Journal of Psychiatric Research 47 (2013) 1572–1580

Contents lists available at ScienceDirect

Journal of Psychiatric Research

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

Psychiatric patient stratification using biosignatures based on cerebrospinal fluid protein expression clusters

Giuseppina Maccarrone [a,1], Claudia Ditzen [a,1], Alexander Yassouridis [a], Christiane Rewerts [a], Manfred Uhr [a], Mathias Uhlen [b], Florian Holsboer [a], Christoph W. Turck [a,*]

[a] Max Planck Institute of Psychiatry, Kraepelinstr. 2, D-80804 Munich, Germany
[b] Department of Biotechnology, Albanova University Center, Royal Institute of Technology, SE-10691 Stockholm, Sweden

ARTICLE INFO

Article history:
Received 13 May 2013
Received in revised form
11 July 2013
Accepted 11 July 2013

Keywords:
Psychiatric patient group stratification
Biosignatures
Multiplex analysis
Cerebrospinal fluid
Protein levels
Protein clusters

ABSTRACT

Psychiatric disorders are caused by perturbed molecular pathways that affect brain circuitries. The identification of specific biosignatures that are the result of altered pathway activities in major depression, bipolar disorder and schizophrenia can contribute to a better understanding of disease etiology and aid in the implementation of diagnostic assays.

In the present study we identified disease-specific protein biosignatures in cerebrospinal fluid of depressed (*n*: 36), bipolar (*n*: 27) and schizophrenic (*n*: 35) patients using the Reverse Phase Protein Microarray technology. These biosignatures were able to stratify patient groups in an objective manner according to cerebrospinal fluid protein expression patterns. Correct classification rates were over 90%. At the same time several protein sets that play a role in neuronal growth, proliferation and differentiation (NEGR1, NPDC1), neurotransmission (SEZ6) and protection from oxidative damage (GPX3) were able to distinguish diseased from healthy individuals (*n*: 35) indicating a molecular signature overlap for the different psychiatric phenotypes. Our study is a first step toward implementing a psychiatric patient stratification system based on molecular biosignatures. Protein signatures may eventually be of use as specific and sensitive biomarkers in clinical trials not only for patient diagnostic and subgroup stratification but also to follow treatment response.

1. Introduction

At present the diagnosis of most psychiatric disorders is based on a personal psychiatric interview according to international guidelines and classification systems (ICD-10; DSM-IV) (World Health Organization, 2010; American Psychiatric Association, 1994). Since the differentiation of psychiatric phenotypes is exclusively based on presented clinical symptoms and does not account for biological differences, the distinction of patient subgroups remains difficult. Especially in primary care settings the correct classification of a psychiatric disorder during its first onset can be very challenging in the absence of valid and objective biomarkers. For example, a first episode of major depression could either be the beginning of recurrent MDD or represent the onset of BPD. Likewise, a severe depressive episode with psychotic symptoms could be misinterpreted as a schizophrenic episode.

The classification of mental disorders would greatly benefit from objective biological markers as part of the diagnostic process (Schwarz and Bahn, 2008). This would enable improved therapeutic intervention and could result in a reduction of psychiatric patients' relapse rates. Enhanced patient stratification would also enable personalized treatment strategies in alignment with individual patient pathophysiology. In addition, and of equal importance, the knowledge of molecular signatures, including protein expression level differences, would ultimately augment our understanding of affected molecular pathways characteristic of clinical manifestation and disease course (Ditzen et al., 2012).

It is widely accepted that mood disorders and schizophrenia are of multifactorial origin (Falkai et al., 2008; aan het Rot et al., 2009). Aside from the genetic component diverse other factors can ultimately lead to a dysfunction of central nervous system pathways. Clinical symptoms of different psychiatric disorders as well as the intensity and course of the disease are likely reflected in distinct molecular patterns. Overlapping molecular signatures on the other

* Corresponding author. Tel.: +49 89 30622 317; fax: +49 89 30622 610.
E-mail address: turck@mpipsykl.mpg.de (C.W. Turck).
[1] The authors have equally contributed.

0022-3956/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jpsychires.2013.07.021

hand may illuminate common pathophysiological pathways between different psychiatric disorders (Alaerts and Del-Favero, 2009; Prasad et al., 2010; Cichon et al., 2011; Binder and Nemeroff, 2010; Blackwood et al., 2007; Moskvina et al., 2009).

Due to its close proximity cerebrospinal fluid (CSF) was chosen for the identification of brain disorder molecular biosignatures (Zougman et al., 2008). Next to blood-derived proteins CSF also contains proteins and other molecules originating from the brain which can be used to examine pathological mechanisms associated with a psychiatric disorder (Huang et al., 2006). Specifically, the analysis of CSF proteome alterations can be exploited to distinguish between disease and health as well as different types of psychiatric disorders (Yuan et al., 2002). In previous studies we have carried out proteomic analyses of CSF from psychiatric patients that have resulted in a list of biomarker candidates (Ditzen et al., 2012; Maccarrone et al., 2004). Fifty nine proteins from this list were chosen based on literature reports that have implicated them in the etiology of mood disorders, schizophrenia and neurodegenerative disease (Table S1). The goal of the present study was to interrogate these biomarker candidates with regard to their ability to distinguish controls from patients and stratify individual disease groups.

For a sensitive and reproducible analysis of CSF proteins we have chosen the Reverse Phase Protein Microarray (RPPM) technology (Voshol et al., 2009) that allows the detection and quantification of protein analytes in a high throughput manner. Furthermore, the simultaneous measurement of a great number of patient samples enhances analytical accuracy. With the help of the RPPM technology and statistical evaluation we were able to correctly distinguish controls from diseased patients and stratify psychiatric patient groups based on CSF protein expression patterns.

2. Materials and methods

2.1. Cerebrospinal fluid specimens

Lumbar puncture was performed in sitting position at 8 am using an atraumatic needle. CSF was collected from sober patients suffering from MDD, BPD, SCZ and from healthy controls. In order to remove cellular debris, CSF samples were centrifuged, aliquoted and stored at −80 °C. Patients provided informed consent during hospital admission that their CSF can be used for scientific purposes following routine clinical analyses. The use of CSF samples for scientific studies has been approved by the *Bayerische Aerztekammer*. Samples are stored anonymously in the CSF bank of the *Max Planck Institute of Psychiatry*. For the present study CSF samples from MDD ($n = 40$), BPD ($n = 40$) and SCZ ($n = 40$) patients fulfilling ICD-10 criteria for a major depressive episode (F32/33.0–9), bipolar disorder (F31.0–9) and schizophrenia (F20.0–9), respectively, and healthy control subjects (n: 40) were selected (Supplementary Table S2). The BPD group included manic and depressed individuals. Controls represent subjects visiting the 'Neurological Outpatient Unit' at the *Max Planck Institute of Psychiatry*. Controls had unspecific and temporary complaints like vertigo or headache. For all control subjects any neurological or internal disease had been ruled out and subjects were declared "healthy". These individuals were then asked to take part in the study. After agreeing they were subjected to a psychiatric interview in order to exclude psychiatric disorders.

In addition to diagnosis, other patient information includes previous and current medication, results of physical and neurological examination and hematological, clinical chemistry as well as CSF sample laboratory data. Clinical diagnostic classification is performed by experienced psychiatrists during hospitalization of the patients. CSF samples with abnormal levels of glucose, lactate, cells, positive oligoclonal bands or disturbance of the blood–brain-barrier as well as blood-contaminated samples were not considered. Due to these strict exclusion criteria the initial number of samples for the analyses had to be reduced accordingly: MDD, $n = 36$ (age: 44 +/− 16; 56% female), BPD, $n = 27$ (age: 44 +/− 15; 56% female), SCZ, $n = 35$ (age: 37 +/− 14; 57% female), controls $n = 35$ (age: 40 +/− 16; 66% female).

2.2. Antibodies

Based on our previous studies we selected protein analytes that have been implicated in the etiology of mood disorders and SCZ. For RPPM screening 51 antibodies were provided by the *Human Protein Atlas* (HPA) program (Albanova University Center, Royal Institute of Technology, Stockholm, Sweden) (Table 1A). The polyclonal antibodies specific for the human proteins were produced by cloning and expression of 'Protein Epitope Signature Tags' (PrESTs). The experimental procedure followed for cloning, immunization and affinity purification has been described elsewhere (Agaton et al., 2003; Nilsson et al., 2005). The antibodies were provided as affinity purified reagents in Tris–HCl buffer, supplemented with 50% glycerol and 0.02% sodium azide as preservatives, and stored at −20 °C. The total protein concentration was determined by absorbance measurements at 280 nm. The binding specificity of purified antibodies was ascertained with a PrEST-array. Eight polyclonal anti-human antibodies specific for chromogranin A, cystatin C, fibronectin, glutathione peroxidase 1, glutathione peroxidase 3, presenilin-1, synapsin and synaptophysin (C-terminus) were commercially obtained. The antibodies and their suppliers are listed in Table 1B.

2.3. Immunoblot analysis

Three μg human CSF proteins were resolved by SDS-PAGE. The proteins were electroblotted onto PVDF membrane (Millipore, Schwalbach, Germany). The membranes were incubated with primary antibody (anti-fibronectin (1:1000), anti-chromogranin A (1:500) and anti-cystatin C (1:8000) in TBS-T buffer (20 mM Tris–HCl, 137 mM NaCl, 0.05% Tween 20 pH 8) for 1 h at RT, followed by a 2 h incubation at RT with a 1:2000 dilution of secondary antibody IgG-HRP (GE Healthcare, Piscataway, NJ). Proteins were visualized by enhanced chemiluminescence using the ECL reagent (GE Healthcare) and exposed to autoradiography films. For quality assessment of the commercial antibodies instructions for the immunoblots published by the manufacturers were used. Antibodies produced by the PrEST method were quality controlled by Western blot and immunohistochemistry using human tissue or cells (http://www.proteinatlas.org). Western blots for some antibodies used in the RPPM screening are depicted in Supplementary Fig. S1.

2.4. Reverse Phase Protein Microarray screening

The RPPM screening analysis was performed by NMI Technologies Transfer GmbH (NMI-TT, Reutlingen, Germany) using a protein microarray platform (Zeptosens, Witterswil, Switzerland) (Pawlak et al., 2002).

2.4.1. CSF sample preparation

Sample preparation for microarray printing was carried out by NMI-TT. CSF protein concentration was estimated using Bradford protein assay (BioRad, Hercules, CA).

CSF samples were diluted 1:2 in PBS buffer at pH 7.0 with 10% glycerol. This dilution was chosen for a final print protein concentration < 0.4 mg/ml. Two samples had to be diluted fourfold because the protein concentration was above 0.8 mg/ml. This is the upper limit of printed sample protein concentration for the RPPM method to guarantee assay signals with linear characteristics.

Table 1
A: List of polyclonal antibodies used for the Reverse Phase Protein Microarray screens of CSF samples. Antibodies were produced with the PrEST technology and were affinity purified. Ab ID: antibody; ENSG ID: Ensembl gene; WB: Western blot; Dil IHC: dilution used for immunohistochemistry; Dil RPPM: dilution used for Reverse Phase Protein Microarray. B: List of commercial antibodies used for the Reverse Phase Protein Microarray screening of CSF samples. WB: Western blot; Dil WB: dilution used for Western blot; Dil RPPM: dilution used for Reverse Phase Protein Microarray.

Ab ID	ENSG ID (idty: 100%)	Protein name	Gene name	Swiss-Prot ID	WB	Dil IHC	Dil RPPM
PRN2M	00000198898	F-actin-capping protein subunit alpha-2	CAPZA2	P47755	Yes	100	50
PRN7M	00000165119	Heterogenous nuclear ribonucleoprotein K	HNRPK	P61978	Yes	250	100
PRN15M	00000114353	Guanine nucleotide-binding protein G(i)	GNAI2	P04899	Yes	25	100
PRN18M	00000159164	Synaptic vesicle glycoprotein 2A	SV2A	Q7L0J3	Yes	150	3000
PRN20M	00000132692	Brevican core protein precursor (brain-enriched hyaluronan-binding protein)	BCAN	Q96GW7	Yes	200	1000
PRN27M	00000196743	Ganglioside GM2 activator precursor	GM2A	P17900	Yes	300	3000
PRN33M	00000095713	Cartilage acidic protein 1 precursor	CRTAC1	Q9NQ79	No	10	400
PRN34M	00000104805	Nucleobindin-1 precursor (CALNUC)	NUCB1	Q02818	Yes	250	100
PRN37M	00000106588	Proteasome subunit alpha type-2	PSMA2	P25787	Yes	200	100
PRN38M	00000107281	Neural proliferation differentiation and control protein 1 precursor	NPDC1	Q9NQX5	Yes	40	400
PRN41M	00000134265	Gamma-soluble NSF attachment protein	NAPG	Q99747	No	150	1000
PRN42M	00000136854	Syntaxin-binding protein 1	STXBP1	P61764	Yes	75	100
PRN46M	00000153956	Dihydropyridine-sensitive L-type calcium channel subunits alpha-2/delta precursor	CACNA2D1	P54289	No	75	400
PRN47M	00000138028	Cell growth regulator with EF hand domain protein 1	CGREF1	Q99674	No	100	2000
PRN50M	00000153956	Voltage-dependent calcium channel subunit alpha-2/delta-1	CACNA2D1	P54289	No	25	100
PRN54M	00000089220	Phosphatidylethanolamine-binding protein 1	PEBP1	P30086	Yes	500	2000
PRN56M	00000163531	Neurofascin precursor	NFASC	O94856	Yes	10	20
PRN58M	00000150656 (idty: 99%)	Carnosine dipeptidase 1 (metallopeptidase M20 family)	CNDP1	Q96KN2	Yes	100	20,000
PRN62M	00000153310	Protein FAM49B	FAM49B	Q9NUQ9	Yes	100	100
PRN67M	00000110436	Excitatory amino acid transporter 2	SLC1A2	P43004	Yes	50	200
PRN72M	00000170606	Heat shock 70 kDa protein 4	HSPA4	P34932	Yes	250	100
PRN73M	00000175161	Cell adhesion molecule 2	CADM2	Q8N3J6	Yes	25	200
PRN79M	00000068903	NAD-dependent deacetylase sirtuin-2	SIRT2	Q8IXJ6	No	100	200
PRN80M	00000070961	Plasma membrane calcium-transporting ATPase 1	ATP2B1	P20020	No	75	200

A

Ab ID	ENSG ID (idty: 100%)	Protein name	Gene name	Swiss-Prot ID	WB	Dil IHC	Dil RPPM
PRN81M	00000168140	Vasorin precursor	VASN	Q6EMK4	No	35	2000
PRN85M	00000156219	Ecto-ADP-ribosyltransferase 3 precursor	ART3	Q13508	No	250	2000
PRN88M	00000135046	Annexin A1	ANXA1	P04083	Yes	1250	250
PRN89M	00000135046	Annexin A1	ANXA1	P04083	Yes	125	50
PRN90M	00000165868	Heat shock 70 kDa protein 12A	HSPA12A	O43301	Yes	100	50
PRN93M	00000063015	Seizure related 6 homolog	SEZ6	Q53EL9	Yes	25	200
PRN94M	00000145642	Membrane protein FAM159B	FAM159B	A6NKW6	Yes	25	250
PRN95M	00000050165	Dickkopf-related protein 3 precursor	DKK3	Q9UBP4	Yes	100	8000
PRN99M	00000171951	Secretogranin-2 precursor	SCG2	P13521	Yes	375	150
PRN100M	00000172260	Neuronal growth regulator 1 precursor.	NEGR1	Q7Z3B1	Yes	150	50
PRN103M	00000159082	Synaptojanin-1	SYNJ1	O43426	Yes	125	200
PRN104m	00000162729	Immunoglobulin superfamily member 8 precursor	IGSF8	Q969P0	Yes	15	100
PRN116M	00000089199	Secretogranin-1 precursor	CHGB	P05060	Yes	250	250
PRN117M	00000105767	Cell adhesion molecule 4 precursor	CADM4	Q8NFZ8	Yes	350	150
PRN120M	00000115306	Spectrin beta chain, brain 1	SPTBN1	Q01082	Yes	75	200
PRN121M	00000126861	Oligodendrocyte myelin glycoprotein	OMG	P23515	Yes	100	400
PRN124M	00000165280	Transitional endoplasmic reticulum ATPase	VCP	P55072	Yes	600	200
PRN130M	00000078018	Microtubule-associated protein 2	MAP2	P11137	Yes	80	40
PRN135M	00000070961	Plasma membrane calcium-transporting ATPase	ATP2B1	P20020	Yes	50	50
PRN140M	00000166922	Neuroendocrine protein 7B2 precursor	SCG5	P05408	Yes	150	100
PRN142M	00000181195	Proenkephalin A precursor	PENK	P01210	Yes	35	100
PRN146M	00000153446	UPF0764 protein C16orf89	C16orf89	Q6UX73	Yes	50	250
PRN156M	00000159720	Vacuolar ATP synthase subunit d 1	ATP6V0D1	P61421	Yes	25	1000
PRN74M	00000149218	Endonuclease domain-containing 1 protein	ENDOD1	O94919	Yes	650	250
PRN148M	00000214851	Putative uncharacterized protein C12orf33	C12orf33	Q8N6U2	Yes	150	150
PRN149M	00000106278	Receptor-type tyrosine-protein phosphatase zeta precursor	PTPRZ1	P23471	No	150	600
PRN150M	00000174684	N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase	B3GNT1	O43505	No	35	1000

B

Antibody ID	Protein name	Gene name	Swiss-Prot ID	Supplier	WB	Dil WB	Dil RPPM
PRNFNm	Fibronectin	FN1	P02751	Sigma (Germany)	Yes	1000	1,000,000
PRNCAm	Chromogranin A	CHGA	P10645	Abcam (Germany)	Yes	500	250
PRNGP1m	Glutathione peroxidase 1	GPX1	P07203	Abcam (Germany)	Yes	1000	200
PRNGP3m	Glutathione peroxidase 3	GPX3	P22352	Aviva Systems Biology (USA)	Yes	1000	200
PRNPRm	Presenilin-1	PSEN1	P49768	Cell Signaling (USA)	Yes	500	50
PRNRCm	Cystatin C	CST3	P01034	Abcam (Germany)	Yes	8000	1,000,000
PRNSYm	Synapsin	SYN1	P17600	Cell Signaling (USA)	Yes	1000	50
PRNSPm	Synaptophysin	SYP	P08247	Epitomics (USA)	Yes	10,000	200

2.4.2. Array printing

Diluted CSF samples were printed in duplicate onto a ZeptoMARK protein chip (Zeptosens) which consists of 6 microarrays. Each microarray includes 320 spots, a sufficient number to perform all experiments. Together with the CSF samples, reference fluorescence-labeled protein (Cy5-labeled bovine serum albumin, conjugated in house) was co-arrayed into three separate rows of landing marks. The microarrays were printed under clean-room conditions, by depositing a single drop with a volume of 400 pl using a NanoPlotter NP2 piezoelectric arrayer (GeSim GmbH, Grosserkmannsdorf, Germany). After spotting, the microarrays were blocked with 3% (w/v) bovine serum albumin (BSA), washed with distilled water, dried under a nitrogen stream and then stored in the dark at 4 °C until further use.

For quality control blank assays were prepared on separate arrays. The CSF sample spots were arrayed under identical conditions without the marker-specific primary antibodies to assess non-specific signal contributions caused by the secondary antibody.

2.4.3. Assay procedure

The detection of the respective protein analyte in the CSF sample spots was performed using a direct two-step sequential immunoassay. In the first step an analyte-specific antibody in assay buffer (50 mM imidazole–HCl, 150 mM NaCl, 0.1% Tween 20, 0.005% sodium azide, 5% (w/v) BSA, pH 7.4) was added onto the dried RPP microarray at the indicated dilution followed by an overnight incubation at RT (25 °C). Two antibody dilutions were used to assess the 133 CSF sample protein expression profiles (Table 1). After excess primary antibody removal by washing with assay buffer, the microarrays were incubated with secondary, Cy5-labeled anti-rabbit IgG (Jackson Immuno Research, West Grove, PA, USA), 1500-fold diluted from the stock solution in assay buffer to a final concentration of 0.5 μg/ml, for 1 h at 25 °C in the dark.

Subsequently the arrays were washed and imaged in assay buffer with the ZeptoREADER® imager instrument (Zeptosens). Fluorescence emission readout was performed at excitation/emission wavelengths of 635/670 nm, red detection channels with an exposure time of 0.5–16 s. For each microarray six fluorescence images at different exposure times were recorded and the resulting images analyzed with the ZeptoVIEW™ PRO software (Zeptosens).

2.4.4. RPPM image analysis

Microarray images were selected at the highest exposure time without reaching the saturated pixel. The spot diameter of the array analysis grid was set constant at 130 μm and cross-aligned to the microarrays. In order to analyze the array images the background-corrected and the "Referenced Fluorescent Intensity" mean signals (RFI) were calculated for each sample spot. The background-corrected RFI was calculated as the ratio of the sample spot signal and the local reference signal. The RFIs of each duplicate spot pair were averaged and generated the "Mean Fluorescence Intensity" signals (MFI) for each CSF sample (raw signal). The determined errors correspond to the variations of the duplicates across the array (coefficient variation 3%). The blank-corrected MFI values were normalized for the protein content in the spot resulting in the normalized fluorescent intensity (NFI) values. The NFI signal values were used for statistical analysis. Protein amount estimated by Bradford assay was used for the normalization of the raw MFI.

2.4.5. Statistical analysis of RPPM data

The statistical analysis was performed using the protein array NFI values generated from four study groups [1: BPD ($n = 27$); 2: MDD ($n = 36$); 3: SCZ ($n = 35$); 4: control ($n = 35$)]. Because inferential statistics on a large number of variables requires among other prerequisites also a strong level of significance correction that would make the identification of relevant proteins very difficult, we performed the statistical analysis as follows. We first applied the explorative 'Measure of Relevance' (MoR) method (Yassouridis et al., 2012) in order to identify the most informative proteins, i.e. those proteins that can characterize individual disease groups. The explorative MoR method, which is based on a measure containing relevant information of the distribution form, location and dispersion parameters of the samples, enables reasonable reduction of data dimensionality without the need for test decisions, corrections of significance levels and other presumptions. To further improve the results we followed up with a reliability investigation by applying the MoR method to smaller study groups. For the purpose of demonstrating reproducibility 500 different sub-samples were used with patients ($n = 22$) randomly chosen from each group. Those variables among the 500 repetitions proven to be informative with a frequency over 0.75 and 0.90, defined as reliable and highly reliable biomarker candidates, respectively, are able to distinguish between two study groups. Inferential statistics using multivariate analysis of covariance (MANCOVA) with age and sex as covariates were then performed on the reliable biomarker candidates. Those biomarkers that pointed to significant differences between two or more disease groups in MANCOVA, after Bonferroni adjustments, were designated as most significant biomarkers. To illustrate the discrimination power of the reliable and highly reliable biomarkers (frequency: 0.75 and 0.90, respectively) they were subjected to a discriminant analysis in a final step.

3. Results

The present study was performed with the goal to identify patient group specific biosignatures for MDD, BPD, SCZ and controls based on CSF protein level differences. After evaluation of the CSF samples with regard to cytological features, oligoclonal bands, glucose, lactate and total protein levels as well as albumin ratios, RPPM analyses were performed as outlined in Fig. 1. CSF expression levels for the 59 proteins listed in Supplementary Table S1 were measured. For each antibody six array images were acquired and analyzed and the assay with the best MFI signal to background ratio from two screens was selected. The assays showed low background signals and homogenous spot morphologies (except for the CRPMP5 antibody). In most cases the best assay conditions were at high antibody dilution (Table 1A). The NFIs of the RPPM data were then statistically analyzed using the MoR method (Yassouridis et al., 2012) and resulted in 33 proteins that were able to distinguish between at least two study groups. The reliability of these biomarker candidates was ascertained by running 500 MoR method repeat cycles with random sample selection and $n = 22$ items per group. From the 33 biomarker candidates the reliability investigation identified 24 proteins as reliable biomarker candidates (Table 2).

Fig. 2 depicts the results of the reliability investigation and lists reliable and highly reliable biomarker candidates with relative frequencies over 75% and 90%, respectively. Among the 59 proteins 24 (CHGA; NEGR1; STXBP1; FN1; GM2A; ATP6V0D1; CRTAC1, OMG; CNDP1; SPTBN1; SEZ6; ATP2B1; CADM2; NPDC1; SCG5; NAPG; GPX3; PSMA2; FAM159B; DKK3; GNAI2; PTPRZ1; CAPZA2; SYNJ1) represent reliable biomarker candidates and among those 16 (CRTAC1, OMG; CNDP1; SPTBN1; SEZ6; ATP2B1; CADM2; NPDC1; SCG5; NAPG; GPX3; DKK3; CAPZA2; GNAI2; PTPRZ1; NEGR1) are highly reliable biomarker candidates. Table 2 shows means and standard error (SEMs) of the NFI values of all reliable biomarker proteins within each group; highly reliable biomarker candidates are highlighted in bold. The magnitude of CSF protein level differences is depicted in Fig. 3. The 24 reliable protein biomarker candidates were subsequently subjected to MANCOVA with age and sex as covariates to proof by inferential means the capability of

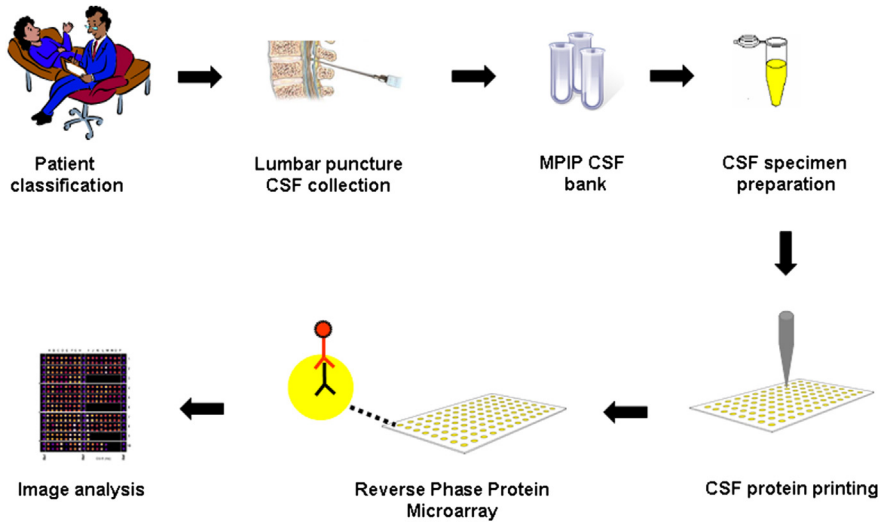

Fig. 1. Workflow of the experimental steps from patient interview to the CSF Reverse Phase Protein Microarray analysis.

these proteins to distinguish between the disease groups. MANCOVA revealed a significant group effect [Wilks multivariate tests of significance: $F(72,312) = 8.28$, sig. of $F < 0.0001$] for all reliable protein biomarker candidates (Table 2; univariate *F*-tests, $p < 0.0001$). Therefore, all reliable biomarker candidates identified by the MoR method and the subsequent reliability investigation were designated as *significant* biomarkers. Interestingly, the covariates sex and age showed also significant associations [effects of the covariates age and sex: $F(48,208) = 2.32$, sig of $F < 0.0001$] for most of the reliable biomarker proteins. Since in a MANCOVA factor effects, in our case disease group effects, on the dependent variables, in our case reliable protein biomarker candidates, are adjusted from covariate effects, we can claim that the significant group differences identified by the reliable protein biomarker candidates are solely caused by the disease, irrespective of age and sex of the patients. However, most of the reliable proteins are only able to distinguish diseased individuals from controls (Table 2), whereas only a subset of the candidates has the power to distinguish between two disease groups. FN1, GM2A, CHGA and DKK3 are proteins that can distinguish between SCZ and BPD, CHGA also distinguishes between MDD and SCZ patients. Likewise, only a few proteins can differentiate individual psychiatric disorders from controls (SYNJ1, PTPRZ1, ATP6V0D1, PSMA2, STXBP1 and GNAI2) (Table 2). Discriminant analysis performed on the 24 reliable and 16 highly reliable proteins revealed for the four groups a total correct classification rate of 90.23% and 85.7%, respectively. Fig. 4 reports the topographic distribution of the four study groups resulting from the discrimination analysis of the 24 reliable and significant biomarker proteins. The correct classification rates for the BPD, MDD, SCZ and C groups were 92.6%, 83.3%, 100% and 85.7%, respectively. Applying the 24 significant biomarkers to discrimination analyses of disease group vs. control pairs, resulted in the following correct classification rates: 100% and 100% for BPD and controls, respectively; 100% and 97% for MDD and controls, respectively; 100% and 91% for SCZ and controls, respectively. We therefore conclude that our method results in high sensitivity and specificity values.

4. Discussion

The pathophysiology of psychiatric disorders remains elusive (Falkai et al., 2008; aan het Rot et al., 2009). The long prevailing monoamine hypothesis that forms the basis for the majority of currently used antidepressant drugs cannot account for the rather long delay of a therapeutic effect observed in psychiatric patients (Gjerris, 1988; De Bellis et al., 1993). Alterations in neuroanatomical structures and the immune and neuroendocrine systems seem to indicate that other mechanisms are likely involved in psychiatric disorders (Bremner et al., 2000; Meisenzahl et al., 2008; Carpenter et al., 2004; Holsboer and Ising, 2010). Although a great number of genetic studies have been performed major risk factors for the development of a psychiatric disorder have not been identified (Alaerts and Del-Favero, 2009; Prasad et al., 2010; Cichon et al., 2011; Binder and Nemeroff, 2010; Blackwood et al., 2007). All this has severely hampered efforts to come up with more precise diagnostic tools.

In the present study we have searched for biosignatures based on CSF protein level differences from three psychiatric patient groups and compared them to controls. We have been able to delineate shared and disease-specific CSF protein signatures. Patient group stratification was achieved at a 90.23% correct classification rate. The best discrimination power (nearly 100%) was observed for the SCZ patient group. The majority of expression level differences was observed between healthy and diseased individuals (Table 2). They include proteins that have been reported to play a role in neuronal growth, proliferation and differentiation (NEGR1, NPDC1), neurotransmission (SEZ6) as well as protection from oxidative damage (GPX3).

Although expression level variations were also found between disease groups most of the differences only reached significance when compared to the control group. This finding could be a reflection of shared molecular pathway alterations among psychiatric disorders and supports the notion that they in fact represent a continuum of overlapping phenotypes as has been recently suggested based on a large scale genetic study (Cross-Disorder Group of the Psychiatric Genomics Consortium et al., 2013). This finding questions the validity of the currently used ICD-10 and DSM-IV classification systems for psychiatric disorders. Based on our CSF biosignature results only a selected few proteins exhibit expression level differences able to discriminate between disease groups. These include CHGA (pituitary secretory protein I), DKK3, fibronectin and GM2A. DKK3, GM2A and fibronectin were able to differentiate BPD from SCZ patients. DKK3 is an inhibitor of the Wnt signaling pathway that is important for neuronal development and axon guidance (Krupnik et al., 1999). GM2A is a protein linked to Tay-Sachs disease, a disorder characterized by the accumulation of

Table 2
List of biomarkers with relative frequency over 75% and over 90% (bold font). Group means and standard error (SEMs) of the corresponding NFI values are shown (BPD: bipolar disorder, MDD: major depressive disorder, SCZ: schizophrenia, C: control). Significant group effects (p-values) on the biomarker proteins were identified at a Bonferroni-corrected level of significance by univariate F-tests in a multivariate analysis of covariance (MANCOVA) with age and sex as covariates. Pairs of groups with significant differences were identified by Bonferroni Post-Hoc tests.

Protein name	Gene code	Swiss-Prot ID	Significant differences between groups	Group effects: p-values	BPD (group 1)		MDD (group 2)		SCZ (group 3)		Controls (group 4)	
					Mean (NFI, n: 27)	SEM	Mean (NFI, n: 36)	SEM	Mean (NFI, n: 35)	SEM	Mean (NFI, n: 35)	SEM
Neuronal growth regulator 1	**NEGR1**	**Q7Z3B1**	**1 vs 4 2vs 4**	**<0.0001**	**3.50**	**0.26**	**3.66**	**0.15**	**3.36**	**0.16**	**2.72**	**0.24**
Synaptojanin-1	SYNJ1	O43426	3 vs 4	<0.0001	4.54	0.37	5.29	0.28	5.66	0.29	4.13	0.25
Spectrin beta chain, brain 1	**SPTBN1**	**Q01082**	**1 vs 4 2 vs 4 3 vs 4**	**<0.0001**	**14.36**	**1.13**	**16.62**	**0.82**	**16.93**	**0.86**	**7.89**	**0.99**
Oligodendrocyte myelin glycoprotein	**OMG**	**P23515**	**1 vs 4 2 vs 4 3 vs 4**	**<0.0001**	**7.14**	**0.47**	**8.84**	**0.41**	**9.35**	**0.47**	**17.19**	**1.25**
Neuroendocrine protein 7B2	**SCG5**	**P05408**	**1 vs 4 2 vs 4 3 vs 4**	**<0.0001**	**5.66**	**0.40**	**6.32**	**0.27**	**6.06**	**0.31**	**3.71**	**0.36**
Receptor-type tyrosine-protein phosphatase zeta	**PTPRZ1**	**P23471**	**1 vs 4**	**<0.0001**	**6.42**	**0.49**	**7.33**	**0.32**	**7.31**	**0.36**	**9.15**	**0.44**
Vacuolar ATP synthase subunit d 1	ATP6V0D1	P61421	1 vs 4	=0.0001	4.40	0.38	5.04	0.31	6.07	0.36	6.75	0.38
Guanine nucleotide-binding protein G(i), alpha-2 subunit	**GNAI2**	**P04899**	**2 vs 4**	**<0.0001**	**1.85**	**0.14**	**1.97**	**0.08**	**1.77**	**0.08**	**1.48**	**0.10**
Ganglioside GM2 activator	GM2A	P17900	1 vs 3	=0.0001	7.28	0.48	7.08	0.36	5.30	0.23	7.64	0.68
F-actin-capping protein subunit alpha-2	**CAPZA2**	**P47755**	**2 vs 4 1 vs 4**	**<0.0001**	**3.48**	**0.28**	**3.83**	**0.21**	**4.25**	**0.23**	**5.50**	**0.28**
Cartilage acidic protein 1	**CRTAC1**	**Q9NQ79**	**1 vs 4 2 vs 4 3 vs 4**	**<0.0001**	**5.38**	**0.36**	**5.21**	**0.22**	**4.96**	**0.26**	**13.22**	**1.25**
Proteasome subunit alpha type-2	PSMA2	P25787	3 vs 4	<0.0001	8.56	0.69	8.49	0.35	7.89	0.40	10.81	0.54
Neural proliferation differentiation and control protein 1	**NPDC1**	**Q9NQX5**	**1 vs 4 2 vs 4 3 vs 4**	**<0.0001**	**6.02**	**0.48**	**6.08**	**0.24**	**5.97**	**0.30**	**9.29**	**0.54**
Gamma-soluble NSF attachment protein	**NAPG**	**Q99747**	**1 vs 4 2 vs 4 3 vs 4**	**<0.0001**	**6.28**	**0.50**	**6.84**	**0.28**	**6.66**	**0.32**	**9.83**	**0.48**
Syntaxin-binding protein 1	STXBP1	P61764	2 vs 4	<0.0001	6.46	0.44	6.35	0.24	5.83	0.26	5.04	0.36
Beta-Ala-His dipeptidase	**CNDP1**	**Q96KN2**	**1 vs 4 2 vs 4 3 vs 4**	**<0.0001**	**5.03**	**0.61**	**5.10**	**0.63**	**2.84**	**0.24**	**10.83**	**0.89**
Cell adhesion molecule 2	**CADM2**	**Q8N3J6**	**1 vs 4 2 vs 4 3 vs 4**	**<0.0001**	**9.19**	**0.70**	**10.43**	**0.47**	**10.31**	**0.51**	**5.86**	**0.63**
Plasma membrane calcium-transporting ATPase 1	**ATP2B1**	**P20020**	**2 vs 4 3 vs 4**	**<0.0001**	**2.50**	**0.24**	**2.73**	**0.15**	**3.10**	**0.18**	**1.83**	**0.15**
Seizure related 6 homolog	**SEZ6**	**Q53EL9**	**1 vs 4 2 vs 4 3 vs 4**	**<0.0001**	**10.32**	**0.71**	**11.17**	**0.46**	**12.29**	**0.55**	**5.97**	**0.68**
Membrane protein FAM159B	FAM159B	A6NKW6	2 vs 4 3 vs 4	<0.0001	18.53	1.36	19.02	0.78	19.02	0.86	14.37	0.96
Dickkopf-related protein 3 precursor	**DKK3**	**Q9UBP4**	**1 vs 3 3 vs 4**	**<0.0001**	**7.26**	**0.97**	**5.65**	**0.38**	**3.74**	**0.17**	**9.49**	**1.56**
Glutathione peroxidase 3	**GPX3**	**P22352**	**1 vs 4 2 vs 4 3 vs 4**	**<0.0001**	**6.28**	**0.47**	**6.38**	**0.24**	**5.69**	**0.25**	**4.31**	**0.35**
Chromogranin A	CHGA	P10645	1 vs 4 2 vs 4 3 vs 4	<0.0001	0.87	0.07	0.73	0.05	0.52	0.03	0.98	0.10
Fibronectin	FN1	P02751	1 vs 3	=0.0018	19.46	1.64	17.12	1.21	13.12	0.77	18.44	1.74

GM2 gangliosides in neuronal cells (Mahuran, 1999). We found the levels of both proteins to be reduced in SCZ patients' CSF. This is in line with an earlier study where decreased DKK3 mRNA levels in superior temporal cortex were detected in SCZ individuals (Ftouh et al., 2005).

CHGA protein levels were able to distinguish MDD from SCZ patients. CHGA belongs to the granin protein family with important roles in secretory pathways responsible for controlled delivery of neurotransmitters and growth factors (Bartolomucci et al., 2011).

Our CSF RPPM results were unable to distinguish the BPD and MDD patient groups. This could be due to the fact that the analyzed proteins are state rather than trait markers and are indicative of a depressive episode. In this context genetic studies have also failed to provide a clear distinction between BPD and MDD classifications and a long-term study showed that the conversion rate from unipolar to bipolar disorder is 1.25% throughout the life span (McMahon and Insel, 2012; Angst et al., 2005). The majority of the BPD patients included in the present study were suffering from a

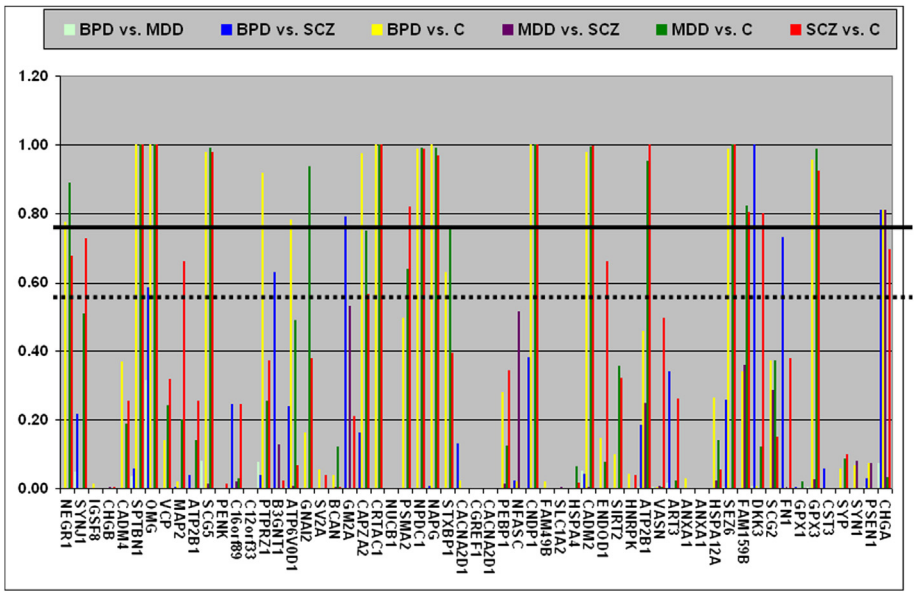

Fig. 2. Reliability investigation: plot of relative frequencies of proteins showing relevant 'Measure of Relevance' (MoR) scores among 500 repeats of random sample selection for $n = 22$ items per group (BPD: bipolar disorder, MDD: major depressive disorder, SCZ: schizophrenia, C: control). The horizontal dashed and solid lines indicate relative frequencies of 0.75 and 0.90, respectively.

depressive episode at the time of CSF collection, which may explain our inability to distinguish the two groups.

Although age and sex correlate significantly with almost all reliable proteins, MANCOVA revealed that these covariates did not influence the group effects on these proteins. This was an important finding since our analysis was limited by a slight gender difference between patients and controls (13% more females in the control group). Another limitation of the study is the fact that the majority of patients were medicated. Since psychiatric patients are usually only admitted to a hospital after several outpatient treatment trials have failed, most of them are already medicated at the time of admission. It is conceivable that variables like treatment nature or number of different medications could influence the group effects on the reliable biomarker proteins. In order to address this issue we applied MANCOVA to these proteins with type and number of drugs as additional covariates besides age and sex, which does not result in a change of our findings. We also find a significant group effect [Wilks multivariate tests of significance: $F(72,272) = 6.97$, sig. of $F < 0.0001$] that is remarkably significant at the corrected level of significance for all reliable protein biomarker candidates (univariate F-tests, $p < 0.0001$) except for PTPR21 (Table 2) that only showed marginal significance at the Bonferroni-corrected significance level. The complementary covariates did not show significant associations with the reliable biomarker proteins and therefore did not notably affect the group differences. Although nominal or categorized covariates could also be used with the MoR method to determine the variability they account for, we instead considered covariates in the variance analysis to make statements about their effects in terms of inferential and not explorative statistics.

Due to the difficulty of obtaining a sufficient number of CSF samples for a second independent analysis we performed a reliability analysis. For each of the original samples 500 different subsamples were used. Although this does not fulfill all the requirements for a true replication (the sub-samples in the reliability analysis are not distinct and independent) we submit that it adds to the robustness of our results. This was shown by performing additional permutation t-tests using Monte Carlo samplings of originally 10,000 permutations between group pairs. With this

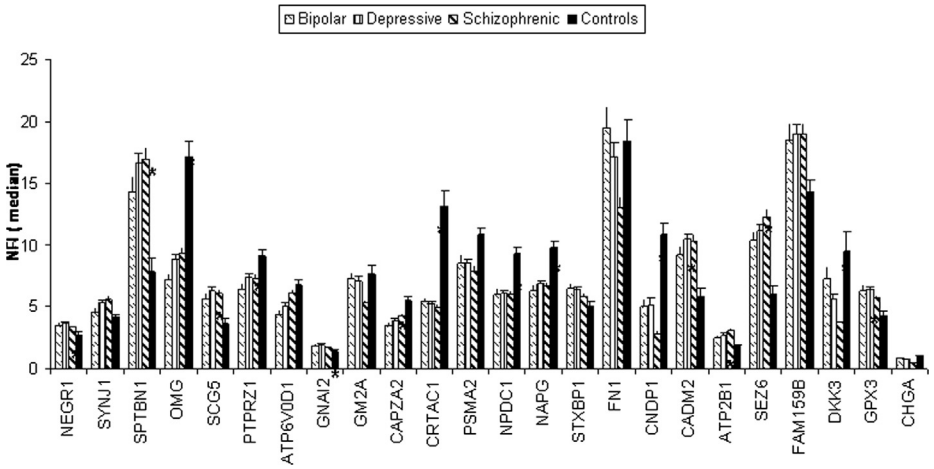

Fig. 3. Normalized fluorescent intensity (NFI) median value (+/−SEM) plot of proteins with relevant 'Measure of Relevance' (MoR) scores within the various disease groups. The protein biomarker candidates have a relative frequency of 75% after 500 repeats of random sample selection ($n = 22$ items for each group).

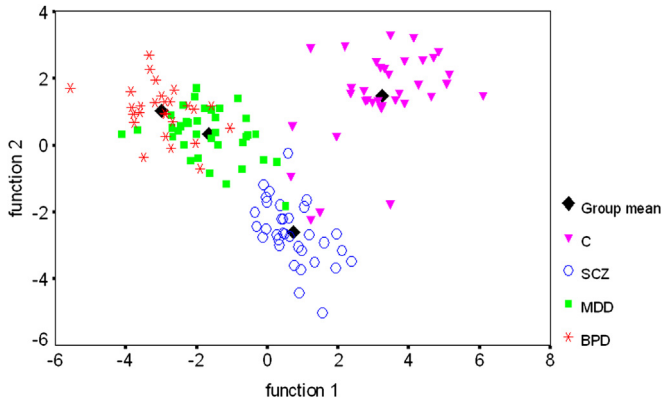

Fig. 4. Topographic distribution of group items (BPD: bipolar disorder, MDD: major depressive disorder, SCZ: schizophrenia, C: control) by applying discriminant function analysis to the identified protein biomarkers with 'Measure of Relevance' (MoR) reliability over 75%.

method we obtained as significant classifiers the same proteins that had been identified as reliable biomarkers with the MoR method followed by the reliability analysis. However, whereas the significant classifiers of the permutation approach were identified on the confirmatory level, the reliable biomarkers of the MoR method and the reliability analysis were identified on the explorative level and can be further used for inferential methods like MANCOVA.

Ultimately, it is our goal to extend the analyses to blood as the preferred peripheral specimen in the clinical laboratory. Since CSF and blood components are exchanged there is a good chance that the protein alterations we have identified in CSF will also be detectable in blood, albeit at lower levels. Our study is a first step toward implementing a more precise psychiatric patient stratification system based on molecular biosignatures. In combination with genetics, metabolomics, EEG and other data psychiatric medicine may eventually realize an individual treatment routine with great benefits for the patient (Holsboer, 2008). Protein signatures may also serve as biomarkers in clinical trials not only for patient stratification but also to follow treatment response.

Conflicts of interest

The authors declare to have no conflict of interest.

Role of funding source

This study was supported by NGFNplus MooDS funding and the *Max Planck Society*. The funding sources had no further role in study design, collection, analysis and interpretation of data, in writing of the report and in the decision to submit the paper for publication.

Contributors

Maccarrone: Designed the protocol, carried out Western blot image searches and checked antibodies produced by the PrEST method; wrote part of the manuscript, prepared figures and tables; interpreted NMI data.

Ditzen: Carried out literature searches and analyses; wrote part of the manuscript; prepared tables; analyzed the clinical data.

Yassouridis: undertook the statistical analysis; prepared figures.

Rewerts: technical assistance.

Uhr: responsible for the CSF bank of the *Max Planck Institute of Psychiatry*; managed sample collection and clinical data analysis.

Uhlen: provided the antibodies produced by PrEST from Human Protein Atlas (HPA) program.

Holsboer: initiated the study and edited the manuscript.

Turck: Designed the study, edited the manuscript before submission.

All authors contributed to and have approved the final manuscript.

Acknowledgments

This study was supported by NGFNplus MooDS funding and the *Max Planck Society*.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jpsychires.2013.07.021.

References

aan het Rot M, Mathew SJ, Charney DS. Neurobiological mechanisms in major depressive disorder. Canadian Medical Association Journal 2009;180:305–13.

Agaton C, Galli J, Hoiden Guthenberg I, Janzon L, Hansson M, Asplund A, et al. Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. Molecular Cellular Proteomics 2003;2:405–14.

Alaerts M, Del-Favero J. Searching genetic risk factors for schizophrenia and bipolar disorder: learn from the past and back to the future. Human Mutation 2009;30:1139–52.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed. Washington, DC: American Psychiatric Press; 1994.

Angst J, Sellaro R, Stassen HH, Gamma A. Diagnostic conversion from depression to bipolar disorders: results of a long-term prospective study of hospital admissions. Journal of Affective Disorders 2005;84:149–57.

Bartolomucci A, Possenti R, Mahata SK, Fischer-Colbrie R, Loh YP, Salton SR. The extended granin family: structure, function, and biomedical implications. Endocrine Reviews 2011;32:755–97.

Binder EB, Nemeroff CB. The CRF system, stress, depression and anxiety-insights from human genetic studies. Molecular Psychiatry 2010;15:574–88.

Blackwood DH, Pickard BJ, Thomson PA, Evans KL, Porteous DJ, Muir WJ. Are some genetic risk factors common to schizophrenia, bipolar disorder and depression? Evidence from DISC1, GRIK4 and NRG1. Neurotoxicity Research 2007;11:73–83.

Bremner JD, Narayan M, Anderson ER, Staib LH, Miller HL, Charney DS. Hippocampal volume reduction in major depression. American Journal of Psychiatry 2000;157:115–8.

Carpenter LL, Heninger GR, Malison RT, Tyrka AR, Price LH. Cerebrospinal fluid interleukin (IL)-6 in unipolar major depression. Journal of Affective Disorders 2004;79:285–9.

Cichon S, Mühleisen TW, Degenhardt FA, Mattheisen M, Miró X, Strohmaier J, et al. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. American Journal of Human Genetics 2011;88:372–81.

Cross-Disorder Group of the Psychiatric Genomics Consortium, Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet 2013 Apr 20;381(9875):1371–9. http://dx.doi.org/10.1016/S0140-6736(12)62129-1.

De Bellis MD, Geracioti Jr TD, Altemus M, Kling MA. Cerebrospinal fluid monoamine metabolites in fluoxetine-treated patients with major depression and in healthy volunteers. Biological Psychiatry 1993;33:636–41.

Ditzen C, Tang N, Jastorff AM, Teplytska L, Yassouridis A, Maccarrone G, et al. Cerebrospinal fluid biomarkers for major depression confirm relevance of associated pathophysiology. Neuropsychopharmacology 2012;37:1013–25.

Falkai P, Mike O, Inez MG, Paul H, Andras BG, Sophia F, et al. A roadmap to disentangle the molecular etiology of schizophrenia. European Psychiatry 2008;23:224–32.

Ftouh S, Akbar MT, Hirsch SR, de Belleroche JS. Down-regulation of Dickkopf 3, a regulator of the Wnt signalling pathway, in elderly schizophrenic subjects. Journal of Neurochemistry 2005;94:520–30.

Gjerris A. Baseline studies on transmitter substances in cerebrospinal fluid in depression. Acta Psychiatrica Scandinavica 1988;346(Suppl.):1–35.

Holsboer F. How can we realize the promise of personalized antidepressant medicines? Nature Reviews Neuroscience 2008;9:638–46.

Holsboer F, Ising M. Stress hormone regulation: biological role and translation into therapy. Annual Review of Psychology 2010;61:81–109. C1–11.

Huang JT, Leweke FM, Oxley D, Wang L, Harris N, Koethe D, et al. Disease biomarkers in cerebrospinal fluid of patients with first-onset psychosis. PLOS Medicine 2006 Nov;3(11):e428.

Krupnik VE, Sharp JD, Jiang C, Robison K, Chickering TW, Amaravadi L, et al. Functional and structural diversity of the human Dickkopf gene family. Gene 1999;238:301–13.

Maccarrone G, Birg I, Malisch E, Rosenhagen MC, Ditzen C, Chakel JA, et al. In-depth analysis of the human CSF proteome using protein prefractionation. Clinical Proteomics 2004;1:333–64.

Mahuran DJ. Biochemical consequences of mutations causing the GM2 gangliosidoses. Biochimica et Biophysica Acta 1999;1455:105–38.
McMahon FJ, Insel TR. Pharmacogenomics and personalized medicine in neuropsychiatry. Neuron 2012;74:773–6.
Meisenzahl EM, Koutsouleris N, Bottlender R, Scheuerecker J, Jäger M, Teipel SJ, et al. Structural brain alterations at different stages of schizophrenia: a voxel-based morphometric study. Schizophrenia Research 2008;104:44–60.
Moskvina V, Craddock N, Holmans P, Nikolov I, Pahwa JS, Green E, et al. Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. Molecular Psychiatry 2009;14:252–60.
Nilsson P, Paavilainen L, Larsson K, Odling J, Sundberg M, Andersson AC, et al. Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. Proteomics 2005;5:4327–37.
Pawlak M, Schick E, Bopp MA, Schneider MJ, Oroszlan P, Ehrat M. Zeptosens' protein microarrays: a novel high performance microarray platform for low abundance protein analysis. Proteomics 2002;2:383–93.
Prasad KM, Talkowski ME, Chowdari KV, McClain L, Yolken RH, Nimgaonkar VL. Candidate genes and their interactions with other genetic/environmental risk factors in the etiology of schizophrenia. Brain Research Bulletin 2010;83:86–92.
Schwarz E, Bahn S. Cerebrospinal fluid: identification of diagnostic markers for schizophrenia. Expert Review of Molecular Diagnostics 2008 Mar;8(2):209–16. http://dx.doi.org/10.1586/14737159.8.2.209.
Voshol H, Ehrat M, Traenkle J, Bertrand E, van Oostrum J. Antibody-based proteomics: analysis of signaling networks using reverse protein arrays. FEBS Journal 2009;276:6871–9.
World Health Organization. International statistical classification of diseases and related health problems; 2010. 10th Revision.
Yassouridis A, Ludwig T, Steiger A, Leisch F. A new way of identifying biomarkers in biomedical basic-research studies. PLOS ONE 2012;7:e35741.
Yuan X, Russell T, Wood G, Desiderio DM. Analysis of the human lumbar cerebrospinal fluid proteome. Electrophoresis 2002;23:1185–96.
Zougman A, Pilch B, Podtelejnikov A, Kiehntopf M, Schnabel C, Kumar C, et al. Integrated analysis of the cerebrospinal fluid peptidome and proteome. Journal of Proteome Research 2008;7:386–99.