BD and pre-diagnostic SCZ (USA Military). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

recruited within NESDA and found 34 analytes significantly changed. Five of the analytes are also present on the BD biomarker panel (carcinoembryonic antigen, EN-RAGE, growth-regulated alpha protein, interleukin-1 receptor antagonist and matrix metalloproteinase-3) (Romi et al., 2012). The study tested 171 analytes for significance of which 101 are in common with the 115 analytes available in the discovery phase of this study.

When compared to the diagnostic SCZ biomarker panel reported by Schwarz et al., we found three overlapping biomarkers (cluster of differentiation 40, interleukin-10 and carcinoembryonic antigen) of the described 34 biomarkers (19 analytes measured in common) (Schwarz et al., 2011b). The second, more comprehensive, SCZ panel, published by Chan et al., describes a 26 analyte panel and used a similar statistical approach to the one in this article. Out of 60 commonly measured analytes, three analytes are in common with the BD panel (apolipoprotein A1, interleukin-10 and interleukin 1 receptor antagonist) (Chan et al., 2015).

The molecular mechanisms of the selected analytes, which can distinguish BD patients in different disease stages (pre-diagnostic, first depressive episode and established patients), potentially reveal new insights into the progressive nature of BD.

Importantly, three analytes (Matrix-metalloproeinase-3 (MMP-3), MMP-7 and MMP-9) belong to a structurally related family of secreted proteases that play an important role in extracellular matrix degradation (Chopra et al., 2014). Matrix-metalloproteinases (MMPs) have been implicated in the regulation of cell survival, angiogenesis, cell signalling and the maintenance of an intact blood–brain barrier. Alterations in serum levels of several MMPs have been reported for neurodegenerative and neuroinflammatory diseases like Alzheimer's disease and multiple sclerosis (MS) (Romi et al., 2012; Waubant et al., 1999).

Overexpression of MMPs, including MMPs secreted by T-cells and macrophages, are known to damage and open the blood brain barrier (Chopra et al., 2014; Leppert et al., 1995; Rosenberg et al., 1992). MMP-3, -7 and -9 expression is increased in astrocytes, microglia and neurons surrounding white matter lesions in MS (Cossins et al., 1997; Maeda and Sobel, 1996). MMP-9 expression levels in brain tissue of MS patients were reported to be correlated with increased serum levels (Waubant et al., 1999). We found increased levels of MMP-9 and -7 in serum from BD patients, while MMP-3 levels were decreased. Interestingly, MS patients have a significantly higher life-time risk to develop BD (Carta et al., 2014). In addition to MMPs, antipsychotic treatment might further contribute to blood–brain barrier dysfunction (Zetterberg et al., 2014).

Over half of the analytes (11/20) in the panel were associated with inflammatory processes. Seven of these analytes are pro-inflammatory mediators and the remaining four are anti-inflammatory (see Table 2). Levels of three of the seven pro-inflammatory analytes (CD40 Ligand, Growth-Regulated alpha protein and Receptor for advanced glycosylation end products) were decreased and in addition, three anti-inflammatory analytes (Carcinoembryonic Antigen, Interleukin-1 receptor antagonist and Interleukin-10) were found to be increased in BD patients. These findings do not support some published results, which report a predominant increase of pro-inflammatory analytes in BD (Barbosa et al., 2014; Rosenblat and McIntyre, 2015; Kupka et al., 2002; O'Brien et al., 2006). However, a few studies support our findings of an imbalance in pro-inflammatory and anti-inflammatory mediators in BD (Kim et al., 2007; Knijff et al., 2007). In addition, a recent study also reported elevated levels of carcinoembryonic antigen (CEA) in BD patients (Kaplan et al., 2015). CEA is involved in the activation and production of anti-inflammatory cytokines such as IL-10 (Thomas et al., 2011). The authors suggest that CEA could have utility for diagnosing BP (Kaplan et al., 2015).

A third functional group of changing serum proteins in BD relate to decreased levels of Apolipoprotein A1 and A2 and increased levels of Lipoprotein (a), which facilitate cholesterol and triglyceride transport, but also play a role in regulation of inflammation (Keeney et al., 2013). Decreased levels of Apolipoprotein A1 have been found in serum and cerebrospinal fluid of patients suffering from a wide range of neurodegenerative disorders (Keeney et al., 2013). These findings are also consistent with an increased risk for cardiovascular events in BD, which support the understanding that BD, like many other chronic disorders such as diabetes, cardiovascular disease and neurodegenerative disorders, presents with prominent immune alterations. Interestingly, the immune changes that we identified for bipolar disorder appear to be distinct from those seen in schizophrenia and major depressive disorder (Papakostas et al., 2013; Chan et al., 2015; Lakka et al., 2002).

The ideal study to distinguish biomarkers which are either cause or effect of a BD disease process, would be to conduct a prospective study to monitor the levels of the various analytes in the same probands/patients over time and to relate these analytes with clinical symptomology and treatment. However, this would be a substantial undertaking and unfortunately this question can't be addressed in the present study. The key goal of our study was to try to evaluate the utility of a biomarker panel for the diagnosis of bipolar disorder.

However, the fact that the same protein signature can identify individuals who have (as yet) not developed overt symptoms and were diagnosed with bipolar disorder (military cohort; NESDA misdiagnosed group) provides at least some evidence that the identified changes are not secondary to disease state or medication and may be aetiologically linked to the disorder (see Supplementary Table 6 for a representative overview of BD patient medication). However, it remains unclear whether the identified alterations are secondary to developmental, environmental and/or genetic factors. Unfortunately, this is also the case for other chronic, late onset disorders such as cardiovascular disease, diabetes and many others.

Nevertheless, it is possible that the inflammatory dysregulation is precipitating the onset of psychopathology and therefore may represent a therapeutic target (Chan et al., 2014). Ayorech et al. have reviewed sixteen studies using off target anti-inflammatory medication and conclude that there is indeed evidence for efficacy of anti-inflammatory medications in BD treatment (Ayorech et al., 2014).

There are several limitations to the present study. Firstly, we did not take different bipolar sub-classifications (i.e. BP1 versus BD2) and mood states at the time of sample collection into account. Secondly, we cannot rule out that a small number of the biomarkers in the panel may be associated with medication rather than disease status. However, we have demonstrated that medication as cofounding factor is very unlikely to have a strong impact on the biomarker panel. If the biomarker panel was discriminating between bipolar patients and controls primarily based on medication, we would expect this panel to have a limited predictive performance in the antipsychotic naïve pre-diagnostic (Supplementary Material 1) and 'misdiagnosed' (NESDA; only 2 of 12 patients were on antidepressant medication and none received mood stabilising or antipsychotic medication) bipolar disorder patient cohorts. In contrast, we see a fair to good predictive performance.

A second general limitation of this study is the lack of consistent demographic data across all clinical centres. As age and sex were the only covariates available across the eight case–control studies used in the meta-analysis, we identified a diagnostic test without considering other relevant clinical variables, like symptom scores. Of all the studies analysed, the most relevant for a diagnostic test was the nested study from NESDA. However, only a small number

of BD patients who were initially diagnosed with MDD and later developed BD symptomatology were available, larger studies will be necessary to validate the findings from NESDA.

In conclusion, the 20 analytes reported here, represent a preliminary panel of validated biomarkers from which a definitive signature for the diagnosis of BD could be developed. The ultimate goal will be to implement a low-cost blood test that can be routinely used in primary and secondary clinical care settings in conjunction with a clinical interview for the diagnosis of BD before the development of hypomanic or manic symptoms.

## Conflict of interest

Jason D. Cooper has been a consultant for Myriad-Genetics Inc. and is a consultant for Psynova Neurotech Ltd.

Robert Yolken is a member of the Stanley Medical Research Institute Board of Directors and Scientific Advisory Board. The terms of this arrangement are being managed by the Johns Hopkins University in accordance with its conflict of interest policies.

David W. Niebuhr, David N. Cowan and Natalya S. Weber do not have any actual or potential conflict of interest to declare. The views expressed are those of the authors and should not be construed to represent the positions of the US Department of the Army or Department of Defense.

Sabine Bahn has been a consultant for Myriad-Genetics Inc. and is a director of Psynova Neurotech Ltd.

All other authors have no conflict of interest to declare. The authors have no additional financial interests.

## Acknowledgments

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbi.2015.10.001.

## References

Ayorech, Z., Tracy, D.K., Baumeister, D., Giaroli, G., 2014. Taking the fuel out of the fire: evidence for the use of anti-inflammatory agents in the treatment of bipolar disorders. J. Affect. Disord. 174C, 467–478.

Barbosa, I.G., Morato, I.B., de Miranda, A.S., Bauer, M.E., Soares, J.C., Teixeira, A.L., 2014. A preliminary report of increased plasma levels of IL-33 in bipolar disorder: further evidence of pro-inflammatory status. J. Affect. Disord. 157, 41–44.

Bertenshaw, G.P., Yip, P., Seshaiah, P., et al., 2008. Multianalyte profiling of serum antigens and autoimmune and infectious disease molecules to identify biomarkers dysregulated in epithelial ovarian cancer. Cancer Epidemiol. Biomarkers Prev. 17, 2872–2881.

Bilello, J.A., Thurmond, L.M., Smith, K.M., et al., 2015. MDDScore: confirmation of a blood test to aid in the diagnosis of major depressive disorder. J. Clin. Psychiatry 76, e199–e206.

Carta, M.G., Moro, M.F., Lorefice, L., et al., 2014. The risk of bipolar disorders in multiple sclerosis. J. Affect. Disord. 155, 255–260.

Chan, M.K., Gottschalk, M.G., Haenisch, F., et al., 2014. Applications of blood-based protein biomarker strategies in the study of psychiatric disorders. Prog. Neurobiol. 122C, 45–72.

Chan, M.K., Krebs, M.-O., Cox, D., et al., 2015. Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. Transl. Psychiatry 5, e601.

Chopra, K., Baveja, A., Kuhad, A., 2014. MMPs: a novel drug target for schizophrenia. Expert Opin. Ther. Targets, 1–9.

Clelland, C.L., Read, L.L., Panek, L.J., Nadrich, R.H., Bancroft, C., Clelland, J.D., 2013. Utilization of never-medicated bipolar disorder patients towards development and validation of a peripheral biomarker profile. PLoS ONE 8. http://dx.doi.org/10.1371/journal.pone.0069082.

Colom, F., Vieta, E., Daban, C., Pacchiarotti, I., Sánchez-Moreno, J., 2006. Clinical and therapeutic implications of predominant polarity in bipolar disorder. J. Affect. Disord. 93, 13–17.

Cossins, J.A., Clements, J.M., Ford, J., et al., 1997. Enhanced expression of MMP-7 and MMP-9 in demyelinating multiple sclerosis lesions. Acta Neuropathol. 94, 590–598.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33, 1–22.

Ghaemi, S.N., Sachs, G.S., Chiou, A.M., Pandurangi, A.K., Goodwin, F.K., 1999. Is bipolar disorder still underdiagnosed? Are antidepressants overutilized? J. Affect. Disord. 52, 135–144.

James, Gareth, Witten, Daniela, Hastie, Trevor, Tibshirani, Robert, 2013. An Introduction to Statistical Learning: With Applications in R, Corr. 4th. Springer.

Johnson, W.E., Li, C., Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8, 118–127.

Kaplan, I., Bulut, M., Atli, A., Güneş, M., Kaya, M.C., Çolpan, L., 2015. Serum levels of carcinoembryonic antigen (CEA) in patients with bipolar disorder. Acta Neuropsychiatr. 27, 177–181.

Keeney, J.T.R., Swomley, A.M., Förster, S., Harris, J.L., Sultana, R., Butterfield, D.A., 2013. Apolipoprotein A-I: insights from redox proteomics for its role in neurodegeneration. Proteomics Clin. Appl. 7, 109–122.

Kim, Y.-K., Jung, H.-G., Myint, A.-M., Kim, H., Park, S.-H., 2007. Imbalance between pro-inflammatory and anti-inflammatory cytokines in bipolar disorder. J. Affect. Disord. 104, 91–95.

Knijff, E.M., Nadine, B.M., Kupka, R.W., et al., 2007. An imbalance in the production of IL-1β and IL-6 by monocytes of bipolar patients: restoration by lithium treatment. Bipolar Disord. 9, 743–753.

Kupka, R.W., Breunis, M.N., Knijff, E., Ruwhof, C., Nolen, W.A., Drexhage, H.A., 2002. Immune activation, steroid resistancy and bipolar disorder. Bipolar Disord. 4 (Suppl. 1), 73–74.

Lakka, H.-M., Laaksonen, D.E., Lakka, T.A., et al., 2002. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. JAMA 288, 2709–2716.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E., Storey, J.D., 2012. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 28, 882–883.

Le-Niculescu, H., Kurian, S.M., Yehyawi, N., et al., 2009. Identifying blood biomarkers for mood disorders using convergent functional genomics. Mol. Psychiatry 14, 156–174.

Leppert, D., Waubant, E., Galardy, R., Bunnett, N.W., Hauser, S.L., 1995. T cell gelatinases mediate basement membrane transmigration in vitro. J. Immunol. 154, 4379–4389.

Maeda, A., Sobel, R.A., 1996. Matrix metalloproteinases in the normal human central nervous system, microglial nodules, and multiple sclerosis lesions. J. Neuropathol. Exp. Neurol. 55, 300–309.

Merikangas, K.R., Akiskal, H.S., Angst, J., et al., 2007. Lifetime and 12-month prevalence of bipolar spectrum disorder in the National Comorbidity Survey replication. Arch. Gen. Psychiatry 64, 543–552.

Niebuhr, D.W., Millikan, A.M., Yolken, R., Li, Y., Weber, N.S., 2008. Results from a hypothesis generating case-control study: herpes family viruses and schizophrenia among military personnel. Schizophr. Bull. 34, 1182–1188.

O'Brien, S.M., Scully, P., Scott, L.V., Dinan, T.G., 2006. Cytokine profiles in bipolar affective disorder: focus on acutely ill patients. J. Affect. Disord. 90, 263–267.

Padmos, R.C., Hillegers, M.H.J., Knijff, E.M., et al., 2008. A discriminating messenger RNA signature for bipolar disorder formed by an aberrant expression of inflammatory genes in monocytes. Arch. Gen. Psychiatry 65, 395–407.

Papakostas, G.I., Shelton, R.C., Kinrys, G., et al., 2013. Assessment of a multi-assay, serum-based biological diagnostic test for major depressive disorder: a pilot and replication study. Mol. Psychiatry 18, 332–339.

Penninx, B.W.J.H., Beekman, A.T.F., Smit, J.H., et al., 2008. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int. J. Methods Psychiatr. Res. 17, 121–140.

Perdue, C.L., Cost, A.A.E., Rubertone, M.V., Lindler, L.E., Ludwig, S.L., 2015. Description and utilization of the United States department of defense serum repository: a review of published studies, 1985–2012. PLoS ONE 10, e0114857.

R Core Team, 2014. R: A Language and Environment for Statistical Computing.

Romi, F., Helgeland, G., Gilhus, N.E., 2012. Serum levels of matrix metalloproteinases: implications in clinical neurology. Eur. Neurol. 67, 121–128.

Rosenberg, G.A., Kornfeld, M., Estrada, E., Kelley, R.O., Liotta, L.A., Stetler-Stevenson, W.G., 1992. TIMP-2 reduces proteolytic opening of blood-brain barrier by type IV collagenase. Brain Res. 576, 203–207.

Rosenblat, J.D., McIntyre, R.S., 2015. Are medical comorbid conditions of bipolar disorder due to immune dysfunction? Acta Psychiatr. Scand. http://dx.doi.org/10.1111/acps.12414.

Rubertone, M.V., Brundage, J.F., 2002. The Defense Medical Surveillance System and the Department of Defense Serum Repository: glimpses of the future of public health surveillance. Am. J. Public Health 92, 1900–1904.

Schwarz, E., Guest, P.C., Rahmoune, H., et al., 2011a. Identification of a blood-based biological signature in subjects with psychiatric disorders prior to clinical manifestation. World J. Biol. Psychiatry. http://dx.doi.org/10.3109/15622975.2011.599861.

Schwarz, E., Guest, P.C., Rahmoune, H., et al., 2011b. Identification of a biological signature for schizophrenia in serum. Mol. Psychiatry. http://dx.doi.org/10.1038/mp.2011.42, published online April.

Thomas, P., Forse, R.A., Bajenova, O., 2011. Carcinoembryonic antigen (CEA) and its receptor hnRNP M are mediators of metastasis and the inflammatory response in the liver. Clin. Exp. Metastasis 28, 923–932.

Vieta, E., Berk, M., Wang, W., Colom, F., Tohen, M., Baldessarini, R.J., 2009. Predominant previous polarity as an outcome predictor in a controlled treatment trial for depression in bipolar I disorder patients. J. Affect. Disord. 119, 22–27.

Waubant, E., Goodkin, D.E., Gee, L., et al., 1999. Serum MMP-9 and TIMP-1 levels are related to MRI activity in relapsing multiple sclerosis. Neurology 53, 1397–1401.

Zetterberg, H., Jakobsson, J., Redsäter, M., et al., 2014. Blood-cerebrospinal fluid barrier dysfunction in patients with bipolar disorder in relation to antipsychotic treatment. Psychiatry Res. 217, 143–146.