Ph.D. Dissertation of Medicine

Differential diagnosis of major psychiatric disorders using mass spectrometry-based blood proteomic analysis

질량 분석기 기반 혈액 단백체 분석을 이용한 주요정신질환의 감별진단

August 2022

Graduate School of Medicine
Seoul National University
Psychiatry Major

Sang Jin Rhee

Differential diagnosis of major psychiatric disorders using mass spectrometry-based blood proteomic analysis

질량 분석기 기반 혈액 단백체 분석을 이용한 주요정신 질환의 감별진단

Name of Examiner: Yong Min Ahn

Submitting a Ph.D. Dissertation of
Medicine

April 2022

Graduate School of Medicine
Seoul National University
Psychiatry Major

Sang Jin Rhee

Confirming the Ph.D. Dissertation written by
Sang Jin Rhee

July 2022

Chair _____(Seal)
Vice Chair _____(Seal)
Examiner _____(Seal)
Examiner _____(Seal)
Examiner _____(Seal)

Abstract

Sang Jin Rhee
Department of Medicine (Psychiatry)
The Graduate School
Seoul National University

Background: Major depressive disorder (MDD), bipolar disorder (BD), and schizophrenia (SCZ) are representative major psychiatric disorders that are known to be associated with life-long disability and mortality. These disorders are difficult to distinguish, as their diagnosis is based on subjective symptoms and behavioral observations. Recent studies suggest that profiling and targeted quantification of proteomes might help in objective differentiation between these disorders. Thus, this study was conducted to compare and differentiate these disorders based on the quantification of peripheral proteins.

Methods: Mass spectrometry-based proteomic profiling analysis was performed on serum samples from psychotropic drug-free 15 MDD and 10 BD patients. *T*-tests were performed with pairwise comparisons to detect differentially expressed proteins (DEPs) (Study 1). The study was expanded to plasma samples of 174 MDD, 170 BD, 171 SCZ, and 160 healthy controls Both targeted proteomics and proteomic profiling were performed to quantify and verify proteomic candidate targets that differentiated these disorders. Through repeated LASSO regression with feature extraction and weighted

model averaging of targeted proteomics, multiprotein-marker (MPM) models were developed to differentiate MDD, BD, and SCZ. The performance of ensemble models that combined MPM models and the Symptom Checklist-90-Revised was compared with clinician rater score-based models (Study 2). In both studies, functions and pathways related to differential proteins were predicted with bioinformatics analysis.

Results: Fourteen DEPs were statistically significant between drug-free MDD and BD. RAB7A, ROCK2 were significantly overexpressed in MDD, and EPO7 was significantly overexpressed in BD (Study 1). Each MPM model developed for pairwise patients group comparison (MDD vs BD, MDD vs SCZ, BD vs SCZ) demonstrated reasonable or good differentiation performance in independent test sets (AUROC=0.74~0.82). In addition, the ensemble models performances (AUROC=0.77~0.90) were overall comparable to those of clinician rater score-based models (AUROC=0.74~0.94) in independent test sets (Study 2). Further, the differential proteins in both studies were associated with cellular functions and immune/inflammatory pathways.

Conclusions: In this study, the viability of proteomic quantification and its integration with clinical data in comparing and differentiating major psychiatric disorders is proposed. The results indicate that these approaches have potential in differentiating MDD, BD and SCZ. Further studies with longitudinal designs are warranted.

Keyword: major depressive disorder, bipolar disorder, schizophrenia, proteomics, multiple reaction monitoring, proteomic profiling

Student Number: 2013-21686

*The following articles have been published, and integrated in the dissertation

1. Rhee SJ, Han D, Lee Y, Kim H, Lee J, Lee K, Shin H, Kim H, Lee TY, Kim M, Kim SH, Ahn YM, Kwon JS, Ha K. Comparison of serum protein profiles between major depressive disorder and bipolar disorder. BMC Psychiatry. 2020 Apr 3;20(1):145. doi: 10.1186/s12888-020-02540-0. PMID: 32245436

The article based on comparing drug-free MDD and BD is integrated in Chapter 2. The author contributed as the 1st author.

2. Shin D*, Rhee SJ*, Shin D, Joo EJ, Jung HY, Roh S, Lee SH, Kim H, Bang M, Lee KY, Kim SH, Lee J, Kim Y, Yeo I, Kim Y, Kim J, Kwon JS, Ha K, Ahn YM*, Kim Y. Integrating proteomic and clinical data to discriminate major psychiatric disorders: Applications for major depressive disorder, bipolar disorder, and schizophrenia. Clin Transl Med. 2022; Jun;12(6):e929. doi: 10.1002/ctm2.929. PMID: 35758551

The article based on comparing MDD, BD, and SCZ is integrated in Chapter 3. The author contributed as the co-1st author.

Table of Contents

List of tables

List of figures

List of abbreviations

ACN: acetonitrile

AIC: Akaike's information criterion

ANCOVA: a blend of analysis of variance and regression

ANOVA: analysis of variance

AUROC: area under the receiver operating characteristics

BD: bipolar disorder

BD-I: bipolar I disorder

BD-II: bipolar II disorder

BMI: body mass index

BPRS: Brief Psychiatric Rating Scale

BZD/HNT: benzodiazepines/hypnotics

CE: collision energy

CNS: central nervous system

CRSB: clinician rater score-based

DEP: differentially expressed protein

DSM: Diagnostic and Statistical Manual of Mental Disorders

DTT: dithiothreitol

ES: ensemble

FDR: false discovery rate

HAM-A: Hamilton Anxiety Rating Scale

HAM-D: Hamilton Depression Rating Scale

HC: healthy controls

HCD: higher-energy collisional dissociation

HSD: honestly significant difference

IAA: iodoacetamide

IPA: Ingenuity Pathway Analysis

LASSO: least absolute shrinkage and selection operator

LC: liquid chromatography

MADRS: Montgomery-Asberg Depression Rating Scale

MDD: major depressive disorder

MINI: Mini-International Neuropsychiatric Interview

MPM: multiprotein marker

MRM: multiple reaction monitoring

MS: mass spectrometry

NCE: normalised collision energy

PAR: peak area ratio

PCA: primary component analysis

SCLB: symptom checklist-based

SCL-90-R: Symptom Checklist-90-Revised

SCZ: schizophrenia

SIS: stable isotope-labeled internal standard

SNUH: Seoul National University Hospital

TCEP: tris(2-carboxyethyl)phosphate

TFA: trifluoroacetic acid

YMRS: Young Mania Rating Scale

Chapter 1. Introduction

1.1. Study Background

Major depressive disorder (MDD), bipolar disorder (BD), and schizophrenia (SCZ) are major psychiatric disorders that are known to be associated with life-long disability and morbidity, which poses substantial burdens on patients and economies (1-5). MDD is characterized by an episode of loss of interest and/or depressive mood. The diagnosis needs to exclude a history of (hypo)mania. BD is a chronic disorder with recurrent episodes of depression and (hypo)mania. SCZ is a heterogenous disorder that includes positive (hallucination/delusion) symptoms, negative symptoms, and cognitive impairment as core symptoms. All 3 disorders are common—MDD is known to have a lifetime prevalence of 16.6%, followed by BD of 3.9% and SCZ of 0.7% (6-8). The burden of these psychiatric disorders is increasing, and according to the Global Burden of Disease Study 2019, MDD, SCZ, and BD ranked first, third, and fifth, respectively, for the highest mental disorder burdens, based on the Years Lived with Disability (YLD) (9). All of these disorders ranked in the top 30 leading causes of disability when considering other conditions/diseases, and especially MDD was ranked second overall (9). Moreover, mortality rates including suicide are significantly increased than the general population (10).

As the diagnosis of these disorders is based on subjective symptoms and behavioral observations, and as our knowledge of the biological basis of these

disorders is limited, differentiating them can be challenging. The complexity and heterogeneity of each disorder, and the shared symptoms between them further complicate accurate diagnosis. BD can be misdiagnosed as SCZ during a manic psychotic episode, and as MDD during a depressive episode. Additionally, for MDD or BD, a depressive episode with psychotic features can be misdiagnosed as SCZ. Nearly 40% of BD patients are known to be initially diagnosed as MDD (11-13). Additionally, around 30% of BD patients are known to be diagnosed as SCZ or other psychotic disorders (14). Misdiagnosis leads to delayed or inappropriate treatment, which can be a serious problem (15, 16). Therefore, researchers and clinicians have been interested in finding objective markers that can help to differentiate these disorders. As the diagnosis of these disorders depend on clinical evaluation of self-reported symptoms and behavioral observations, the integration of molecular biomarkers and clinical symptoms could further enable objective differential diagnosis.

Considering the limitations of genomic and transcriptomic studies in previous research of psychiatric disorders, interests have increased in proteomics-based research, as proteomes can reflect biological function (17-20). Especially, recent advances in mass spectrometry (MS)-based proteomic research have improved the development of high-throughput techniques for quantifying multiple proteins simultaneously (21, 22). Thus, MS-based proteomics is suitable for discovering and quantifying multiple proteins that are associated with certain disorders (23). MS-based proteomic profiling

analyzes global proteomes in biological samples and is performed to discover molecular biomarkers, while MS-based targeted proteomics, including multiple reaction monitoring (MRM), detects and quantifies proteins of interest with high accuracy and reproducibility (21-23).

Traditional proteomic studies for psychiatric disorders have focused on the analysis of proteomes in the central nervous system (CNS) including post-mortem brain tissues and cerebrospinal fluid, however accessibility and invasiveness remains a major challenge (17, 18, 24). Thus, quantitative proteomics has sought to analyze peripheral blood samples. Initial efforts focused on comparing psychiatric diseases and healthy controls (HC) (24-28). Recent studies have focused on differentiating MDD with BD (29-33), BD with SCZ (34-38), and SCZ with MDD (39). There was a previous report that compared the proteomes of MDD, BD and SCZ in peripheral blood mononuclear cells, however it was based on few samples (40).

1.2. Purpose of Research

Considering the current background, the study was conducted as follows. In Study 1, proteomic profiling was performed in psychotropic drug-free MDD and BD patients. This was done to compare proteins that were significantly different in peripheral serum. In Study 2, the research protocol was expanded and both targeted quantification and protein profiling were performed in MDD, BD, SCZ patients and HC. Markers based on targeted

data quantification from plasma proteins were developed to differentiate MDD, BD, SCZ. Clinical variables were integrated to compare differentiative performances. Proteomic profiling was performed to compare significancy and consistency of expression patterns between the two types of proteomic data. Finally, biological function analysis was performed in both studies.

Chapter 2. Differentially expressed serum proteins between drug free major depressive disorder and bipolar disorder

2.1. Methods

Clinical samples

The study population initially comprised 18 MDD and 15 BD patients, who were enrolled between May 2012 and September 2017. Ages ranged from 16 to 42 years. The BD patients consisted of 4 BD-I (bipolar I disorder), 10 BD-II (bipolar II disorder), and 1 bipolar disorder not otherwise specified. Patients were recruited from Seoul National University Hospital (SNUH) and Seoul National University Bundang Hospital. Patients were diagnosed clinically with the Diagnostic and Statistical Manual of Mental Disorders, Fourth or Fifth edition (DSM-IV or DSM-5) by psychiatric specialists. Final statistical analysis was based on 25 subjects who were psychotropic drug free for at least 2 weeks, and with no missing values for clinical and demographic data.

Patients were excluded with the following criteria: co-diagnosis with substance-related disorders, history of physical illnesses such as hypertension, hypercholesteremia, liver diseases, epilepsy, and endocrine diseases including diabetes and thyroid diseases, evidence of intellectual disability, organic brain disease, and difficulties interpreting the Korean language.

The study design was approved by the Institutional Review Board of SNUH (IRB No. 1704-075-846). The study was performed in accordance

with the Declaration of Helsinki. Written informed consent was obtained from each participant, and for those who were under 18, from both the participant and their parents/guardians.

Serum samples from each individual subject were obtained in the morning, after overnight fasting (> 8 h). Samples were centrifuged at 3000 rpm for 10 min at 4 °C. The supernatant was collected and stored in Eppendorf tubes at < −70 °C until use.

Demographics and clinical features

The demographics that were considered for the patients in this study were age, sex, body mass index (BMI), and current smoking status. Age and BMI were analysed as continuous variables, and sex (men/women), current smoking status (yes/no) were analysed as dichotomous variables. Symptom severity was assessed with the Hamilton Depression Rating Scale (HAM-D) (41). All MDD patients were in mild/moderate depressive state, while the BD patients were depressed (n = 5), depressive with irritability (n = 4), or hypomanic (n = 1). The chronicity of the disease and medication was assessed as continuous variables; the duration from first onset (years), and the duration from first medication (years).

Serum sample preparation

Serum samples were prepared in line with the previous method reported by Geyer et al. (2016) (42) with some modifications. Briefly,

digestion buffer (8 M urea, 5 mM tris(2-carboxyethyhl)phosphate (TCEP), 20 mM chloroacetaldehyde in 0.1 M avidin-biotin complex) was added to 2 μL of each serum sample. To denature and alkylate the proteins, the mixture was heated at 60 °C for 25 min, and then was cooled to room temperature. The first protein digestion was conducted at 37 °C, overnight with a trypsin/LysC mixture at a 100:1 protein-to-protease ratio. The second digestion was conducted at 37 °C, for 2 h with trypsin (enzyme-to-substrate ratio [w/w] of 1:1000). All of the resulting peptides were acidified with 10% trifluoroacetic acid (TFA) and desalted with homemade C18-StageTips as described in previous studies (43, 44). The desalted samples were dried with a vacuum dryer, and stored at −80 °C.

Construction of a matching library

To establish a spectral library for matching between runs (45), the MARS-14 column (Agilent Technologies, Santa Clara, CA, USA) was used to deplete the 14 blood proteins with highest abundance, according to the manufacturer's instructions. The depleted samples were processed with a two-step, filter-aided, sample preparation, as described in previous studies (43, 44). All of the resulting peptides were desalted with homemade C18-StageTips. For the deep serum data set, 25 μg of purified and digested peptides were fractionated with an Agilent 1260 bioinert HPLC (Agilent Technologies) equipped with an analytical column (4.6 × 250 mm, 5-μm particle). High-pH reversed-phase peptide fractionation was performed at a

flow rate of 0.8 mL/min over a 60-min gradient with two solvents (15 mm ammonium hydroxide in water, 15 mM ammonium hydroxide in 90% acetonitrile (ACN)). The fractions were desiccated in a vacuum centrifuge and stored at −80 °C until LC(liquid chromatography)-MS/MS analysis.

LC-MS/MS analysis

Profiling analysis was performed with a Quadrupole Orbitrap mass spectrometer; the Q-Exactive Plus (Thermo Fisher Scientific, Waltham, MA, USA) coupled with an Ultimate 3000 RSLC system (Dionex, Sunnyvale, CA, USA) equipped with a nano-electrospray ion source, as described in previous studies (43, 46). The peptide samples were separated on a two-column system, with a trap column (75 μm I.D. × 2 cm, C18 3 μm, 100 Å) and an analytical column (50 μm I.D. × 15 cm, C18 1.9 μm, 100 Å). The dried peptide samples were re-dissolved in Solvent A (2% ACN and 0.1% formic acid), prior to sample injection.

The peptides separation on a 90-min gradient from 8% to 30% Solvent B (100% ACN and 0.1% formic acid) was applied to all samples. The spray voltage was set to 2.0 kV in positive ion mode, and the temperature of the heated capillary was set to 320 °C. Mass spectra were collected in data-dependent mode with the top 15 method on the Q Exactive. The Orbitrap analyser scanned precursor ions with a mass range of 300–1650 m/z and resolution of 70,000 at 200 m/z. Higher-energy collisional dissociation (HCD) scans were acquired at a resolution of 17,500, and HCD peptide fragments

were acquired at a normalised collision energy (NCE) of 28. The maximum ion injection time for the full MS/MS scans was 20 and 120 ms, respectively.

Data processing

MaxQuant (version 1.5.3.1) was used to process mass spectra (47), The Andromeda engine was utilized to match MS/MS spectra with the Uniprot human database (December 2014, 88,657 entries) (48). For total protein level analysis, primary searches were done with a 6-ppm precursor ion tolerance. The search parameters were as follows; MS/MS ion tolerance at 20 ppm, *N*-acetylation, and methionine oxidation as variable modifications, cysteine carbamidomethylation as fixed modification, full enzymatic digestion with trypsin, peptides with a minimal length of six amino acids, and up to two missed cleavages. A false discovery rate (FDR) was set to less than 1%, at peptide, protein, and modification levels. Matching between runs was performed with the depleted sample as a library, to maximize quantification events across samples.

Statistical analysis

Among the 33 patients, 25 were psychotropic drug free for at least 2 weeks, with no missing values for demographic and clinical data. These drug free patients were considered for the following analysis. Demographic and clinical differences between MDD and BD were analysed with the Mann-Whitney U-test for continuous variables and the Fisher's exact test for

dichotomous variables. Statistical analyses for the data-dependent acquisition (DDA) data were performed using the Perseus software (49). At first, proteins which were identified as only identified by site, reverse, and contaminants were removed. Then, the expression levels were calculated with the Maxquant software by their Intensity Based Absolute Quantification (iBAQ) values. Log_2 transformation was conducted to decrease the skewed distribution. Proteins with a minimum of 70% quantified values in at least the MDD group or the BDD group were considered valid. Missing values were imputed based on a normal distribution (width = 0.3, down-shift = 1.8). To detect differentially expressed proteins (DEPs), *t*-tests were performed for pairwise proteome comparisons. Protein abundances were subjected to z-normalisation, followed by hierarchical clustering with the Pearson's correlation distance. Linear regression was performed to control covariates that differed between patient groups.

Canonical pathways and diseases/functions associated with the DEPs were evaluated by Ingenuity Pathway Analysis (IPA, QIAGEN, Hilden, Germany) (50) based on corresponding gene names. Additionally, the top protein network and associated diseases/functions were predicted. The analytical algorithms in IPA use lists of DEPs to predict biological processes and pathways with the Fisher's exact test.

All statistical tests were two-sided and $P < 0.05$ was considered statistically significant.

2.2. Results

Demographic and clinical characteristics

In total, 15 MDD patients and 10 BD patients who were psychotropic drug free for at least 2 weeks were included in the final analysis. There were no significant differences between MDD and BD regarding age, sex, BMI, smoking, duration from first onset, and duration from first medication. For depressive symptoms, the MDD group was more severe than the BD group (HAM-D score; 15.33 ± 4.61 vs 12.10 ± 3.54, $P = 0.02$). The summary of demographic and clinical characteristics is presented in Table 1-1.

Quantitative analysis

Comprehensive serum profiling with the depleted, pooled serum samples generated a peptide matching library, which consisted of 1616 proteins/16,505 peptides. Thirty-three serum samples were analysed by unbiased single-shot approaches, and "match between runs" functionality with the constructed peptide library. This leaded to 481 quantified serum proteins with at least two peptides. The 268 proteins that were quantified at least 70% in either the drug-free MDD or BD group were subjected to statistical analysis. *T*-tests revealed 14 DEPs between drug-free MDD and BD (Figure 1-1). When adjusting for multiple comparisons (with the Benjamini-Hochberg FDR adjusted p-value), the levels of RAB7A, ROCK2, and XPO7 were still significantly different (Table 1-2). Hierarchical clustering revealed

that the two disorders generally clustered together (Figure 1-2). The association between expression levels of the 3 DEPs and diagnosis (MDD or BD) was still significant, after controlling the total HAMD score as a covariate in linear regression.

Bioinformatics analysis

The 14 DEPs were subjected to bioinformatics analysis. The tree-map of the diseases/functions is presented in Figure 1-3. The DEPs were especially enriched in organismal injury and abnormalities, cell-to cell signalling and interaction, cellular function and maintenance, inflammatory response, and cellular assembly and organization. The top five canonical pathways were LXR/RXR activation, IL-12 signalling and production in macrophages, clathrin-mediated endocytosis signalling, actin cytoskeleton signalling, and the extrinsic prothrombin activation pathway (Figure 1-4).

The top network of the 14 DEPs consisted of six proteins (RAB7A, ROCK2, CD14, ANG, SELENOP, and KIF20B). The related diseases/functions with the network incorporated cellular movement, haematological system development and function, and immune cell trafficking (Figure 1-5).

2.3. Discussion

In this preliminary study, 14 DEPs between psychotropic drug-free MDD and BD patients were detected. RAB7A and ROCK2 were overexpressed in MDD, and XPO7 was overexpressed in BD. Bioinformatics analysis revealed that cellular functions and inflammation/immune response pathways were significantly different between MDD and BD.

Reviewing previous studies, Chen et al. (2015) analysed plasma samples of drug-naïve patients and identified 25 DEPs, whereas Ren et al. (2017) analysed plasma samples of drug-free patients and identified 9 DEPs (29, 30). Recently, Idemoto et al. (2021) analysed serum samples of drug-free patients and identified 44 DEPs (32). The present study found only one protein, ROCK2, as a duplicate protein (29) from these previous studies. The discrepancies are probably related to the different study designs, demographic and clinical characteristics of the population, and mass quantification platforms. The present study was based on serum samples, which can differ with plasma protein quantification (51). Moreover, the depressive symptoms were milder in this study, which might have led to different peripheral profiles.

The most significant DEP based on fold-change and statistical significance was RAB7A. This protein is known to be a key regulator of endo-lysosomal trafficking (52), and is found both in the CNS and peripheral blood (53, 54). Investigation of its expression in Alzheimer disease has revealed that it was increased in the cerebrospinal fluid (54), and its gene expression was

significantly changed in both the hippocampus and peripheral blood (53). A previous study revealed that its gene expression was upregulated in the post-mortem brain of those with depression who committed suicide (55). Other functions of RAB7A, including endoplasmic reticulum stress modulation, growth-factor-mediated cell signalling, and lipid metabolism (56, 57), are all proposed pathophysiological mechanisms of mood disorders (58-61).

ROCK2 was also overexpressed in MDD, and the direction of expression was consistent with the previous results of Chen et al. (29). ROCK is a serine/threonine kinase, and is known to be a crucial regulator of actin cytoskeleton and cell polarity (62). Specifically, the inhibition of ROCK increases neurite outgrowth and axonal regeneration, and activates pro-survival protein kinase B in the CNS (62). The isoform ROCK2, is especially distributed in the brain, spinal cord, and heart (63). For mood disorders, there is a report that placental ROCK2 is downregulated in depressed women (64). Additionally, there is evidence that ROCK2 is upregulated during sleep deprivation (65) and is involved in the circadian variation of vascular contractility (66). Interestingly, circadian rhythm dysregulation is also a proposed marker specific for BD, when compared to MDD (67).

XPO7 is known to mediate the nuclear export of proteins (68). A genome-wide association study of alcohol dependence identified XPO7 as a significant gene (69). However, in a confirmation study, it was significantly changed only in patients with BD and comorbid alcohol dependence, but not in those without alcohol dependence (70). Nevertheless, the specific

pathophysiology is currently unclear, because of the sparse literature between XPO7 and psychiatric manifestations.

Bioinformatics analysis of the DEPs and its network revealed that cellular functions and inflammation/immune pathways were significantly altered between MDD and BD. The most significant canonical pathway was LXR/RXR (liver X receptor/retinoic acid receptor) activation. LXRs is known to form heterodimers with RXRs, which regulates the expression of genes that control lipid homeostasis (71). Interestingly, this pathway was reported to be significantly associated with plasma DEPs in an animal model of depression (72). Moreover, specific pathways of inflammation and immune dysfunction were significantly altered between MDD and BD, which are known to be involved in the pathophysiology of mood disorders (73-75).

There were several limitations of this study. First, as the study focused on drug-free patients, the sample size was small. Larger sample sizes would increase statistical power to detect DEPs, which would increase the precision of bioinformatics analysis. Second, the study lacked a control group without mental disorders. This made it difficult to interpret the expression direction of DEPs between MDD and BD. Third, as the study was cross-sectional, causal relationship between DEPs and mood disorders could not be determined. Additionally, the late manifestation or non-report of hypomanic/manic episodes might mislead the diagnosis of BD to MDD. Fourth, the study lacked validation. Validation in an independent set, or with other platforms like enzyme-linked immunosorbent assay (ELISA) would

confirm the reliability of the results. However, the insufficiency of samples limited validation in other populations and with alternative assays. Fifth, depressive symptom severity could have influenced the DEPs significance of the study. However, the association between expression levels of the three DEPs and diagnosis (MDD or BD) was still significant, after controlling the total HAMD score as a covariate in linear regression. Sixth, treatment histories could have influenced the results of the study. Even though only those whom were drug-free were analysed, prior treatment medication regimens could have influenced the results. Finally, other uncontrolled covariates such as exercise, and alcohol drinking might have influenced the protein profiles. However, despite these limitations, the study's strength is that it focused on potential DEPs in the difference between MDD and BD, which is an understudied subject. Additionally, only drug-free patients were included, and blood collection time/fasting time was controlled.

In conclusion, the preliminary study demonstrated that proteomic profiles in the serum differed between MDD and BD. RAB7A, ROCK2, and XPO7 proteins were differentially expressed between these disorders after controlling for multiple comparisons. These DEPs might enable differentiation, and expand the understanding of the pathophysiology of these disorders. Additional studies involving a larger sample size, with a control group, with more information on covariates that can influence the proteomic profiles, as well as including validation designs, are warranted. Furthermore, longitudinal designs to determine protein profiles from those whom are

initially diagnosed with MDD but develop hypomanic/manic symptoms later, are needed.

Chapter 3. Differentiation of major psychiatric disorders with plasma proteome and clinical data [Study 2]

3.1. Methods

Clinical samples

The study population comprised 174 MDD, 170 BD, 171 SCZ patients and 160 HC, who were enrolled between August 2018 and December 2020. Ages ranged from 19 to 65 years. The BD patients consisted of 75 BD-I, 84 BD-II, and 11 other specified bipolar and related disorder. Patients were recruited from SNUH, Nowon Eulji Medical Center, Eulji University; Seoul Metropolitan Government Seoul National University Boramae Medical Center, Hanyang University Seoul Hospital, Inha University Hospital, and Cha University Bundang Medical Center. Patients were diagnosed with the DSM-5, confirmed by the Mini-International Neuropsychiatric Interview (MINI). HCs were recruited via advertisement at SNUH. HC had to have no psychiatric diagnosis confirmed by the MINI, and no known familial psychiatric history within second-degree relatives. Patients had to have a Clinical Global Impression - Severity ≥ 3 to participate.

Patients and HC were excluded with the following criteria: use of anti-inflammatory analgesics including nonsteroidal anti-inflammatory drugs (NSAIDs) and steroids in the past two weeks (acetaminophen was permitted); history of neuromodulation including deep brain stimulation,

electroconvulsive therapy, transcranial direct current stimulation, and transcranial magnetic stimulation; history of neurosurgery; CNS diseases including epilepsy, meningitis, parkinsonism and stroke; cancer; tuberculosis; lactation/pregnancy; history of substance use disorder excluding alcohol, caffeine, and nicotine; intensive psychotherapy in the past two months; evidence of intellectual disability; and difficulty interpreting the Korean language. Exclusion criteria was based on previous evidence of the association between certain conditions/diseases and protein expression (76-84). Those who were on psychotherapy were excluded to confine treatment effects to psychotropic medication.

The study design was approved by the Institutional Review Board of SNUH (IRB no. 1806-106-951) and all the other hospitals that participated. The study was performed in accordance with the Declaration of Helsinki. Written informed consent was obtained from each participant.

Plasma samples from each individual subject were obtained in a 6-mL ethylenediaminetetraacetic acid (EDTA) tube (Ref 367863, Becton, Dickinson and Company, Trenton, NJ). Samples were centrifuged at 1100−1300g for 10−15 min at 4°C or room temperature. The supernatant was collected and stored in Eppendorf tubes at < −70 °C until use.

Demographics and clinical features

The collected demographics that were considered for both patient groups and HC were age, sex, BMI, blood collection time, fasting time,

current exercise, alcohol and smoking status. Age and BMI were analyzed as continuous variables. Sex (men/women), blood collection time (AM, PM), fasting time (< 8 hours, ≥ 8 hours), current exercise status (yes/no), current alcohol use (yes/no), and current smoking status (yes/no), were analyzed as dichotomous variables. Current exercise status was defined as at least 30 min, once a week of moderate-intensity physical activity by the World Health Organization (WHO) recommendation (85). Current alcohol use was defined as at least 1 drink, once a week.

Symptom severity was assessed with the Brief Psychiatric Rating Scale (BPRS) (86), Young Mania Rating Scale (YMRS) (87), Montgomery-Asberg Depression Rating Scale (MADRS) (88), and Hamilton Anxiety Scale (HAM-A) (89). The subjective symptoms were checked with the Symptom Checklist-90-Revised (SCL-90-R) (90). As bipolar disorder has different mood states, patients with total YMRS score > 12 points were classified as having current hypomanic/manic/mixed symptoms (91).

Medication use was analyzed as a dichotomous variable for antipsychotics (APs), mood stabilizers (MSs), antidepressants (ADs), and benzodiazepines/hypnotics (BZD/HNT). The chronicity of the disease or medication was assessed as continuous variables; the duration from first onset (years) and duration from first medication (years).

Plasma sample preparation

For the targeted proteomic analysis, there were 5 preparation batches. In each batch, the samples were randomly distributed and assigned identification numbers to blind the researchers. The 6 highest abundance proteins from 44 μL of each plasma sample, was depleted with the MARS-6 column (Agilent Technologies, Santa Clara, CA, USA). The depleted plasma was concentrated with a 3000-Da molecular weight cutoff (MWCO) filter (Amicon Ultra-4 3K, Millipore, Burlington, MA, USA), and quantified by the Pierce™ BCA Protein Assay Kit (Thermo Scientific, Rockford, IL, USA). 100 μg of protein was reduced with 40 ul solution of 0.2% RapiGest, 20 mM dithiothreitol (DTT) at 60°C for 1h, and alkylated with 20 μL 100 mM iodoactamide (IAA) in the dark at room temperature for 30 min. Then the samples were digested with trypsin solution at 37°C for 4 h. Digestion was completed by adding 10% formic acid. The sample was centrifuged at 15,000 rpm for 1h at 4°C to remove insoluble chemicals. The supernatants were spiked with crude stable isotope-labeled internal standard (SIS) peptides, in which a C-terminal lysine or arginine was heavy-isotope-labeled ($^{13}C_6{}^{15}N_2$ or $^{13}C_6{}^{15}N_4$) [purity: crude (>70%), JPT, Berlin, Germany].

For proteomic profiling analysis, the remaining depleted individual plasma samples after targeted proteomic preparation were pooled for each group. In addition, equal amounts of pooled plasma samples for each group were integrated to generate a peptide spectral library. A total of 5 pooled samples were centrifuged at 15,000 rpm for 20 min at 4°C. The supernatant of 100 μl was dissolved in 300 μL lysis buffer (4% SDS; 0.1M TEAB, pH 8.5;

2 mM TCEP) and incubated at 100°C for 30 min. Protein concentrations were measured using a BCA reducing agent-compatible kit (Thermo Fisher Scientific, Waltham, MA, USA). A total of 300 μg of extracted protein was precipitated with cold acetone, denatured with 35 μL 100 mM DTT at 100°C for 35 min, and alkylated with 50 mM IAA in the dark at room temperature for 1 h. The proteins were digested at 37°C for 18 hours with trypsin (enzyme:substrate ratio [w/w] of 1:50) and 4% ACN. The digested peptides were measured and equalized. Then the equalized peptides were acidified with 10 μL 10% TFA, and desalted with homemade C18-StageTips (92). The desalted samples were dried with a vacuum dryer, and stored at −80 °C. To increase the number of identified proteins, high-pH reversed-phase peptide fractionation was performed for each pooled plasma sample (43, 93). The fractions were desiccated in a vacuum centrifuge and stored at −80 °C until LC-MS/MS analysis.

Determination of detectable and quantifiable targeted proteins

For targeted proteomic analysis, three sources were integrated: 1) new targets for major psychiatric disorders (MDD, BD, and SCZ), 2) established targets for mood disorders (MDD and BD), and 3) laboratory-established targets.

New targets for major psychiatric disorders were collected from 5 databases: PsyGeNET (http://www.psygenet.org), Schizophrenia Gene Resource 2 (https://bioinfo.uth.edu/SZGR/), Laboratory of Neurophenomics

(http://www.neurophenomics.info/), Comprehensive Database for Schizophrenia (http://www.SPRdb.org), and The Stanley Neuropathology Consortium Integrative Database (http://sncid.stanleyresearch.org) (94-97). As a result, 8081 genes were established as initial target proteins. After including targets that are known to be detectable in the blood with the Human Blood Proteins Atlas and Plasma Proteome Database (98, 99), 1462 proteins were selected. To further filter targets with matching MS/MS spectra and unique peptides, 8 MS/MS spectral libraries from the Institute for Systems Biology (https://www.systemsbiology.org), the National Institute of Standards and Technology (https://www.nist.gov), and the SWATHAtlas database (www. SWATHAtlas.org) were utilized. The peptide length was confined from 6 to 20 amino acids. Eventually, 407 proteins were selected; the highest intensity unique peptide for each protein, and the top 10 transitions for each peptide.

The established candidate targets of mood disorders (MDD and BD) were drawn from a previous study (100), which included a step of evaluating detectability and quantifiability of the DEPs of Study 1 (101). The laboratory-established candidate targets included proteins approved by the US Food and Drug Administration, and laboratory developed tests and proteins which were developed in previous research unrelated to psychiatric disorders. In total, 1667 proteins/2283 peptides were merged as candidate targets.

To examine targets that were detectable and quantifiable in plasma samples of psychiatric disorders, LC-MRM-MS analysis was performed on a

pooled plasma sample that consisted of 50 participants in each group including HC. Targets were considered detectable if: 1) at least five transitions were observed for LC-MRM-MS, 2) they had the same elution patterns within the predicted retention time of ± 5 min, 3) library dot product > 0.6, 4) RTs and dot products were equal between light and heavy peptides. After including the top highest intensity transition per peptide based on the rank of intensity that was filtered by AuDIT (automated detection of inaccurate and imprecise transitions) (102), the analysis resulted in a total of 642 target peptides selected as being quantifiable.

LC-MS analyses for targeted proteomics and proteomic profiling

Targeted proteomic analysis (LC-MRM-MS) was performed with a 1260 Infinity HPLC system equipped with a Jetstream electrospray source, coupled to an Agilent 6490 triple quadrupole MS (Agilent Technologies, Santa Clara, CA, USA). The sample vials of the autosampler were maintained at 4°C, and the guard and analytical column was maintained at 40°C. For each digested sample, 40 µl was injected into a guard column (2.1 × 15.0 mm, 1.8 µm, 80 Å) (Agilent Technologies, Santa Clara, CA, USA), and online desalting was conducted in 3% solvent B (0.1% formic acid/ACN (v/v)) at 50 µl/min for 10 min. After the position of valve was switched, the sample was transferred to the analytical column (0.5 × 35.0 mm, 3.5 µm, 80 Å) (Agilent Technologies, Santa Clara, CA, USA) in 3% solvent B, at 40 µL/min

for 5 min. Bound peptides were separated on the column and eluted with a linear gradient of 3% to 35% solvent B at 40 μL/min for 50 min.

The mass spectra were generated in positive ion mode, based on the following parameters: 2500 V ion spray capillary voltage, 2000 V nozzle voltage, 5 V cell accelerator voltage, 200 V delta EMV, and 380 V fragmented voltage. The drying gas was sprayed at 15 L/min at 250°C, and the sheath gas flow was 12 L/min at 350°C. Collision energy (CE) was optimized by adding the intensities of individual transitions that resulted in the largest peak area. Before individual sample analysis, SIS peptides were pooled and analyzed to evaluate their RTs. The RTs were compared with those of endogenous target peptides by spiking the pooled mixture of SIS peptides with 100 fmol of heavy β-galactosidase peptide. Subsequently, the final targets were quantified in individual blood samples. LC-MRM-MS analysis was performed once per sample (1 replicate for each sample).

Proteomic profiling analysis was performed with an Easy-nLC 1000 system equipped with a nano-electrospray ion source, coupled to a Q-Exactive MS (Thermo Fisher Scientific, Waltham, MA, USA), as described in a previous study (93). The peptide samples were separated on a two-column system, with a trap column (75 μm I.D. x 2 cm, 3-μm Acclaim PepMap100 C18 beads) and an analytical column (75 μm I.D. x 50 cm, 3-μm ReproSil-Pur C18-AQ beads). Lyophilized peptide samples were re-dissolved in Solvent A (2% ACN and 0.1% formic acid) prior to sample injection.

The peptides separation on a 180-min gradient from 6% to 40% Solvent B (100% ACN and 0.1% formic acid) was applied for all samples. The spray voltage was set to 2.2 kV in positive ion mode, and the temperature of the heated capillary was set to 275°C. Mass spectra were collected in data-dependent mode with the top 15 method on the Q Exactive. The Orbitrap analyser scanned precursor ions with a mass range of 350–1700 m/*z,* and resolution of 70,000 at 200 m/z. HCD scans were acquired at a resolution of 17,500, and HCD peptide fragments were acquired at a NCE of 27, 30, and 33. The maximum ion injection time for the full MS/MS scans was 20 ms and 80 ms, respectively. All samples were analyzed in 3 technical replicates.

Processing of targeted proteomic data

The raw data from the LC-MRM-MS analysis was processed in Skyline (version 19.1.0) (MacCoss Lab, Seattle, WA, USA). After manual inspection of peptide-transition peaks, the peak area ratio (PAR); the ratio of the endogenous (Light) to SIS (Heavy) peptide peak area; was calculated. A total of 54 unstable target peptides with low intensities (intensity < 1000), unequal RTs between light and heavy peptide, and skewed peaks were further excluded. Subsequently, the final PAR values of 588 target peptides across 675 samples were normalized by the area of heavy β-galactosidase peptide, to reduce technical variability from sample preparations. The normalized PAR values were $\log_2$ transformed. Potential batch effects between five

sample preparation batches were corrected with the Combat algorithm using the R package proBatch (103).

Sequence-based search for plasma proteome

Proteome Discoverer (version 2.2) was used to process mass spectra. The SEQUEST-HT algorithm was utilized to match MS/MS spectra with a modified version of the Uniprot human database (December 2014, 88,717 protein entries). The search parameters were as follows: a precursor ion mass tolerance value at 20 ppm, a fragment ion mass tolerance value at 0.02 Da, dynamic modification values at 42.01 Da for *N*- acetylation, 15.99 Da for methionine oxidation, and 57.02 Da for cysteine carbamidomethylation, full enzymatic digestion with trypsin, and peptides up to two missed cleavages. FDR was set to less than 1%, at the peptide and protein levels. Peptides were mapped and linked with the "Feature Mapper" node to increase the number of identified proteins. Proteins were quantified and normalized by calculating the relative intensities for peptide-spectrum matches and using the "Precursor Ions Quantifier" node.

Analysis of demographics and clinical characteristics

The study participants with MDD, BD, and SCZ were distributed into training, validation, and independent test sets (6:2:2). The HC were analyzed as a reference set. Demographic and clinical characteristics were compared between patient groups and HC by chi-squared tests for dichotomous

variables, and by one-way analysis of variance (ANOVA) with Tukey's HSD (honest significant difference) test for continuous variables.

Determination of targeted proteomic candidate features in the training set

Initially, proteomic features within the range of raw PAR ≤ 0.05 or raw PAR ≥ 100 in over 5% of individuals, were excluded in each training set.

Next, proteomic features that had significant relations with demographics, medication use, and chronicity of disease and medication were excluded as follows. At first, univariate analysis with the *t*-test for dichotomous variables and Pearson's correlation analysis for continuous variables were performed between each covariate and proteomic feature. For each significant covariate, the disease type was further controlled by univariate analysis of covariance (ANCOVA). All proteomic features that had a statistical significance with a covariate and not with the disease type were excluded.

Finally, proteomic features with a variance inflation factor (VIF) > 5 relative to other features were excluded. The final proteomic candidate features were used to develop MPM (multiprotein marker) models. AUROC (area under the receiver operating characteristics) analysis was performed for each proteomic candidate feature, and fold-changes were calculated by

subtracting the average batch-corrected PAR value between patient groups/HC.

Development of MPM models

Repeated 5-fold cross-validation (100 repetitions) LASSO (least absolute shrinkage and selection operator) regression was performed in the training sets for each pairwise group comparison, with the R package glmnet (104). Feature extraction and model averaging based on previous studies (100, 105) were performed to develop MPM models as follows. For each model that was generated with the repetition process, the bias-corrected version of the AIC (Akaike's information criteria) was calculated as :

$$\mathrm{AIC_c} = AIC + \frac{2K(K+1)}{n-K-1}$$

where n is the sample size, and K is the number of model parameters.

The Akaike weight (w_m) was calculated with the bias-corrected version of AIC (AIC_c), to represent the probability of each model being the best model (100, 105), as :

$$w_m = \frac{e^{\left(-\frac{1}{2}AIC_{c(m)}\right)}}{\sum_{j=1}^{M} e^{\left(-\frac{1}{2}AIC_{c(j)}\right)}}$$

where $\mathrm{AIC_{c(j)}}$ is the AIC_c for model j = 1 to M.

The frequency proportion across the 100 models for each proteomic candidate feature was defined as the selection fraction. Features that had a selection fraction = 1 or ≥ 0.8 were combined, in each pairwise group comparison. The

weighted coefficient for each feature was defined by summing the product of the Akaike weight and coefficient of each feature, of each unique model, as :

$$\beta = \sum_{j=1}^{M} w_j \beta_j$$

where w_j and β_j are the Akaike weight and coefficient estimates for a feature of interest in model j, and β is the weighted average of β_j across model j = 1 to M.

The above process generated two MPM models based on the selection fraction, for each pairwise group comparison. The performance of each MPM model was evaluated in the validation sets. Then, considering the number of combined features and differentiation performance in the validation sets, the final MPM models were selected. Pearson's correlation analysis was performed to analyze the association between clinical symptoms with proteomic candidate features from the final MPM models. For significant correlations, ANCOVA was additionally performed to determine if the features were associated with current symptoms or disease types. Finally, the performances of the final MPM models were tested in the independent test sets.

Expression levels of the proteomic features in the final MPM models

Expression levels of the targeted proteomic features in the final MPM models were examined in the total study populations. ANOVA was

performed for each pairwise group comparison—1) MDD versus BD versus HC, 2) MDD versus SCZ versus HC, and 3) BD versus SCZ versus HC—, followed by Tukey's HSD for post-hoc analysis. Violin plots for the final MPM model value based on the proteomic features, were constructed to compare mean/variances between groups in the total study population.

Development of SCLB, CRSB, and ES models in the training sets

Symptom checklist-based (SCLB) models were determined by considering all combinations of the dimensions and the overeating item of the SCL-90-R. The models with the highest differentiation performance by binary logistic regression in the training sets, were selected as the final SCLB models. Further, all total scores of the clinician rater scales (BPRS, MADRS, YMRS, and HAM-A) were used to develop clinician rater score-based (CRSB) models by binary logistic regression in the training sets. Both were performed with the “glm” function of the R package e1071 (106). Finally, ensemble (ES) models were constructed by combining MPM and SCLB models, based on the stacking ensemble strategy (107). This was performed by combining prediction values (represented as the probability) of the MPM and SCLB models in the training sets.

Differentiation and diagnostic performance of the models

The differentiation performance of each model was calculated with AUROC values by the R package pROC (108) in the training, validation,

independent test, and total sets. The optimal cutoff of ES models were determined by the Youden Index as follows: J = max (Sensitivity + Specificity – 1) (109). Sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) were calculated at the optimal cutoff.

Processing of proteomic profiling data

Proteins had to have all 3 normalized values positive within a group, and at least one unique peptide determined to be quantified. After $\log_2$ transformation, the technical variation was measured by Pearson's correlation, and principal component analysis (PCA) was performed to examine the segregation of groups. Then, ANOVA with Tukey's HSD was performed to identify DEPs between patient groups and HCs. After z-score normalization of DEP levels, k-means hierarchical clustering analysis was performed. Fold-changes were calculated by subtracting the average normalized protein amount between patient groups/HC.

Bioinformatics analysis

The protein networks with a network score ≥ 20 were merged by IPA, based on all of the proteins that were included in the final MPM models for each pairwise group comparison, with their corresponding gene names. Diseases/functions and canonical pathways associated with the merged network were evaluated. The analytical algorithms in IPA use lists of proteins

to predict protein networks and their corresponding diseases/functions and canonical pathways, with the Fisher's exact test.

Statistical analysis

The statistical analyses were performed with SPSS (version 25.0, IBM, Armonk, NY, USA) and R (version 4.1.0). Additionally, Perseus (version 1.5.8.5) was used for proteomic profiling data. Statistical tests were two-tailed, and $P < 0.05$ was considered to be statistically significant.

3.2. Results

Demographics and clinical characteristics

In total, 174 MDD, 170 BD, 171 SCZ patients, and 160 HC were included in the final analysis. There were significant differences between patients and HC groups in all demographic and clinical characteristics except sex (Table 2-1). Further, these characteristics were compared between patient groups in the training, validation, and independent test sets (Tables 2-2, 2-3, and 2-4). Particularly in the training set, demographics except for sex, fasting time, exercise, and smoking, were significantly different between patient groups. Medication use was different excluding BZD/HNT. All of the total clinician rater scores, and dimensions of the SCL-90-R except the paranoid ideation and overeating item were different between patient groups (Table 2-2).

Batch effect correction of the targeted proteomic data

As there were multiple preparation batches, its effect was corrected for the targeted proteomic data to reduce technical variability. PCA plots were examined based on the adjusted PAR after batch effect correction. There were no outliers identified, and sample preparation batch effects were minimized (Figure 2-1a). Additionally, there was no significant batch effect regarding hospitals (Figure 2-1b).

Proteomic candidate features selection in the training set

After excluding proteomic features that had a significant relationship only with the covariates and not the disease type based on univariate analysis and ANCOVA, and that had a variance inflation factor (VIF) > 5 relative to other features, 23, 29, and 30 candidates were identified for the differentiation between MDD with BD, MDD with SCZ, and BD with SCZ, respectively (Table 2-5).

These features overall showed consistent expression level patterns across disease types. In addition, correlations between expression levels with demographic and clinical characteristics were low; the absolute median Pearson's correlations (r) were 0.05, 0.07, and 0.07 for each pairwise group comparison, respectively. AUROC values of each candidate proteomic feature ranged from 0.5 to 0.7 (Figure 2-2). At last, there was low interdependence between proteomic candidate features; the absolute median Pearson's correlations (r) were 0.08, 0.09, and 0.08 for each pairwise group comparison, respectively (Figure 2-3).

Development of MPM models

For each pairwise group comparison, 100 models were generated with LASSO in the training sets. The model probability (Akaike weight) was 0.98 for one unique model, absolutely supporting it for the MPM model of MDD versus BD. In contrast, the highest model probability for unique models were 0.32 and 0.31 for the MPM model of MDD versus SCZ, and BD versus

SCZ (Table 2-6). For the selection fraction of 1 and ≥ 0.8, model averaging was performed to develop MPM models in the training sets. When compared with selection fraction = 1, selection fraction ≥ 0.8 resulted in additional 6, 4, and 8 proteomic features in the MPM models for each pairwise group comparison. The AUROC values of the models in differentiating MDD versus BD, MDD versus SCZ, and BD versus SCZ, based on selection fraction = 1, were 0.84, 0.87, and 0.88 in the training sets and 0.73, 0.74, and 0.72 in the validation sets, respectively. For selection fraction ≥ 0.8, the AUROC values were 0.86, 0.88, and 0.90 in the training sets and 0.75, 0.77, and 0.79 in the validation sets, respectively (Figure 2-4 and Figure 2-5). Even though there was a slight increase in AUROC values for the MPM models based on selection fraction ≥0.8 in the validation sets, the MPM models based on selection fraction = 1 were determined as the final models, to generate a simpler combination with the most important proteomic features (Figure 2-4).

For each final MPM model, the direction of the average coefficient of each proteomic feature was consistent with the direction of alteration of expression in the training set (Figure 2-4). Additionally, ANCOVA was performed in the training set to analyze the relationship between each proteomic feature with clinical symptoms. The following proteomic features were associated with the following clinical symptoms, and not with differential diagnosis: ALDOC was associated with the paranoid ideation and

hostility dimension of the SCL-90-R, ARMD4 was associated with the obsessive-compulsive dimension of the SCL-90-R, and CTND1 was associated with the total BPRS score in the MPM model for MDD versus BD. ALDOC was associated with the psychoticism, paranoid ideation, and hostility dimension of the SCL-90-R, and IBP3 was associated with the obsessive-compulsive dimension of the SCL-90-R in the MPM model for MDD versus SCZ. Finally, GPR37 and UROM were associated with the total BPRS score in the MPM model for BD versus SCZ (Table 2-7)

The final MPM models consisted of 17, 20, and 17 proteomic features for each pairwise group comparison. The AUROC values in differentiating MDD versus BD, MDD versus SCZ, and BD versus SCZ in the independent test sets were 0.74, 0.82, and 0.78, respectively (Figure 2-4). The models of MDD versus BD and BD versus SCZ, were additionally used to differentiate MDD/SCZ with subgroups of BD in the total set. The MPM model for MDD versus BD had an AUROC value of 0.78 in differentiating MDD from BD-II/other specified bipolar and related disorder, and 0.80 in differentiating MDD from BD without current hypomanic/manic/mixed symptoms. The MPM model for BD versus SCZ had an AUROC value of 0.82 in differentiating BD-I from SCZ (Figure 2-6). The violin plots for the final MPM model values in the total population are presented in Figure 2-7.

Proteins of MPM models in the total study population

The mass spectral information of the proteomic features from the final MPM models is presented in Table 2-8. Expression levels of the proteomic features in the final MPM models were examined in the total study population including HC (Table 2-9).

In the final MPM model of MDD versus BD, 10 of the 17 proteins differed significantly between MDD versus BD versus HC. In the post-hoc analysis for MDD versus BD, ALDOC, CETP, DDR1, and ITIH2 were upregulated in MDD, and ARMD4, C1RL, CTND1, and IC1 were upregulated in BD. Additionally for MDD versus HC, ALDOC and AMPN were upregulated in MDD, and ARMD4 and DOPO were upregulated in HC. Finally, for BD versus HC, C1RL and CTN1 were upregulated in BD, and CETP and ITIH2 were upregulated in HC.

In the final MPM model of MDD versus SCZ, 16 of 20 proteins differed significantly between MDD versus SCZ versus HC. In the post-hoc analysis for MDD versus SCZ, ALDOC, AT1A1, CBG, CRYM, GPX3, IBP3, IBP5, ITIH2, PROC, and TFPI1 were upregulated in MDD, and CATS, CBPB2, PROS, and SAA1 were upregulated in SCZ. Additionally for MDD versus HC, ITIH2 was upregulated in MDD. Finally, for SCZ versus HC, ALDOC, CATS, CBPB2, PROS, and SAA1 were upregulated in SCZ, and ALS, AT1A1, CBG, CRYM, IBP5, and PROC were upregulated in HC.

In the final MPM model of BD versus SCZ, 15 of 17 proteins differed significantly between BD versus SCZ versus HC. In the post-hoc analysis for BD versus SCZ, AACT, AMPN, CFAB, HEP2, and PSMD1 were

upregulated in SCZ, and C1RL, CLD3, DOPO, GPR37, IBP5, ICI, MBL2, TFPI1, and UROM were upregulated in BD. Additionally for BD versus HC, C1RL was upregulated in BD, and PSMD1 was upregulated in HC. Finally, for SCZ versus HC, AMPN, BPIB1 and CFAB were upregulated in SCZ, and CLD3, DOPO, IBP5, ICI, TFPI1, and UROM were upregulated in HC.

Overlapping proteomic features for the final MPM models and their expression patterns were further examined (Table 2-9 and Figure 2-8). ALDOC, COAA1, ITIH2, and SAA1 overlapped between the final MPM models for MDD versus BD and MDD versus SCZ. The expression levels of ALDOC and ITIH2 were significantly different between diseases in both MPM models. AMPN, C1RL, DOPO, IC1, and NPC2 overlapped between the final MPM models for MDD versus BD and BD versus SCZ. The expression levels of C1RL and IC1 significantly different between diseases in both MPM models. IBP5 and TFPI1 overlapped between the final MPM models for MDD versus SCZ and BD versus SCZ, and both expression levels were significantly different between disorders in both MPM models. However, there was no protein that overlapped in all of the final MPM models.

Differentiation performances of SCLB, ES, and CRSB models

The SCLB models that were constructed with the combinations of the SCL-90-R dimensions are presented in Table 2-10 and its differentiation performance is plotted in Figure 2-9. When compared with MPM models, the differentiation performances of the SCLB models were generally lower for

pairwise group comparisons of MDD versus BD, and BD versus SCZ, and generally higher for pairwise group comparison of MDD versus SCZ (Figure 2-4 and Figure 2-9). After integrating the final MPM and SCLB models, the ES models AUROC ranges for differentiating MDD versus BD, MDD versus SCZ, and BD versus SCZ were 0.77~0.84, 0.83~0.91, and 0.73~0.89, respectively (Figure 2-10). The CRSB models AUROC ranges for differentiating MDD versus BD, MDD versus SCZ, and BD versus SCZ were 0.74~0.83, 0.91~0.95, and 0.78~0.82, respectively (Figure 2-11). The differentiation performances in each set and diagnostic performances in the independent test set of the ES and CRSB models were overall comparable (Figure 2-12).

Network analysis of proteomic features in the final MPM models

All of the 43 proteins from the final MPM models were subjected to network analysis. The two networks that satisfied network score ≥ 20 were merged, comprising 70 molecules, with 31 proteins from the final MPM models (Table 2-11 and Figure 2-13). Diseases/functions that were associated with the merged network included cellular movement (P = 7.87E-21 to 1.61E-7), cell-to-cell signaling and interaction (P = 9.14E-10 to 1.61E-7), immune cell trafficking (P = 2.3E-12 to 1.3E-7), neurological disease (P = 7.47E-12 to 8.17E-8), and psychological disorder (P = 6.09E-12 to 3.89E-2). Furthermore, the merged network was associated with the following

canonical pathways— production of nitric oxide and reactive oxygen species in macrophages, neuroinflammation signaling pathway, reelin signaling in neurons, synaptogenesis signaling pathway, CREB signaling in neurons, opioid signaling pathway, axonal guidance signaling, FXR/RXR activation, LXR/RXR activation, and acute phase response signaling (Figure 2-13).

Plasma proteome profiling of psychiatric disorders

A total of 902 proteins were quantified for proteomic profiling, and the dynamic range spanned across 7 orders of magnitude (Figure 2-14a). The technical variations across groups were low; the Pearson's correlation value ranged from 0.93 to 0.98 (Figure 2-14b). In addition, the median coefficient of variation (CV) values for the normalized abundance of each group was less than 1.5 % in the technical replicates (Figure 2-14c). PCA plots revealed that plasma proteome composition was grossly different between groups (Figure 2-14d).

In the pairwise groups comparisons of MDD versus BD versus HC, 267 DEPs were determined, falling into 4 clusters. For MDD versus SCZ versus HC, 347 DEPs were determined, falling into 5 clusters, and for BD versus SCZ versus HC, 339 DEPs were determined, falling into 5 clusters (Figure 2-15).

Consistency of the proteomic features from the final MPM models between targeted proteomics and proteomic profiling data

The consistency in the statistical significance and expression patterns of the proteomic features in the final MPM models between targeted proteomics and proteomic profiling was further examined. In the MPM model of MDD versus BD, 4 proteins (ITIH2, TRFE, ALDOC, and SAA1) were included as DEPs. Only ITIH2 had consistent statistical significance [Post-hoc *P*-value = 0.001 for both proteomic platforms] and expression patterns [upregulated in MDD for both proteomic platforms] (Figure 2-15a, Table 2-9, and Table 2-12).

In the MPM model of MDD versus SCZ, 6 proteins (PROS, TFPI1, ITIH2, CBG, ALDOC, and SAA1) were included as DEPs. TFPI1 [Post-hoc *P*-value = 0.005 and 0.043 for targeted proteomics and proteomic profiling, respectively] and ITIH2 [Post-hoc *P*-value = 0.025 and 0.003 for targeted proteomics and proteomic profiling, respectively] had consistent statistical significance and expression patterns [upregulated in MDD for both proteomic platforms] (Figure 2-15b, Table 2-9, and Table 2-12).

Finally, in the MPM model of BD versus SCZ, only C1RL was included as DEPs. C1RL had consistent statistical significance [Post-hoc *P*-value = 0.002 and 0.003 for targeted proteomics and proteomic profiling, respectively] and expression patterns [upregulated in BD for both proteomic platforms] (Figure 2-15c, Table 2-9, and Table 2-12).

Consequently, TFPI1, ITIH2, and C1RL were determined to be important proteins that satisfied the consistency requirement. Alterations in their expression patterns of the profiling data were compared (Figure 2-16).

TFPI1 was upregulated in MDD and SCZ but downregulated in BD when compared with HC (MDD>SCZ>HC>BD). ITIH2 had no expression difference between MDD and HC but was downregulated in BD and SCZ when compared with HC (MDD≈HC>SCZ>BD). C1RL had no expression difference between BD and HC but was downregulated in SCZ and MDD when compared with HC (BD≈HC>MDD>SCZ).

3.3. Discussion

The diagnosis of major psychiatric disorders, including MDD, BD, and SCZ, primarily relies on subjective symptoms and behavioral observations. However, due to the complexity, heterogeneity of each disorder, and shared symptoms between them, it is sometimes difficult to differentiate these disorders objectively with high accuracy. Thus, the main aim of this study was to develop MPM models for differentiating these disorders. An additional aim was to integrate proteomic and self-report clinical features to compare its differentiation performance with clinician assessed clinical features.

When developing the final MPM models, it was important to consider the high-dimensional characteristics of the proteomic data when compared to a relatively smaller sample number (110). Thus, the efforts to overcome overfitting was important (111). In this study, initially discarding features irrelevant to the differentiation of the disorders, developing repetitive models based on machine learning methods with cross-validation, and the process of feature extraction/weighted model averaging, yielded generalizable MPM models. The applicability of these methods was demonstrated in a previous study (100). Consequently, the final MPM models for each pairwise group comparison performed reasonably in each data set. Furthermore, the models demonstrated similar performance in the subgroup analyses of BD, suggesting the model performance is unaffected by subtype or episode states of BD.

When focusing on proteomic feature selection, the need to control effects of certain covariates including demographics, medication, and chronicity of the disease/medication was important, as these differed between disease types. Thus, in the initial stage of developing MPM models, the study confined candidate proteomic features to those significantly associated with disease types if they were associated with significant covariates. In addition, the final features had low interdependence with current symptoms, suggesting its differentiation power when combined in the MPM models. Particularly for BD, the features were unrelated to depressive or manic symptoms, which should be recognized considering the dynamic differences of these symptoms. To summarize, majority of the markers could be considered as a trait marker, enabling differentiation between the disorders.

Subsequently, integration of the models based on current subjective symptoms and proteomic data enabled enhancement in its performance, and was comparable with models based on clinician rater scores. Of course, these models were constructed with current symptoms and proteomic data, so the lifetime diagnosis of MDD, BD, and SCZ based on these models has its limitations. To enhance performance, integrating known risk factors in differentiating these disorders like seasonality, atypical symptoms, family history, etc. should be considered. Additionally, controlling procedure batches and hospital effects would be important. Nevertheless, the ES models have potential clinical applicability, enabling objective differentiation

between these disorders. With further validation, these models could help differential diagnosis in situations when it is confusing or complicated.

Compared with a previous study, ITIH2 was replicated as a significantly increased protein in MDD when compared to BD (100). Smirnova et al. (2021) discovered AACT in the profiling data of SCZ but not in BD, which is in line with this study, as AACT was higher in SCZ versus BD (34). However, the direction of HEP2 was the opposite when compared with the study of Santa Cruz et al. (2019) (36). The differences in proteomic quantification approaches can be a reason for the discrepancy. Thus, the comparison with proteomic profiling data helps to focus on more consistent, reliable proteins. The level of TFPI1 was significantly different between each group. This protein is known to have an important role in inhibiting the extrinsic coagulation pathway (112). Its function in cancer has been previously studied (113), but its ability to differentiate major psychiatric disorders is a novel finding. The level of ITIH2 was upregulated in MDD and HC, when compared to BD and SCZ. As discussed previously, ITIH2 was a consistent key protein in differentiating MDD and BD (100). ITIH2 is a serine protease inhibitor with known anti-inflammatory properties (114). A previous study reported that the levels of ITIH2 were decreased in MDD when compared to HC, but the study sample of it was smaller (115). The level of C1RL was upregulated in BD and HC, when compared to MDD and SCZ. A recent study reported that the level of C1RL at age 12 was higher in those who experienced psychotic symptoms at the age of 18, when compared to

those without psychotic symptoms (116). The chronicity could explain the discrepancy of expression direction, because the complement pathway activity might differ before and after the onset of psychotic symptoms.

Although the proteomic features were grossly different when compared to previous reports, bioinformatics analysis results were generally replicated. Earlier studies differentiating MDD/SCZ from BD also implicated dysregulation of the coagulation and complement cascades (29, 36) which was significant in this study. Other significant pathways included neural signaling, and oxidative and inflammatory pathways, which have all been recapitulated in previous studies (100, 101). Although these pathways were related to the differentiation between MDD and BD (100, 101), the results of the current study suggest that they are also associated with SCZ. Another interesting result was the significance of the reelin signaling pathway. This pathway is known to have several important functions in the CNS. This includes the regulation of neuronal migration and synaptogenesis, and the pathway has been previously linked to major psychiatric disorders, including MDD, BD, and SCZ (117). However, the results of the bioinformatics analysis need cautious interpretation because they were based on proteins from peripheral blood, and not the CNS.

Approximately 500 ml of CSF enters the circulation each day (118), and in certain situations, the dysfunction of the blood-brain barrier suggests the occurrence of protein exchange between the CNS and the peripheral system (119). However, the expression of each individual protein differs, and

the blood is “neither perfectly correlated, nor uncorrelated” with the CNS (120). There is substantial evidence of the leaky blood-brain barrier in psychiatric diseases, and evidence of neural-derived exosomes which passes the blood-brain barrier and exits in peripheral blood (121). To summarize, evidence supports the possibility of differentiating psychiatric disorders with plasma proteins, but each individual protein needs its own consideration.

When matching the 43 plasma proteins of MPM models with a public database (THE HUMAN PROTEIN ATLAS) (122), 41 (except for MBL2 and UROM) were reported as detectable proteins in the human brain. However, when comparing consistency with a previous study of the cerebrospinal fluid (123), only 2 proteins (AMBP, ITIH2) of the MPM models for MDD vs BD were DEPs of the cerebrospinal fluid study, and the expression direction was opposite. Seven proteins (ALS, CBPB2, CBG, IBP3, ITIH2, PROC1, PROS) of the MPM models for MDD vs SCZ were DEPs of the cerebrospinal fluid study, with CBPB2 and PROS expressing consistent upregulation in SCZ. There was no consistent protein for BD vs SCZ. However, considering the skewed expression patterns of DEPs in certain pairwise comparisons (i.e. MDD vs BD and MDD vs SCZ) of the previous cerebrospinal fluid study, as well as heterogeneous study designs, sample type/size, and analytical methods between both studies, future studies that focus on correlations between the proteomes of blood and CNS are warranted.

There were several limitations of this study. First, despite the expanded sample size compared with previous studies, it remained a major

limitation. As proteomic features are known to be associated with symptom severity and other diseases, the necessity to exclude certain conditions resulted in a smaller sample size than expected. Even though multiple efforts including feature extraction and model averaging were performed, overfitting occurred, as the performance in other sets were lower than in the training set. Second, there could be other potential confounders. Specifically, medication categorization was broad, and specific dosages and durations were not controlled. Although the study confined proteomic features that were associated with differential diagnosis when controlling related covariates, other covariates might have influenced the differentiation performance. Third, the cross-sectional design limited causality interpretation. Longitudinal studies are required to observe diagnostic alterations of MDD, BD, and SCZ, and serial quantitation of proteins at various time points would enable differentiation between state and trait markers. Finally, although the study tried to determine important protein candidates for differentiating MDD, BD, and SCZ, other targets might have been overlooked.

Nevertheless, the study has several strengths. It is the first report to differentiate MDD, BD and SCZ with demonstrating diagnostic potential by integrating proteomic data and clinical symptom data. Through high-throughput MS-based proteomics, numerous targets were quantified simultaneously in 675 individual samples, greater than most previous proteomic studies. Moreover, important and reliable proteomic features were

selected by comparing the consistency of the direction of expression pattern, and statistical significance between different quantification platforms.

In conclusion, the study demonstrated the differential potential of plasma proteomic data, and the viability of integrating proteomic and clinical data in differentiating MDD, BD and SCZ. The proteomic features, and related functions and pathways, might expand the understanding of the pathophysiology of these disorders. Additional studies involving a larger sample size, with longitudinal designs to determine protein profiles from those whose diagnosis changes afterwards, are needed.

Chapter 4. Final conclusion

Through Study 1 and 2, the viability of differentiating major psychiatric disorders based on proteomics data was demonstrated. Study 1 was based on drug-free patients with MDD and BD. As medication can affect proteomic nature, including only those whom are drug-free or drug naïve patients is ideal. However, the relatively smaller sample size also has its limitations in interpreting the results. Thus, in Study 2, medication was controlled as a covariate, and included patients as in the real world. Other covariates that are known to influence proteomic expressions were also carefully considered.

Study 2 was expanded not only from increasing the sample size, but from expanding the sample groups (including SCZ, HC), and using dual quantification platforms. This approach enabled multiple comparison between disease groups, and verification of more consistent and reliable proteomic features. However, the proteomic features did not overlap between the studies. Study 1 was based on statistical significance, and Study 2 was based on differentiation ability. The studies additionally differed in demographic characteristics, and quantification platforms, which could yield different results. Nevertheless, functional analysis with IPA yielded similar results based on broad functions and pathways.

An additional approach in Study 2 was to integrate subjective symptoms with the proteomic data, to see if the differential performance

could be enhanced. As described previously, the association of lifetime diagnosis with current proteomic features and symptoms has its limitations. Nevertheless, this integrative approach enhanced the performance comparable of it to models based on four clinician rater scales.

The results in this study indicate that these approaches have potential in differentiating major psychiatric disorders. Future studies should validate their performance in a larger independent set with longitudinal designs.

Table 1-1. Demographics and clinical characteristics of drug free major depressive disorder and bipolar disorder

Characteristics	MDD (n=15)	BD (n=10)	Statistics	*P*-value[a]
Age, mean ± SD, years	28.53 ± 8.04	25.10 ± 4.91	Z=-1.059	0.29
Sex (Female), n (%)	10 (66.7%)	7 (70.0%)	Fisher's exact test	>0.99
BMI, mean ± SD, kg/m^2	22.42 ± 5.18	22.41 ± 5.17	Z=-0.055	0.96
Current smoker, *n* (%)	1 (6.7%)	3 (30.0%)	Fisher's exact test	0.27
HAM-D, mean ± SD	15.33 ± 4.61	12.10 ± 3.54	**Z=-2.320**	**0.02**
Duration from onset, mean ± SD, years	3.73 ± 4.57	4.80 ± 2.94	Z=-1.654	0.10
Duration from first medication, mean ± SD, years	0.33 ± 0.49	1.90 ± 3.28	Z=-0.974	0.33

[a]*P*-value < 0.05 is considered statistically significant, denoted by bold font

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SD = standard distribution, BMI = body mass index, HAM-D = Hamilton depression rating scale

Table 1-2. Differentially expressed proteins between drug free major depressive disorder and bipolar disorder

Majority protein IDs[a]	Protein names	Gene names	Fold change[b]	t-test statistics	*P*-value[c]
Overexpressed in MDD					
P51149;C9J592;C9J8S3	Ras-related protein Rab-7a	RAB7A	2.729	**5.975**	**< 0.001**
E9PF63;O75116	Rho-associated protein kinase 2	ROCK2	2.321	**4.782**	**< 0.001**
P01763	Ig heavy chain V-III region WEA	IGHV3-48	1.205	3.668	0.001
P08571;D6RFL4	Monocyte differentiation antigen CD14; Monocyte differentiation antigen CD14, urinary form; Monocyte differentiation antigen CD14, membrane-bound form	CD14	1.266	2.875	0.009
P23083	Ig heavy chain V-I region V35	IGHV10R15-1	2.103	2.739	0.011
A0A075B6Q5	Ig heavy variable 3-64	IGHV3-64	2.252	2.698	0.013
P01780	Ig heavy chain V-III region JON	IGHV3-7	0.941	2.404	0.025
Overexpressed in BD					
E7ESC6;Q9UIA9	Exportin-7	XPO7	-2.999	**-4.520**	**< 0.001**
P03950	Angiogenin	ANG	-1.714	-3.636	0.001
A0A075B6J4	Ig lambda variable 3-25	IGLV3-25	-2.550	-3.326	0.003
Q96Q89-4;Q96Q89-3;Q96Q89;Q96Q89-2	Kinesin-like protein KIF20B	KIF20B	-1.298	-2.718	0.012
A0A096LPE2;P35542;A0A087X0E2	Serum amyloid A-4 protein Serum amyloid A protein	SAA2-SAA4;SAA4	-0.660	-2.64	0.015
P05160	Coagulation factor XIII B chain	F13B	-1.120	-2.315	0.030
D6REX5;P49908;D6RIS9	Selenoprotein P	SEPP1;SELENOP	-0.881	-2.088	0.048

[a]UniProt accession number
[b]Fold change calculated by the difference of the logarithmic$_{(2)}$ transferred intensity
[c]*P*-value < 0.05 is considered statistically significant, denoted by bold font
Abbreviations: MDD = major depressive disorder, BD = bipolar disorder

Table 2-1. Demographics and clinical characteristics of the total set of major psychiatric disorders and healthy controls

Characteristics	SCZ n=171	BD n=170	MDD n=174	HC n=160	Statistics	*P*-value[a]	Post-hoc analysis[b]
Sex: Male	74 (43.3%)	58 (34.1%)	61 (35.1%)	48 (30.0%)	$\chi2 = 6.755$	0.08	
Age	39.21 ± 11.80	34.31 ± 12.27	36.31 ± 13.23	35.94 ± 11.20	**F = 4.784**	**0.003**	SCZ > BD
BMI	25.84 ± 4.84	24.57 ± 4.41	23.46 ± 4.04	22.08 ± 2.72	**F = 25.412**	**< 0.001**	SCZ > BD = MDD > HC
Blood collection time: AM	81 (47.4%)	50 (29.4%)	55 (31.6%)	69 (43.1%)	**$\chi2 = 16.516$**	**0.001**	
Fasting time: ≥8 hours	39 (22.8%)	33 (19.4%)	42 (24.1%)	87 (54.4%)	**$\chi2 = 61.650$**	**< 0.001**	
Exercise: moderate intensity	60 (35.1%)	68 (40.0%)	52 (29.9%)	116 (72.5%)	**$\chi2 = 73.476$**	**< 0.001**	
Alcohol drinking: ≥ once a week	32 (18.7%)	60 (35.3%)	63 (36.2%)	75 (46.9%)	**$\chi2 = 30.098$**	**< 0.001**	
Smoking: current smoker	41 (24.0%)	53 (31.2%)	59 (33.9%)	8 (5.0%)	**$\chi2 = 46.018$**	**< 0.001**	
Duration from first onset	12.48 ± 10.02	9.39 ± 8.77	6.73 ± 7.75*		**F = 17.985**	**< 0.001**	SCZ > BD > MDD
Duration from first medication	11.72 ± 9.89	6.91 ± 8.34	3.68 ± 6.03		**F = 41.613**	**< 0.001**	SCZ > BD > MDD
Medication							
Antipsychotics	166 (97.1%)	130 (76.5%)	72 (41.4%)		**$\chi2 = 134.298$**	**< 0.001**	
Lithium/Anticonvulsants	28 (16.4%)	125 (73.5%)	24 (13.8%)		**$\chi2 = 172.782$**	**< 0.001**	
Antidepressant	37 (21.6%)	44 (26.0%)	145 (83.3%)*		**$\chi2 = 166.118$**	**< 0.001**	
Benzodiazepines/hypnotics	104 (60.8%)	106 (62.4%)	117 (67.2%)		$\chi2 = 1.678$	0.43	
Clinician rater score							
BPRS	43.64 ± 11.98	39.47 ± 8.15	40.61 ± 6.98	27.06 ± 3.75	**F = 126.282**	**< 0.001**	SCZ > BD = MDD > HC
YMRS	4.49 ± 5.81	5.68 ± 6.95	1.87 ± 2.55	1.19 ± 2.11	**F = 32.242**	**< 0.001**	SCZ = BD > MDD = HC
MADRS	13.80 ± 9.54	17.48 ± 10.49	26.05 ± 9.89	4.14 ± 4.24	**F = 171.012**	**< 0.001**	MDD > BD > SCZ > HC
HAM-A	8.75 ± 6.67	9.88 ± 5.98	14.94 ± 7.21	2.27 ± 2.01	**F = 130.165**	**< 0.001**	MDD > SCZ = BD > HC
Self-report scale							
Symptom Checklist-90-Revised							

Somatization subscale	0.77 ± 0.75*	0.86 ± 0.74***	1.39 ± 0.94**	0.17 ± 0.21	**F = 81.329**	**< 0.001**	MDD > SCZ = BD > HC
Obsessive-compulsive subscale	1.26 ± 0.90*	1.48 ± 0.87***	1.90 ± 0.84**	0.38 ± 0.36	**F = 111.170**	**< 0.001**	MDD > BD > SCZ > HC
Interpersonal sensitivity subscale	1.22 ± 0.90*	1.29 ± 0.85***	1.63 ± 0.91**	0.30 ± 0.32	**F = 85.622**	**< 0.001**	MDD > SCZ = BD > HC
Depression subscale	1.22 ± 0.96*	1.60 ± 0.94***	2.22 ± 0.92**	0.29 ± 0.35	**F = 154.116**	**< 0.001**	MDD > BD > SCZ > HC
Anxiety subscale	1.00 ± 0.89*	1.15 ± 0.85***	1.62 ± 0.95**	0.13 ± 0.24	**F = 100.632**	**< 0.001**	MDD > SCZ = BD > HC
Hostility subscale	0.73 ± 0.88*	0.97 ± 0.94***	1.18 ± 0.94**	0.11 ± 0.21	**F = 53.672**	**< 0.001**	BD = MDD > SCZ > HC
Phobic anxiety subscale	0.74 ± 0.79*	0.73 ± 0.75***	1.14 ± 0.99**	0.05 ± 0.14	**F = 60.264**	**< 0.001**	MDD > SCZ = BD > HC
Paranoid ideation subscale	1.05 ± 0.97*	0.92 ± 0.83***	1.18 ± 0.91**	0.15 ± 0.29	**F = 54.784**	**< 0.001**	SCZ, BD, MDD > HC MDD > BD, MDD = SCZ, BD = SCZ
Psychoticism subscale	1.03 ± 0.88*	0.94 ± 0.75***	1.22 ± 0.79**	0.09 ± 0.20	**F = 81.867**	**< 0.001**	SCZ, BD, MDD > HC MDD > BD, MDD = SCZ, BD = SCZ
Overeating item	1.13 ± 1.18	1.17 ± 1.34***	1.10 ± 1.28**	0.34 ± 0.54	**F = 19.519**	**< 0.001**	SCZ = BD = MDD > HC

* n=1 missing, ** n=2 missing, *** n=3 missing

[a] *P*-value < 0.05 is considered statistically significant, denoted by bold font

[b] Levels of statistical significance of post-hoc analysis based on Tukey's HSD are presented as equality and inequality sign (=, >).
The equality sign (=) signifies no statistical significance, inequality sign (>) denotes statistical significance between groups.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, HC = healthy controls, BMI = body mass index, BPRS = Brief Psychiatric Rating Scale
YMRS = Young Mania Rating Scale, MADRS = Montgomery-Asberg Depression Rating Scale, HAM-A = Hamilton Anxiety Scale

Table 2-2. Demographics and clinical characteristics of the training set of major psychiatric disorders

Characteristics	SCZ n=102	BD n=102	MDD n=104	Statistics	*P*-value[a]	Post-hoc analysis[b]
Sex: Male	41 (40.2%)	33 (32.4%)	38 (36.5%)	$\chi2 = 1.358$	0.51	
Age	40.37 ± 11.35	34.23 ± 12.58	36.32 ± 13.19	**F = 6.481**	**0.002**	SCZ > BD
BMI	26.58 ± 4.84	24.42 ± 4.35	23.24 ± 3.99	**F = 15.242**	**< 0.001**	SCZ > BD = MDD
Blood collection time: AM	49 (48.0%)	27 (26.5%)	30 (28.8%)	**$\chi2 = 12.669$**	**0.002**	
Fasting time: ≥ 8 hours	29 (28.4%)	20 (19.6%)	24 (23.1%)	$\chi2 = 2.230$	0.33	
Exercise: moderate intensity	32 (31.4%)	37 (36.3%)	33 (31.7%)	$\chi2 = 0.689$	0.71	
Alcohol drinking: ≥ once a week	19 (18.6%)	37 (36.3%)	41 (39.4%)	**$\chi2 = 11.938$**	**0.003**	
Smoking: current smoker	24 (23.5%)	35 (34.3%)	35 (33.7%)	$\chi2 = 3.525$	0.17	
Duration from first onset	13.68 ± 10.38	10.41 ± 9.30	7.01 ± 8.06	**F = 13.275**	**< 0.001**	SCZ > BD > MDD
Duration from first medication	13.04 ± 10.23	7.83 ± 9.07	3.56 ± 6.14	**F = 31.105**	**< 0.001**	SCZ > BD > MDD
Medication						
Antipsychotics	99 (97.1%)	79 (77.5%)	43 (41.3%)	**$\chi2 = 81.304$**	**< 0.001**	
Lithium/Anticonvulsants	19 (18.6%)	69 (67.6%)	10 (9.6%)	**$\chi2 = 92.169$**	**< 0.001**	
Antidepressant	25 (24.5%)	29 (28.4%)	88 (84.6%)	**$\chi2 = 94.037$**	**< 0.001**	
Benzodiazepines/hypnotics	58 (56.9%)	63 (61.8%)	71 (68.3%)	$\chi2 = 2.875$	0.24	
Clinician rater score						
BPRS	44.31 ± 11.50	39.83 ± 7.51	40.41 ± 6.89	**F = 7.733**	**0.001**	SCZ > BD = MDD
YMRS	4.64 ± 5.57	5.29 ± 6.11	1.96 ± 2.76	**F = 12.767**	**< 0.001**	SCZ = BD > MDD
MADRS	14.22 ± 9.62	17.70 ± 10.40	26.06 ± 10.06	**F = 37.995**	**< 0.001**	MDD > BD > SCZ
HAM-A	9.35 ± 7.33	10.06 ± 5.44	15.57 ± 7.22	**F = 26.455**	**< 0.001**	MDD > SCZ = BD
Self-report scale						

Symptom Checklist-90-Revised						
Somatization subscale	0.76 ± 0.68*	0.86 ± 0.70***	1.41 ± 0.95*	**F = 20.424**	**< 0.001**	MDD > SCZ = BD
Obsessive-compulsive subscale	1.24 ± 0.83*	1.51 ± 0.89***	1.95 ± 0.87*	**F = 17.653**	**< 0.001**	MDD > SCZ = BD
Interpersonal sensitivity subscale	1.26 ± 0.84*	1.31 ± 0.84***	1.68 ± 0.88*	**F = 7.617**	**0.001**	MDD > SCZ = BD
Depression subscale	1.21 ± 0.90*	1.67 ± 0.96***	2.22 ± 0.93*	**F = 30.291**	**< 0.001**	MDD > BD > SCZ
Anxiety subscale	1.02 ± 0.88*	1.14 ± 0.82***	1.63 ± 0.96*	**F = 13.719**	**< 0.001**	MDD > SCZ = BD
Hostility subscale	0.71 ± 0.78*	1.00 ± 0.95***	1.19 ± 0.98*	**F = 7.096**	**0.001**	MDD > SCZ
Phobic anxiety subscale	0.78 ± 0.80*	0.67 ± 0.69***	1.16 ± 1.01*	**F = 9.119**	**< 0.001**	MDD > SCZ = BD
Paranoid ideation subscale	1.04 ± 0.87*	0.95 ± 0.82***	1.23 ± 0.96*	F = 2.516	0.08	
Psychoticism subscale	1.02 ± 0.82*	0.93 ± 0.76***	1.22 ± 0.82*	**F = 3.434**	**0.033**	MDD > BD
Overeating item	1.19 ± 1.16	0.98 ± 1.28***	1.06 ± 1.17*	F = 0.757	0.47	

* n=1 missing, *** n=3 missing

[a] *P*-value < 0.05 is considered statistically significant, denoted by bold font

[b] Levels of statistical significance of post-hoc analysis based on Tukey's HSD are presented as equality and inequality sign (=, >).
The equality sign (=) signifies no statistical significance, inequality sign (>) denotes statistical significance between groups.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, HC = healthy controls, BMI = body mass index
BPRS = Brief Psychiatric Rating Scale, YMRS = Young Mania Rating Scale, MADRS = Montgomery-Asberg Depression Rating Scale
HAM-A = Hamilton Anxiety Scale

Table 2-3. Demographics and clinical characteristics of the validation set of major psychiatric disorders

Characteristics	SCZ n=34	BD n=34	MDD n=35	Statistics	*P*-value[a]	Post-hoc analysis[b]
Sex: Male	15 (44.1%)	10 (29.4%)	8 (22.9%)	$\chi2 = 3.741$	0.15	
Age	36.88 ± 11.91	37.21 ± 12.73	36.63 ± 12.94	F = 0.018	0.98	
BMI	25.18 ± 5.38	23.87 ± 3.90	23.61 ± 4.08	F = 1.203	0.31	
Blood collection time: AM	18 (52.9%)	10 (29.4%)	15 (42.9%)	$\chi2 = 3.897$	0.14	
Fasting time: ≥ 8 hours	7 (20.6%)	3 (8.8%)	11 (31.4%)	$\chi2 = 5.431$	0.07	
Exercise: moderate intensity	13 (38.2%)	17 (50.0%)	9 (25.7%)	$\chi2 = 4.326$	0.12	
Alcohol drinking: ≥ once a week	10 (29.4%)	14 (41.2%)	10 (28.6%)	$\chi2 = 1.536$	0.46	
Smoking: current smoker	12 (35.3%)	6 (17.6%)	13 (37.1%)	$\chi2 = 3.767$	0.15	
Duration from first onset	11.15 ± 9.07	9.12 ± 9.12	7.00 ± 7.67	F = 1.989	0.14	
Duration from first medication	10.38 ± 9.54	7.56 ± 8.85	4.83 ± 6.06	**F = 3.889**	**0.024**	SCZ > MDD
Medication						
Antipsychotics	34 (100.0%)	23 (67.6%)	17 (48.6%)	**$\chi2 = 22.992$**	**< 0.001**	
Lithium/Anticonvulsants	4 (11.8%)	26 (76.5%)	6 (17.1%)	**$\chi2 = 38.702$**	**< 0.001**	
Antidepressant	6 (17.6%)	10 (30.3%)	30 (85.7%)*	**$\chi2 = 36.585$**	**< 0.001**	
Benzodiazepines/hypnotics	24 (70.6%)	19 (55.9%)	23 (65.7%)	$\chi2 = 1.659$	0.44	
Clinician rater score						
BPRS	40.15 ± 11.82	36.09 ± 6.94	41.60 ± 6.93	**F = 3.574**	**0.032**	MDD > BD
YMRS	3.47 ± 5.09	5.21 ± 6.07	1.51 ± 1.96	**F = 5.345**	**0.006**	BD > MDD
MADRS	13.47 ± 11.16	15.21 ± 10.19	27.63 ± 8.96	**F = 20.112**	**< 0.001**	MDD > SCZ = BD
HAM-A	7.53 ± 5.82	8.50 ± 5.84	14.03 ± 5.71	**F = 12.712**	**< 0.001**	MDD > SCZ = BD
Self-report scale						

Symptom Checklist-90-Revised						
Somatization subscale	0.83 ± 0.80	0.87 ± 0.84	1.58 ± 0.89	**F = 8.642**	**< 0.001**	MDD > SCZ = BD
Obsessive-compulsive subscale	1.42 ± 1.04	1.30 ± 0.85	2.06 ± 0.77	**F = 7.180**	**0.001**	MDD > SCZ = BD
Interpersonal sensitivity subscale	1.25 ± 0.92	1.15 ± 0.82	1.84 ± 0.84	F = 6.418	**0.002**	MDD > SCZ = BD
Depression subscale	1.45 ± 1.08	1.40 ± 0.88	2.46 ± 0.82	**F = 14.246**	**< 0.001**	MDD > SCZ = BD
Anxiety subscale	1.09 ± 1.02	1.09 ± 0.95	1.89 ± 0.95	**F = 7.722**	**0.001**	MDD > SCZ = BD
Hostility subscale	0.90 ± 1.09	0.84 ± 0.91	1.38 ± 0.90	**F = 3.175**	**0.046**	MDD = BD = SCZ
Phobic anxiety subscale	0.70 ± 0.77	0.72 ± 0.93	1.29 ± 0.98	**F = 4.851**	**0.010**	MDD > SCZ = BD
Paranoid ideation subscale	1.08 ± 1.12	0.87 ± 0.82	1.29 ± 0.79	F = 1.777	0.17	
Psychoticism subscale	1.10 ± 1.02	0.91 ± 0.79	1.43 ± 0.65	**F = 3.432**	**0.036**	MDD > BD
Overeating item	1.03 ± 1.24	1.62 ± 1.37	1.14 ± 1.40	F = 1.851	0.16	

* n=1 missing

[a] *P*-value < 0.05 is considered statistically significant, denoted by bold font

[b] Levels of statistical significance of post-hoc analysis based on Tukey's HSD are presented as equality and inequality sign (=, >).
The equality sign (=) signifies no statistical significance, inequality sign (>) denotes statistical significance between groups.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, HC = healthy controls, BMI = body mass index
BPRS = Brief Psychiatric Rating Scale, YMRS = Young Mania Rating Scale, MADRS = Montgomery-Asberg Depression Rating Scale
HAM-A = Hamilton Anxiety Scale

Table 2-4. Demographics and clinical characteristics of the independent test set of major psychiatric disorders

Characteristics	SCZ n=35	BD n=34	MDD n=35	Statistics	*P*-value[a]	Post-hoc analysis[b]
Sex: Male	18 (51.4%)	15 (44.1%)	15 (42.9%)	$\chi2 = 0.602$	0.74	
Age	38.06 ± 12.86	31.65 ± 10.39	35.94 ± 13.98	F = 2.342	0.10	
BMI	24.33 ± 3.89	25.71 ± 4.92	23.96 ± 4.20	F = 1.547	0.22	
Blood collection time: AM	14 (40.0%)	13 (38.2%)	10 (28.6%)	$\chi2 = 1.153$	0.56	
Fasting time: ≥ 8 hours	3 (8.6%)	10 (29.4%)	7 (20.0%)	$\chi2 = 4.843$	0.09	
Exercise: moderate intensity	15 (42.9%)	14 (41.2%)	10 (28.6%)	$\chi2 = 1.815$	0.40	
Alcohol drinking: ≥ once a week	3 (8.6%)	9 (26.5%)	12 (34.3%)	**$\chi2 = 6.846$**	**0.033**	
Smoking: current smoker	5 (14.3%)	12 (35.3%)	11 (31.4%)	$\chi2 = 4.413$	0.11	
Duration from first onset	10.29 ± 9.55	6.59 ± 5.88	5.62 ± 6.93*	**F = 3.605**	**0.031**	SCZ > MDD
Duration from first medication	9.17 ± 8.72	3.47 ± 3.34	2.91 ± 5.63	**F = 10.483**	**< 0.001**	SCZ > MDD = BD
Medication						
Antipsychotics	33 (94.3%)	28 (82.4%)	12 (34.3%)	**$\chi2 = 33.681$**	**< 0.001**	
Lithium/Anticonvulsants	5 (14.3%)	30 (88.2%)	8 (22.9%)	**$\chi2 = 46.326$**	**< 0.001**	
Antidepressant	6 (17.1%)	5 (14.7%)	27 (77.1%)	**$\chi2 = 37.553$**	**< 0.001**	
Benzodiazepines/hypnotics	22 (62.9%)	24 (70.6%)	23 (65.7%)	$\chi2 = 0.471$	0.79	
Clinician rater score						
BPRS	45.09 ± 13.16	41.76 ± 10.10	40.23 ± 7.41	F = 1.958	0.15	
YMRS	5.06 ± 7.07	7.29 ± 9.64	1.94 ± 2.46	**F = 5.069**	**0.008**	BD > MDD
MADRS	12.89 ± 7.56	19.09 ± 10.96	24.43 ± 10.29	**F = 12.406**	**< 0.001**	MDD = BD > SCZ
HAM-A	8.17 ± 5.22	10.71 ± 7.44	13.97 ± 8.41	**F = 5.797**	**0.004**	MDD > SCZ
Self-report scale						

Symptom Checklist-90-Revised						
Somatization subscale	0.75 ± 0.88	0.85 ± 0.75	1.16 ± 0.94*	F = 2.147	0.12	
Obsessive-compulsive subscale	1.14 ± 0.94	1.60 ± 0.82	1.56 ± 0.75*	**F = 3.129**	**0.048**	MDD = BD = SCZ
Interpersonal sensitivity subscale	1.11 ± 1.03	1.39 ± 0.90	1.25 ± 0.97*	F = 0.727	0.49	
Depression subscale	1.02 ± 0.97	1.62 ± 0.96	2.00 ± 0.93*	**F = 9.262**	**< 0.001**	MDD = BD > SCZ
Anxiety subscale	0.85 ± 0.81	1.23 ± 0.83	1.31 ± 0.86*	F = 2.976	0.06	
Hostility subscale	0.62 ± 0.90	0.99 ± 0.97	0.96 ± 0.84*	F = 1.763	0.18	
Phobic anxiety subscale	0.63 ± 0.78	0.91 ± 0.68	0.91 ± 0.92*	F = 1.368	0.26	
Paranoid ideation subscale	1.05 ± 1.12	0.91 ± 0.89	0.91 ± 0.83*	F = 0.242	0.79	
Psychoticism subscale	1.02 ± 0.92	0.97 ± 0.71	1.02 ± 0.78*	F = 0.040	0.96	
Overeating item	1.06 ± 1.21	1.26 ± 1.42	1.18 ± 1.51*	F = 0.196	0.82	

* n=1 missing

[a] *P*-value < 0.05 is considered statistically significant, denoted by bold font

[b] Levels of statistical significance of post-hoc analysis based on Tukey's HSD are presented as equality and inequality sign (=, >).
The equality sign (=) signifies no statistical significance, inequality sign (>) denotes statistical significance between groups.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, HC = healthy controls, BMI = body mass index
BPRS = Brief Psychiatric Rating Scale, YMRS = Young Mania Rating Scale, MADRS = Montgomery-Asberg Depression Rating Scale
HAM-A = Hamilton Anxiety Scale

Table 2-5. Candidate proteomic features for differentiation of major psychiatric disorders from the training set

MDD vs BD (23)			MDD vs SCZ (29)			BD vs SCZ (30)		
ID[a]	Protein	Peptide	ID[a]	Protein	Peptide	ID[a]	Protein	Peptide
P09972	ALDOC	ALQASALNAWR	P04217	A1BG	LLELTGPK	P04217	A1BG	LLELTGPK
P15144	AMPN	AQIINDAFNLASAHK	P19823	ITIH2	IQPSGGTNINEALLR	P02787	TRFE	ASYLDCIR
Q86TY3	ARMD4	TVVPSITR	P52758	RIDA	AAYQVAALPK	Q99460	PSMD1	VSTAVLSITAK
Q9NZP8	C1RL	GSEAINAPGDNPAK	P35542	SAA4	GPGGVWAAK	P05546	HEP2	TLEAQLTPR
P55290	CAD13	INNTHALVSLLQNLNK	P09972	ALDOC	ALQASALNAWR	P05155	IC1	TTFDPK
P08571	CD14	VLDLSCNR	P24593	IBP5	AVYLPNCDR	P01011	AACT	DEELSCTVVELK
P11597	CETP	ASYPDITGEK	P07225	PROS	NNLELSTPLK	Q8TDL5	BPIB1	ALGFEAAESSLTK
Q96KN2	CNDP1	AIHLDLEEYR	P35858	ALS	DFALQNPSAVPR	P08185	CBG	HLVALSPK
Q03692	COAA1	GTHVWVGLYK	P17936	IBP3	YGQPLPGYTTK	P00751	CFAB	DISEVVTPR
O60716	CTND1	GYELLFQPEVVR	P04070	PROC	TFVLNFIK	P24593	IBP5	AVYLPNCDR
Q08345	DDR1	LHLVALVGTQGR	P08185	CBG	HLVALSPK	P13473	LAMP2	IPLNDLFR
P28845	DHI1	VIVTGASK	P22352	GPX3	FYTFLK	Q8IYB8	SUV3	LLNLEGFPSGSQSR
P09172	DOPO	TPEGLTLLFK	Q15262	PTPRK	QNVVDVFHAVK	Q01082	SPTB2	LTVQTK
P05155	IC1	TTFDPK	P15144	AMPN	AQIINDAFNLASAHK	P15144	AMPN	AQIINDAFNLASAHK
Q9NPH3	IL1AP	NEVWWTIDGK	Q96IY4	CBPB2	DTGTYGFLLPER	P55290	CAD13	INNTHALVSLLQNLNK
P19823	ITIH2	IQPSGGTNINEALLR	Q9Y210	TRPC6	LGILGSHEDLSK	P09172	DOPO	VISTLEEPTPQCPTSQGR
P13473	LAMP2	IPLNDLFR	P10646	TFPI1	IAYEEIFVK	O15551	CLD3	DFYNPVVPEAQK
P61916	NPC2	LVVEWQLQDDK	P04278	SHBG	TSSSFEVR	Q08345	DDR1	LHLVALVGTQGR
O60486	PLXC1	LNTIGHYEISNGSTIK	P54802	ANAG	DFCGCHVAWSGSQLR	P07911	UROM	VLNLGPITR
P62826	RAN	FNVWDTAGQEK	Q03692	COAA1	GTHVWVGLYK	O15354	GPR37	ISPDLPDTIYVLALTYDSAR

P0DJI8	SAA1	FFGHGAEDSLADQAANEWGR	P0DJI8	SAA1	FFGHGAEDSLADQAANEWGR	Q13283	G3BP1	AVYLPNCDR
Q01082	SPTB2	LTVQTK	Q8IYB8	SUV3	LLNLEGFPSGSQSR	O60486	PLXC1	LNTIGHYEISNGSTIK
P02787	TRFE	ASYLDCIR	Q86TY3	ARMD4	TVVPSITR	Q9NZP8	C1RL	GSEAINAPGDNPAK
			P25774	CATS	YTELPYGR	Q96IY4	CBPB2	DTHTYGFLLPER
			P05023	AT1A1	IVEIPFNSTNK	Q96KN2	CNDP1	AIHLDLEEYR
			P20142	PEPC	AECGLGVPTTR	P10646	TFPI1	IAYEEIFVK
			Q14894	CRYM	TVVPVTK	Q9NY15	STAB1	SLEAQGNSSHLDADTVR
			Q13976	KGP1	EEEIQELK	P62888	RL30	SLESINSR
			P14618	KPYM	IYVDDGLISLQVK	P61916	NPC2	LVVEWQLQDDK
						P11226	MBL2	FQASVATPR

[a]UniProt accession number protein entry
Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia

Table 2-6. Unique models originating from the combinations of selected features for each pairwise group comparison

MDD vs BD	Model #	Unique model (combination of selected features)	Number of features	Frequency	Probability
	1	ALDOC_ALQASALNAWR+ AMPN_AQIINDAFNLASAHK+ ARMD4_TVVPSITR+ C1RL_GSEAINAPGDNPAK+ CAD13_INNTHALVSLLQNLNK+ CD14_VLDLSCNR+ CETP_ASYPDITGEK+ CNDP1_AIHLDLEEYR+ COAA1_GTHVWVGLYK+ CTND1_GYELLFQPEVVR+ DDR1_LHLVALVGTQGR+ DHI1_VIVTGASK+ DOPO_TPEGLTLLFK+ IC1_TTFDPK+ IL1AP_NEVWWTIDGK+ ITIH2_IQPSGGTNINEALLR+ LAMP2_IPLNDLFR+ NPC2_LVVEWQLQDDK+ PLXC1_LNTIGHYEISNGSTIK+ RAN_FNVWDTAGQEK+ SAA1_FFGHGAEDSLADQAANEWGR+ SPTB2_LTVQTK+ TRFE_ASYLDCIR+(intercept)	23	97	9.800E-01
	2	ALDOC_ALQASALNAWR+ AMPN_AQIINDAFNLASAHK+ ARMD4_TVVPSITR+ C1RL_GSEAINAPGDNPAK+ CAD13_INNTHALVSLLQNLNK+ CETP_ASYPDITGEK+ CNDP1_AIHLDLEEYR+ COAA1_GTHVWVGLYK+ CTND1_GYELLFQPEVVR+ DDR1_LHLVALVGTQGR+ DHI1_VIVTGASK+ DOPO_TPEGLTLLFK+ IC1_TTFDPK+ IL1AP_NEVWWTIDGK+ ITIH2_IQPSGGTNINEALLR+ LAMP2_IPLNDLFR+ NPC2_LVVEWQLQDDK+ PLXC1_LNTIGHYEISNGSTIK+ RAN_FNVWDTAGQEK+ SAA1_FFGHGAEDSLADQAANEWGR+ SPTB2_LTVQTK+ TRFE_ASYLDCIR+(intercept)	22	1	3.19E-05
	3	ALDOC_ALQASALNAWR+ AMPN_AQIINDAFNLASAHK+ ARMD4_TVVPSITR+ C1RL_GSEAINAPGDNPAK+ CAD13_INNTHALVSLLQNLNK+ CETP_ASYPDITGEK+ COAA1_GTHVWVGLYK+ CTND1_GYELLFQPEVVR+ DDR1_LHLVALVGTQGR+ DOPO_TPEGLTLLFK+ IC1_TTFDPK+ IL1AP_NEVWWTIDGK+ ITIH2_IQPSGGTNINEALLR+ NPC2_LVVEWQLQDDK+ RAN_FNVWDTAGQEK+ SAA1_FFGHGAEDSLADQAANEWGR+ TRFE_ASYLDCIR+(intercept)	17	1	4.69E-09
	4	ALDOC_ALQASALNAWR+ AMPN_AQIINDAFNLASAHK+ ARMD4_TVVPSITR+ C1RL_GSEAINAPGDNPAK+ CAD13_INNTHALVSLLQNLNK+ CETP_ASYPDITGEK+ COAA1_GTHVWVGLYK+ CTND1_GYELLFQPEVVR+ DDR1_LHLVALVGTQGR+ DOPO_TPEGLTLLFK+ IC1_TTFDPK+ IL1AP_NEVWWTIDGK+ ITIH2_IQPSGGTNINEALLR+ NPC2_LVVEWQLQDDK+ RAN_FNVWDTAGQEK+ SAA1_FFGHGAEDSLADQAANEWGR+ TRFE_ASYLDCIR+ LAMP2_IPLNDLFR+(intercept)	18	1	5.65E-06

MDD vs SCZ	Model #	Unique model (combination of selected features)	Number of features	Frequency	Probability
	1	ALDOC_ALQASALNAWR+ ALS_DFALQNPSAVPR+ ANAG_DFCGCHVAWSGSQLR+ AT1A1_IVEIPFNSTNK+ CATS_YTELPYGR+ CBG_HLVALSPK+ CBPB2_DTGTYGFLLPER+ COAA1_GTHVWVGLYK+ CRYM_TVVPVTK+ GPX3_FYTFLK+ IBP3_YGQPLPGYTTK+ IBP5_AVYLPNCDR+ ITIH2_IQPSGGTNINEALLR+ PEPC_AECGLGVPTTR+ PROC_TFVLNFIK+ PROS_NNLELSTPLK+ PTPRK_QNVVDVFHAVK+ RIDA_AAYQVAALPK+ SAA1_FFGHGAEDSLADQAANEWGR+ SAA4_GPGGVWAAK+ TFPI1_IAYEEIFVK+ (intercept)	22	10	6.86E-03
	2	A1BG_LLELTGPK+ ALDOC_ALQASALNAWR+ ALS_DFALQNPSAVPR+ AMPN_AQIINDAFNLASAHK+ ANAG_DFCGCHVAWSGSQLR+ AT1A1_IVEIPFNSTNK+ CATS_YTELPYGR+ CBG_HLVALSPK+ CBPB2_DTGTYGFLLPER+ COAA1_GTHVWVGLYK+ CRYM_TVVPVTK+ GPX3_FYTFLK+ IBP3_YGQPLPGYTTK+ IBP5_AVYLPNCDR+ ITIH2_IQPSGGTNINEALLR+ KGP1_EEEIQELK+ KPYM_IYVDDGLISLQVK+ PEPC_AECGLGVPTTR+ PROC_TFVLNFIK+ PROS_NNLELSTPLK+ PTPRK_QNVVDVFHAVK+ RIDA_AAYQVAALPK+ SAA1_FFGHGAEDSLADQAANEWGR+ SAA4_GPGGVWAAK+ SHBG_TSSSFEVR+ SUV3_LLNLEGFPSGSQSR+ TFPI1_IAYEEIFVK+ (Intercept)	28	13	1.56E-01
	3	A1BG_LLELTGPK+ ALDOC_ALQASALNAWR+ ALS_DFALQNPSAVPR+ AMPN_AQIINDAFNLASAHK+ ANAG_DFCGCHVAWSGSQLR+ ARMD4_TVVPSITR+ AT1A1_IVEIPFNSTNK+ CATS_YTELPYGR+ CBG_HLVALSPK+ CBPB2_DTGTYGFLLPER+ COAA1_GTHVWVGLYK+ CRYM_TVVPVTK+ GPX3_FYTFLK+ IBP3_YGQPLPGYTTK+ IBP5_AVYLPNCDR+ ITIH2_IQPSGGTNINEALLR+ KGP1_EEEIQELK+ KPYM_IYVDDGLISLQVK+ PEPC_AECGLGVPTTR+ PROC_TFVLNFIK+ PROS_NNLELSTPLK+ PTPRK_QNVVDVFHAVK+ RIDA_AAYQVAALPK+ SAA1_FFGHGAEDSLADQAANEWGR+ SAA4_GPGGVWAAK+ SHBG_TSSSFEVR+ SUV3_LLNLEGFPSGSQSR+ TFPI1_IAYEEIFVK+ TRPC6_LGILGSHEDLSK+ (intercept)	30	21	3.23E-01
	4	A1BG_LLELTGPK+ ALDOC_ALQASALNAWR+ ALS_DFALQNPSAVPR+ AMPN_AQIINDAFNLASAHK+ ANAG_DFCGCHVAWSGSQLR+ AT1A1_IVEIPFNSTNK+ CATS_YTELPYGR+ CBG_HLVALSPK+ CBPB2_DTGTYGFLLPER+ COAA1_GTHVWVGLYK+ CRYM_TVVPVTK+ GPX3_FYTFLK+ IBP3_YGQPLPGYTTK+ IBP5_AVYLPNCDR+ ITIH2_IQPSGGTNINEALLR+ PEPC_AECGLGVPTTR+ PROC_TFVLNFIK+ PROS_NNLELSTPLK+ PTPRK_QNVVDVFHAVK+ RIDA_AAYQVAALPK+ SAA1_FFGHGAEDSLADQAANEWGR+ SAA4_GPGGVWAAK+ SHBG_TSSSFEVR+ SUV3_LLNLEGFPSGSQSR+ TFPI1_IAYEEIFVK+ (intercept)	26	7	6.00E-02

5	ALDOC_ALQASALNAWR+ ALS_DFALQNPSAVPR+ AMPN_AQIINDAFNLASAHK+ ANAG_DFCGCHVAWSGSQLR+ AT1A1_IVEIPFNSTNK+ CATS_YTELPYGR+ CBG_HLVALSPK+ CBPB2_DTGTYGFLLPER+ COAA1_GTHVWVGLYK+ CRYM_TVVPVTK+ GPX3_FYTFLK+ IBP3_YGQPLPGYTTK+ IBP5_AVYLPNCDR+ ITIH2_IQPSGGTNINEALLR+ PEPC_AECGLGVPTTR+ PROC_TFVLNFIK+ PROS_NNLELSTPLK+ PTPRK_QNVVDVFHAVK+ RIDA_AAYQVAALPK+ SAA1_FFGHGAEDSLADQAANEWGR+ SAA4_GPGGVWAAK+ SHBG_TSSSFEVR+ SUV3_LLNLEGFPSGSQSR+ TFPI1_IAYEEIFVK+ (intercept)	25	16	7.24E-02
6	A1BG_LLELTGPK+ ALDOC_ALQASALNAWR+ ALS_DFALQNPSAVPR+ AMPN_AQIINDAFNLASAHK+ ANAG_DFCGCHVAWSGSQLR+ ARMD4_TVVPSITR+ AT1A1_IVEIPFNSTNK+ CATS_YTELPYGR+ CBG_HLVALSPK+ CBPB2_DTGTYGFLLPER+ COAA1_GTHVWVGLYK+ CRYM_TVVPVTK+ GPX3_FYTFLK+ IBP3_YGQPLPGYTTK+ IBP5_AVYLPNCDR+ ITIH2_IQPSGGTNINEALLR+ KGP1_EEEIQELK+ KPYM_IYVDDGLISLQVK+ PEPC_AECGLGVPTTR+ PROC_TFVLNFIK+ PROS_NNLELSTPLK+ PTPRK_QNVVDVFHAVK+ RIDA_AAYQVAALPK+ SAA1_FFGHGAEDSLADQAANEWGR+ SAA4_GPGGVWAAK+ SHBG_TSSSFEVR+ SUV3_LLNLEGFPSGSQSR+ TFPI1_IAYEEIFVK+ (intercept)	29	17	2.91E-01
7	A1BG_LLELTGPK+ ALDOC_ALQASALNAWR+ ALS_DFALQNPSAVPR+ AMPN_AQIINDAFNLASAHK+ ANAG_DFCGCHVAWSGSQLR+ AT1A1_IVEIPFNSTNK+ CATS_YTELPYGR+ CBG_HLVALSPK+ CBPB2_DTGTYGFLLPER+ COAA1_GTHVWVGLYK+ CRYM_TVVPVTK+ GPX3_FYTFLK+ IBP3_YGQPLPGYTTK+ IBP5_AVYLPNCDR+ ITIH2_IQPSGGTNINEALLR+ KGP1_EEEIQELK+ PEPC_AECGLGVPTTR+ PROC_TFVLNFIK+ PROS_NNLELSTPLK+ PTPRK_QNVVDVFHAVK+ RIDA_AAYQVAALPK+ SAA1_FFGHGAEDSLADQAANEWGR+ SAA4_GPGGVWAAK+ SHBG_TSSSFEVR+ SUV3_LLNLEGFPSGSQSR+ TFPI1_IAYEEIFVK+ (intercept)	27	10	8.74E-02
8	ALDOC_ALQASALNAWR+ ALS_DFALQNPSAVPR+ ANAG_DFCGCHVAWSGSQLR+ AT1A1_IVEIPFNSTNK+ CATS_YTELPYGR+ CBG_HLVALSPK+ CBPB2_DTGTYGFLLPER+ COAA1_GTHVWVGLYK+ CRYM_TVVPVTK+ GPX3_FYTFLK+ IBP3_YGQPLPGYTTK+ IBP5_AVYLPNCDR+ ITIH2_IQPSGGTNINEALLR+ PEPC_AECGLGVPTTR+ PROC_TFVLNFIK+ PROS_NNLELSTPLK+ PTPRK_QNVVDVFHAVK+ RIDA_AAYQVAALPK+ SAA1_FFGHGAEDSLADQAANEWGR+ SAA4_GPGGVWAAK+ SHBG_TSSSFEVR+ SUV3_LLNLEGFPSGSQSR+ TFPI1_IAYEEIFVK+ (intercept)	24	4	2.71E-03
9	ALDOC_ALQASALNAWR+ ALS_DFALQNPSAVPR+ ANAG_DFCGCHVAWSGSQLR+ AT1A1_IVEIPFNSTNK+ CATS_YTELPYGR+ CBG_HLVALSPK+ CBPB2_DTGTYGFLLPER+ COAA1_GTHVWVGLYK+ CRYM_TVVPVTK+ GPX3_FYTFLK+ IBP3_YGQPLPGYTTK+ IBP5_AVYLPNCDR+ ITIH2_IQPSGGTNINEALLR+ PEPC_AECGLGVPTTR+ PROC_TFVLNFIK+ PROS_NNLELSTPLK+ RIDA_AAYQVAALPK+ SAA1_FFGHGAEDSLADQAANEWGR+ SAA4_GPGGVWAAK+ TFPI1_IAYEEIFVK+ (intercept)	21	2	2.06E-06

BD vs SCZ	Model #	Unique model (combination of selected features)	Number of features	Frequency	Probability
	1	AACT_ALEQDLPVNIK+ AMPN_DFYNPVVPEAQK+ BPIB1_DTGTYGFLLPER+ C1RL_HLVALSPK+ CBG_GSEAINAPGDNPAK+ CBPB2_ALGFEAAESSLTK+ CFAB_LNTIGHYEISNGSTIK+ CLD3_YLSYTLNPDLIR+ CNDP1_DEELSCTVVELK+ DDR1_LTVQTK+ DOPO_INNTHALVSLLQNLNK+ G3BP1_AVYLPNCDR+ GPR37_IPLNDLFR+ HEP2_SLEAQGNSSHLDADTVR+ IBP5_INIPPQR+ IC1_IAYEEIFVK+ LAMP2_ISPDLPDTIYVLALTYDSAR+ MBL2_LLELTGPK+ NPC2_ASYLDCIR+ PLXC1_DISEVVTPR+ PSMD1_SLESINSR+ RL30_VSTAVLSITAK+ SPTB2_LHLVALVGTQGR+ SUV3_VLNLGPITR+ TFPI1_TTFDPK+ TRFE_LVVEWQLQDDK+ UROM_LLNLEGFPSGSQSR+ (intercept)	28	9	9.87E-02
	2	AACT_DEELSCTVVELK+AMPN_AQIINDAFNLASAHK+BPIB1_ALGFEAAESSLTK+C1RL_GSEAINAPGDNPAK+CBG_HLVALSPK+CBPB2_DTGTYGFLLPER+CFAB_DISEVVTPR+CLD3_DFYNPVVPEAQK+CNDP1_AIHLDLEEYR+DDR1_LHLVALVGTQGR+DOPO_VISTLEEPTPQCPTSQGR+G3BP1_INIPPQR+GPR37_ISPDLPDTIYVLALTYDSAR+HEP2_TLEAQLTPR+IBP5_AVYLPNCDR+IC1_TTFDPK+LAMP2_IPLNDLFR+MBL2_FQASVATPR+NPC2_LVVEWQLQDDK+PLXC1_LNTIGHYEISNGSTIK+PSMD1_VSTAVLSITAK+RL30_SLESINSR+SPTB2_LTVQTK+TFPI1_IAYEEIFVK+TRFE_ASYLDCIR+UROM_VLNLGPITR+(intercept)	27	17	3.17E-01
	3	AACT_DEELSCTVVELK+AMPN_AQIINDAFNLASAHK+BPIB1_ALGFEAAESSLTK+C1RL_GSEAINAPGDNPAK+CAD13_INNTHALVSLLQNLNK+CBG_HLVALSPK+CBPB2_DTGTYGFLLPER+CFAB_DISEVVTPR+CLD3_DFYNPVVPEAQK+CNDP1_AIHLDLEEYR+DDR1_LHLVALVGTQGR+DOPO_VISTLEEPTPQCPTSQGR+G3BP1_INIPPQR+GPR37_ISPDLPDTIYVLALTYDSAR+HEP2_TLEAQLTPR+IBP5_AVYLPNCDR+IC1_TTFDPK+LAMP2_IPLNDLFR+MBL2_FQASVATPR+NPC2_LVVEWQLQDDK+PLXC1_LNTIGHYEISNGSTIK+PSMD1_VSTAVLSITAK+RL30_SLESINSR+SPTB2_LTVQTK+SUV3_LLNLEGFPSGSQSR+TFPI1_IAYEEIFVK+TRFE_ASYLDCIR+UROM_VLNLGPITR+(intercept)	29	39	2.65E-01
	4	AACT_DEELSCTVVELK+AMPN_AQIINDAFNLASAHK+BPIB1_ALGFEAAESSLTK+C1RL_GSEAINAPGDNPAK+CBG_HLVALSPK+CBPB2_DTGTYGFLLPER+CFAB_DISEVVTPR+CLD3_DFYNPVVPEAQK+CNDP1_AIHLDLEEYR+DDR1_LHLVALVGTQGR+DOPO_VISTLEEPTPQCPTSQGR+G3BP1_INIPPQR+GPR37_ISPDLPDTIYVLALTYDSAR+HEP2_TLEAQLTPR+IBP5_AVYLPNCDR+IC1_TTFDPK+LAMP2_IPLNDLFR+MBL2_FQASVATPR+NPC2_LVVEWQLQDDK+PLXC1_LNTIGHYEISNGSTIK+PSMD1_VSTAVLSITAK+SPTB2_LTVQTK+TFPI1_IAYEEIFVK+TRFE_ASYLDCIR+UROM_VLNLGPITR+(intercept)	26	27	3.07E-01

5	AACT_DEELSCTVVELK+AMPN_AQIINDAFNLASAHK+BPIB1_ALGFEAAESSLTK+C1RL_GSEAINAPGDNPAK+CAD13_INNTHALVSLLQNLNK+CBG_HLVALSPK+CBPB2_DTGTYGFLLPER+CFAB_DISEVVTPR+CLD3_DFYNPVVPEAQK+CNDP1_AIHLDLEEYR+DDR1_LHLVALVGTQGR+DOPO_VISTLEEPTPQCPTSQGR+G3BP1_INIPPQR+GPR37_ISPDLPDTIYVLALTYDSAR+HEP2_TLEAQLTPR+IBP5_AVYLPNCDR+IC1_TTFDPK+LAMP2_IPLNDLFR+MBL2_FQASVATPR+NPC2_LVVEWQLQDDK+PLXC1_LNTIGHYEISNGSTIK+PSMD1_VSTAVLSITAK+RL30_SLESINSR+SPTB2_LTVQTK+STAB1_SLEAQGNSSHLDADTVR+SUV3_LLNLEGFPSGSQSR+TFPI1_IAYEEIFVK+TRFE_ASYLDCIR+UROM_VLNLGPITR+(intercept)	30	5	1.30E-02
6	AACT_DEELSCTVVELK+AMPN_AQIINDAFNLASAHK+BPIB1_ALGFEAAESSLTK+C1RL_GSEAINAPGDNPAK+CBPB2_DTGTYGFLLPER+CFAB_DISEVVTPR+CLD3_DFYNPVVPEAQK+DOPO_VISTLEEPTPQCPTSQGR+G3BP1_INIPPQR+GPR37_ISPDLPDTIYVLALTYDSAR+HEP2_TLEAQLTPR+IBP5_AVYLPNCDR+IC1_TTFDPK+LAMP2_IPLNDLFR+MBL2_FQASVATPR+NPC2_LVVEWQLQDDK+PLXC1_LNTIGHYEISNGSTIK+PSMD1_VSTAVLSITAK+TFPI1_IAYEEIFVK+TRFE_ASYLDCIR+UROM_VLNLGPITR+(intercept)	22	2	2.88E-09
7	AACT_DEELSCTVVELK+AMPN_AQIINDAFNLASAHK+BPIB1_ALGFEAAESSLTK+C1RL_GSEAINAPGDNPAK+CFAB_DISEVVTPR+CLD3_DFYNPVVPEAQK+DOPO_VISTLEEPTPQCPTSQGR+GPR37_ISPDLPDTIYVLALTYDSAR+HEP2_TLEAQLTPR+IBP5_AVYLPNCDR+IC1_TTFDPK+MBL2_FQASVATPR+NPC2_LVVEWQLQDDK+PLXC1_LNTIGHYEISNGSTIK+PSMD1_VSTAVLSITAK+TFPI1_IAYEEIFVK+UROM_VLNLGPITR+(intercept)	18	1	2.81E-11

Protein_peptide sequence of the selected features is listed for each combination.
Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia

Table 2-7. ANCOVA analysis of proteomic features and significant clinical variables from the final multiprotein marker models in the training set

MPM model	Proteomic features	Covariates	*P*-value[a] (ANCOVA)
MDD vs BD	ALDOC_ALQASALNAWR	Group (MDD vs BD)	0.121
		SCL_PAR	**0.011**
	ALDOC_ALQASALNAWR	Group (MDD vs BD)	0.080
		SCL_ANG	**0.049**
	ARMD4_TVVPSITR	Group (MDD vs BD)	0.073
		SCL_OCD	**0.022**
	CTND1_GYELLFQPEVVR	Group (MDD vs BD)	0.053
		BPRS	**0.036**
MDD vs SCZ	ALDOC_ALQASALNAWR	Group (MDD vs SCZ)	0.124
		SCL_PSY	**0.015**
	ALDOC_ALQASALNAWR	Group (MDD vs SCZ)	0.112
		SCL_PAR	**0.025**
	ALDOC_ALQASALNAWR	Group (MDD vs SCZ)	0.267
		SCL_ANG	**0.011**
	IBP3_YGQPLPGYTTK	Group (MDD vs SCZ)	0.173
		SCL_OCD	**0.008**
BD vs SCZ	GPR37_ ISPDLPDTIYVLALTYDSAR	Group (BD vs SCZ)	0.113
		BPRS	**0.024**
	UROM_VLNLGPITR	Group (BD vs SCZ)	0.141
		BPRS	**0.039**

[a]*P*-value < 0.05 is considered statistically significant, denoted by bold font
Protein_peptide sequence is listed for each feature.
Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia; SCL = Symptom Checklist-90-Revised, PAR = paranoid ideation dimension, ANG = hostility dimension, OCD = obsessive-compulsive dimension, BPRS = Brief Psychiatric Rating Scale, PSY = psychoticism dimension

Table 2-8. Mass spectra information of the proteomic features from the final multiprotein marker models

MDD vs BD				Mass Information								
ID[a]	Protein	Peptide	Gene name	Precursor ion. Light (m/z)	Precursor ion. Heavy (m/z)	Precursor ion charge	Product ion. Light (m/z)	Product ion. Heavy (m/z)	Product ion charge	Product ion type	Collision energy (volt)	Retention time (RT)
P09972	ALDOC	ALQASALNAWR	ALDOC	400.89	404.22	3	508.77	513.77	2	y9	9.60	37.47+/-0.49
P15144	AMPN	AQIINDAFNLASAHK	ANPEP	734.40	739.40	2	277.15	277.15	1	b2	23.80	39.25+/-0.49
Q86TY3	ARMD4	TVVPSITR	ARMH4	436.76	441.77	2	201.12	201.12	1	b2	14.50	28.73+/-0.19
Q9NZP8	C1RL	GSEAINAPGDNPAK	C1RL	670.83	674.83	2	698.35	706.36	1	y7	21.80	24.58+/-0.14
P55290	CAD13	INNTHALVSLLQNLNK	CDH13	598.01	600.68	3	729.43	737.44	1	y6	16.70	44.85+/-0.36
P11597	CETP	ASYPDITGEK	CETP	540.76	544.77	2	759.39	767.40	1	y7	17.80	27.87+/-0.17
Q03692	COAA1	GTHVWVGLYK	COL10A1	387.21	389.88	3	480.28	488.30	1	y4	9.10	33.4+/-0.52
O60716	CTND1	GYELLFQPEVVR	CTNND1	725.39	730.39	2	599.35	609.36	1	y5	23.50	44.94+/-0.5
Q08345	DDR1	LHLVALVGTQGR	DDR1	632.38	637.38	2	1013.61	1023.62	1	y10	20.60	34.94+/-0.25
P09172	DOPO	TPEGLTLLFK	DBH	559.83	563.83	2	509.30	513.31	2	y9	18.40	44.83+/-0.31
P05155	IC1	TTFDPK	SERPING1	354.68	358.69	2	244.17	252.18	1	y2	12.00	22.55+/-0.76
Q9NPH3	IL1AP	NEVWWTIDGK	IL1RAP	624.31	628.31	2	905.45	913.47	1	y7	20.40	40.99+/-0.33
P19823	ITIH2	IQPSGGTNINEALLR	ITIH2	791.93	796.94	2	671.36	676.36	2	y13	25.50	35.16+/-0.23
P61916	NPC2	LVVEWQLQDDK	NPC2	458.24	460.91	3	505.23	513.24	1	y4	11.70	38.92+/-0.52
P62826	RAN	FNVWDTAGQEK	RAN	647.81	651.81	2	633.32	641.33	1	y6	21.10	35.37+/-0.41
P0DJI8	SAA1	FFGHGAEDSLADQAANEWGR	SAA1	726.66	730.00	3	732.34	742.35	1	y6	21.40	39.78+/-0.29
P02787	TRFE	ASYLDCIR	TF	499.24	504.25	2	563.26	573.27	1	y4	16.50	32.04+/-0.2

MDD vs SCZ				Mass Information								
ID[a]	Protein	Peptide	Gene name	Precursor ion. Light (m/z)	Precursor ion. Heavy (m/z)	Precursor ion charge	Product ion. Light (m/z)	Product ion. Heavy (m/z)	Product ion charge	Product ion type	Collision energy (volt)	Retention time (RT)
P09972	ALDOC	ALQASALNAWR	ALDOC	400.89	404.22	3	508.77	513.77	2	y9	9.60	37.47+/-0.49
P35858	ALS	DFALQNPSAVPR	IGFALS	657.84	662.85	2	626.36	636.37	1	y6	21.40	34.77+/-0.45
P54802	ANAG	DFCGCHVAWSGSQLR	NAGLU	593.93	597.26	3	759.34	764.34	2	y13	16.60	34.99+/-0.28
P05023	AT1A1	IVEIPFNSTNK	ATP1A1	631.34	635.35	2	342.20	342.20	1	b3	20.60	35.32+/-0.27
P25774	CATS	YTELPYGR	CTSS	499.75	504.75	2	734.38	744.39	1	y6	16.50	28.78+/-0.18
P08185	CBG	HLVALSPK	SERPINA6	432.77	436.78	2	251.15	251.15	1	b2	14.40	25.22+/-0.23
Q96IY4	CBPB2	DTGTYGFLLPER	CPB2	456.90	460.23	3	401.21	411.22	1	y3	11.60	41.48+/-0.35
Q03692	COAA1	GTHVWVGLYK	COL10A1	387.21	389.88	3	480.28	488.30	1	y4	9.10	33.4+/-0.52
Q14894	CRYM	TVVPVTK	CRYM	372.24	376.24	2	201.12	201.12	1	b2	12.50	24.77+/-0.28
P22352	GPX3	FYTFLK	GPX3	409.73	413.73	2	508.31	516.33	1	y4	13.70	37.78+/-0.24
P17936	IBP3	YGQPLPGYTTK	IGFBP3	612.82	616.82	2	876.48	884.50	1	y8	20.00	29.26+/-0.19
P24593	IBP5	AVYLPNCDR	IGFBP5	554.27	559.27	2	661.27	671.28	1	y5	18.20	28.3+/-0.47
P19823	ITIH2	IQPSGGTNINEALLR	ITIH2	791.93	796.94	2	671.36	676.36	2	y13	25.50	35.16+/-0.23
P20142	PEPC	AECGLGVPTTR	PGC	580.79	585.79	2	630.36	640.37	1	y6	19.00	27.96+/-0.18
P04070	PROC	TFVLNFIK	PROC	491.29	495.30	2	733.46	741.47	1	y6	16.20	45.02+/-0.3
P07225	PROS	NNLELSTPLK	PROS1	564.82	568.82	2	787.46	795.47	1	y7	18.50	34.02+/-0.59
P52758	RIDA	AAYQVAALPK	RIDA	516.30	520.30	2	428.29	436.30	1	y4	17.00	31.05+/-0.43
P0DJI8	SAA1	FFGHGAEDSLADQAANEWGR	SAA1	726.66	730.00	3	732.34	742.35	1	y6	21.40	39.78+/-0.29
P35542	SAA4	GPGGVWAAK	SAA4	421.73	425.74	2	688.38	696.39	1	y7	14.10	26.25+/-0.19
P10646	TFPI1	IAYEEIFVK	TFPI	556.31	560.31	2	506.33	514.35	1	y4	18.20	39.75+/-0.55

BD vs SCZ				Mass Information								
ID[a]	Protein	Peptide	Gene name	Precursor ion. Light (m/z)	Precursor ion. Heavy (m/z)	Precursor ion charge	Product ion. Light (m/z)	Product ion. Heavy (m/z)	Product ion charge	Product ion type	Collision energy (volt)	Retention time (RT)
P01011	AACT	DEELSCTVVELK	SERPINA3	474.57	477.24	3	488.31	496.32	1	y4	12.30	36.4+/-0.28
P15144	AMPN	AQIINDAFNLASAHK	ANPEP	734.40	739.40	2	277.15	277.15	1	b2	23.80	42.82+/-0.3
Q8TDL5	BPIB1	ALGFEAAESSLTK	BPIFB1	662.34	666.35	2	735.39	743.40	1	y7	21.50	37.2+/-0.24
Q9NZP8	C1RL	GSEAINAPGDNPAK	C1RL	670.83	674.83	2	698.35	706.36	1	y7	21.80	24.58+/-0.14
P00751	CFAB	DISEVVTPR	CFB	508.27	513.28	2	787.43	797.44	1	y7	16.80	30.12+/-0.2
O15551	CLD3	DFYNPVVPEAQK	CLDN3	703.85	707.86	2	572.30	580.32	1	y5	22.80	35.45+/-0.24
P09172	DOPO	VISTLEEPTPQCPTSQGR	DBH	559.83	563.83	2	509.30	513.31	2	y9	18.40	44.83+/-0.31
O15354	GPR37	ISPDLPDTIYVLALTYDSAR	GPR37	741.72	745.06	3	825.41	835.42	1	y7	21.90	52.74+/-0.15
P05546	HEP2	TLEAQLTPR	SERPIND1	514.79	519.79	2	814.44	824.45	1	y7	17.00	29.55+/-0.2
P24593	IBP5	AVYLPNCDR	IGFBP5	554.27	559.27	2	661.27	671.28	1	y5	18.20	28.3+/-0.47
P05155	IC1	TTFDPK	SERPING1	354.68	358.69	2	244.17	252.18	1	y2	12.00	22.55+/-0.76
P11226	MBL2	FQASVATPR	MBL2	488.76	493.77	2	701.39	711.40	1	y7	16.20	26.64+/-0.17
P61916	NPC2	LVVEWQLQDDK	NPC2	458.24	460.91	3	505.23	513.24	1	y4	11.70	38.92+/-0.52
O60486	PLXC1	LNTIGHYEISNGSTIK	PLXNC1	582.97	585.64	3	653.33	657.33	2	y12	16.20	33.21+/-0.51
Q99460	PSMD1	VSTAVLSITAK	PSMD1	545.33	549.34	2	731.47	739.48	1	y7	17.90	33.62+/-0.51
P10646	TFPI1	IAYEEIFVK	TFPI	556.31	560.31	2	506.33	514.35	1	y4	18.20	39.75+/-0.55
P07911	UROM	VLNLGPITR	UMOD	491.81	496.81	2	770.45	780.46	1	y7	16.20	35.59+/-0.25

[a]UniProt accession number
Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia

Table 2-9. Expression levels of proteomic features from the final multiprotein marker models in the total study population

MDD vs BD	ID[a]	Protein[b]	Peptide	Gene name	MDD vs BD vs HC		MDD vs BD			MDD vs HC			BD vs HC		
					F-statistics	*P*-value[c]	Fold-change[d]	Upregulated	*P*-value[c]	Fold-change[d]	Upregulated	*P*-value[c]	Fold-change[d]	Upregulated	*P*-value[c]
	P09972	ALDOC	ALQASALNAWR	ALDOC	4.286	**0.014**	-0.527	MDD	**0.043**	0.582	MDD	**0.024**	0.054	BD	0.968
	P15144	AMPN	AQIINDAFNLASAHK	ANPEP	3.597	**0.028**	-0.097	MDD	0.079	0.111	MDD	**0.040**	0.014	BD	0.949
	Q86TY3	ARMD4	TVVPSITR	ARMH4	4.882	**0.008**	0.225	BD	**0.011**	-0.192	HC	**0.041**	0.033	BD	0.907
	Q9NZP8	C1RL	GSEAINAPGDNPAK	C1RL	8.097	**< 0.001**	0.107	BD	**0.015**	0.045	MDD	0.474	0.153	BD	**< 0.001**
	P55290	CAD13	INNTHALVSLLQNLNK	CDH13	1.992	0.137	0.166	BD	0.115	-0.092	HC	0.525	0.075	BD	0.657
	P11597	CETP	ASYPDITGEK	CETP	6.723	**0.001**	-0.185	MDD	**0.002**	0.028	MDD	0.871	-0.158	HC	**0.013**
	Q03692	COAA1	GTHVWVGLYK	COL10A1	1.769	0.172	-0.347	MDD	0.146	0.156	MDD	0.682	-0.190	HC	0.571
	O60716	CTND1	GYELLFQPEVVR	CTNND1	5.810	**0.003**	0.285	BD	**0.006**	-0.019	HC	0.979	0.266	BD	**0.014**
	Q08345	DDR1	LHLVALVGTQGR	DDR1	3.520	**0.030**	-0.316	MDD	**0.025**	0.107	MDD	0.658	-0.209	HC	0.210
	P09172	DOPO	TPEGLTLLFK	DBH	3.385	**0.035**	0.148	BD	0.246	-0.241	HC	**0.028**	-0.093	HC	0.584
	P05155	IC1	TTFDPK	SERPING1	3.679	**0.026**	0.189	BD	**0.020**	-0.112	HC	0.255	0.076	BD	0.535
	Q9NPH3	IL1AP	NEVWWTIDGK	IL1RAP	2.521	0.081	0.119	BD	0.090	-0.020	HC	0.938	0.099	BD	0.198
	P19823	ITIH2	IQPSGGTNINEALLR	ITIH2	6.930	**0.001**	-0.133	MDD	**0.001**	0.038	MDD	0.559	-0.094	HC	**0.032**
	P61916	NPC2	LVVEWQLQDDK	NPC2	1.970	0.141	0.257	BD	0.117	-0.121	HC	0.627	0.136	BD	0.560
	P62826	RAN	FNVWDTAGQEK	RAN	1.519	0.220	-0.185	MDD	0.197	0.067	MDD	0.811	-0.118	HC	0.530

	P0DJI8	SAA1	FFGHGAEDSLADQAANEWGR	SAA1	1.719	0.180	0.249	BD	0.159	-0.091	HC	0.786	0.158	BD	0.490
	P02787	TRFE	ASYLDCIR	TF	2.993	0.051	0.302	BD	0.052	-0.093	HC	0.749	0.209	BD	0.237

MDD vs SCZ	ID[a]	Protein[b]	Peptide	Gene name	MDD vs SCZ vs HC		MDD vs SCZ			MDD vs HC			SCZ vs HC		
					F-statistics	*P*-value[c]	Fold-change[d]	Upregulated	*P*-value[c]	Fold-change[d]	Upregulated	*P*-value[c]	Fold-change[d]	Upregulated	*P*-value[c]
	P09972	ALDOC	ALQASALNAWR	ALDOC	4.981	**0.007**	-0.530	MDD	**0.001**	0.582	MDD	0.983	0.052	SCZ	**0.003**
	P35858	ALS	DFALQNPSAVPR	IGFALS	5.596	**0.004**	-0.133	MDD	0.272	-0.011	HC	0.533	-0.144	HC	**0.028**
	P54802	ANAG	DFCGCHVAWSGSQLR	NAGLU	2.317	0.100	0.237	SCZ	0.119	-0.027	HC	0.940	0.210	SCZ	0.065
	P05023	AT1A1	IVEIPFNSTNK	ATP1A1	7.890	**< 0.001**	-0.234	MDD	**< 0.001**	0.012	MDD	0.792	-0.222	HC	**0.004**
	P25774	CATS	YTELPYGR	CTSS	14.931	**< 0.001**	0.222	SCZ	**0.004**	0.161	MDD	0.058	0.383	SCZ	**< 0.001**
	P08185	CBG	HLVALSPK	SERPINA6	7.439	**0.001**	-0.155	MDD	**< 0.001**	-0.010	HC	0.105	-0.165	HC	**< 0.001**
	Q96IY4	CBPB2	DTGTYGFLLPER	CPB2	3.367	**0.035**	0.070	SCZ	**0.002**	0.049	MDD	0.747	0.119	SCZ	**0.024**
	Q03692	COAA1	GTHVWVGLYK	COL10A1	1.941	0.145	-0.332	MDD	0.121	0.156	MDD	0.870	-0.176	HC	0.564
	Q14894	CRYM	TVVPVTK	CRYM	11.382	**< 0.001**	-0.233	MDD	**0.009**	0.008	MDD	0.958	-0.226	HC	**0.005**
	P22352	GPX3	FYTFLK	GPX3	18.516	**< 0.001**	-0.145	MDD	**0.006**	-0.075	HC	0.260	-0.220	HC	0.314
	P17936	IBP3	YGQPLPGYTTK	IGFBP3	4.749	**0.009**	-0.149	MDD	**0.006**	0.077	MDD	0.260	-0.072	HC	0.314
	P24593	IBP5	AVYLPNCDR	IGFBP5	6.441	**0.002**	-0.292	MDD	**0.004**	0.063	MDD	0.058	-0.229	HC	**< 0.001**
	P19823	ITIH2	IQPSGGTNINEALLR	ITIH2	5.056	**0.007**	-0.115	MDD	**0.025**	0.038	MDD	**0.014**	-0.076	HC	0.966

	P20142	PEPC	AECGLGVPTTR	PGC	4.047	**0.018**	-0.141	MDD	**0.024**	0.018	MDD	0.940	-0.123	HC	0.065
	P04070	PROC	TFVLNFIK	PROC	9.017	**< 0.001**	-0.206	MDD	**< 0.001**	0.027	MDD	0.989	-0.179	HC	**< 0.001**
	P07225	PROS	NNLELSTPLK	PROS1	4.977	**0.007**	0.291	SCZ	**0.002**	0.075	MDD	0.977	0.366	SCZ	**0.002**
	P52758	RIDA	AAYQVAALPK	RIDA	2.976	0.052	0.078	SCZ	0.063	-0.067	HC	0.970	0.011	SCZ	0.948
	P0DJI8	SAA1	FFGHGAEDSLADQAANEWGR	SAA1	8.692	**< 0.001**	0.539	SCZ	**0.044**	-0.091	HC	0.813	0.447	SCZ	**0.009**
	P35542	SAA4	GPGGVWAAK	SAA4	2.511	0.082	0.098	SCZ	0.119	0.011	MDD	0.973	0.110	SCZ	0.200
	P10646	TFPI1	IAYEEIFVK	TFPI	6.287	**0.002**	-0.260	MDD	**0.005**	-0.025	HC	0.562	-0.285	HC	0.103

BD vs SCZ	**ID[a]**	**Protein[b]**	**Peptide**	**Gene name**	**BD vs SCZ vs HC**		**BD vs SCZ**			**BD vs HC**			**SCZ vs HC**		
					F-statistics	***P*-value[c]**	**Fold-change[d]**	**Upregulated**	***P*-value[c]**	**Fold-change[d]**	**Upregulated**	***P*-value[c]**	**Fold-change[d]**	**Upregulated**	***P*-value[c]**
	P01011	AACT	DEELSCTVVELK	SERPINA3	3.695	**0.026**	0.274	SCZ	**0.021**	-0.091	HC	0.658	0.183	SCZ	0.186
	P15144	AMPN	AQIINDAFNLASAHK	ANPEP	4.238	**0.015**	0.105	SCZ	**0.048**	0.014	BD	0.948	0.119	SCZ	**0.023**
	Q8TDL5	BPIB1	ALGFEAAESSLTK	BPIFB1	3.677	**0.026**	0.246	SCZ	0.061	0.019	BD	0.984	0.265	SCZ	**0.043**
	Q9NZP8	C1RL	GSEAINAPGDNPAK	C1RL	9.289	**< 0.001**	-0.133	BD	**0.002**	0.153	BD	**< 0.001**	0.020	SCZ	0.868
	P00751	CFAB	DISEVVTPR	CFB	6.793	**0.001**	0.164	SCZ	**0.004**	-0.005	HC	0.995	0.159	SCZ	**0.006**
	O15551	CLD3	DFYNPVVPEAQK	CLDN3	8.529	**< 0.001**	-0.217	BD	**0.008**	-0.075	HC	0.571	-0.292	HC	**< 0.001**
	P09172	DOPO	VISTLEEPTPQCPTSQGR	DBH	8.002	**< 0.001**	-0.260	BD	**0.012**	-0.093	HC	0.570	-0.353	HC	**< 0.001**
	O15354	GPR37	ISPDLPDTIYVLALTYDSAR	GPR37	5.911	**0.003**	-0.557	BD	**0.002**	0.284	BD	0.196	-0.272	HC	0.223
	P05546	HEP2	TLEAQLTPR	SERPIND1	4.275	**0.014**	0.144	SCZ	**0.016**	-0.030	HC	0.834	0.114	SCZ	0.079

P24593	IBP5	AVYLPNCDR	IGFBP5	11.290	**< 0.001**	-0.395	BD	**< 0.001**	0.166	BD	0.125	-0.229	HC	**0.020**
P05155	IC1	TTFDPK	SERPING1	6.684	**0.001**	-0.241	BD	**0.001**	0.076	BD	0.505	-0.165	HC	**0.043**
P11226	MBL2	FQASVATPR	MBL2	4.366	**0.013**	-0.326	BD	**0.015**	0.262	BD	0.070	-0.064	HC	0.853
P61916	NPC2	LVVEWQLQDDK	NPC2	2.351	0.096	-0.287	BD	0.078	0.136	BD	0.571	-0.151	HC	0.500
O60486	PLXC1	LNTIGHYEISNGSTIK	PLXNC1	1.580	0.207	0.275	SCZ	0.178	-0.132	HC	0.679	0.143	SCZ	0.634
Q99460	PSMD1	VSTAVLSITAK	PSMD1	5.620	**0.004**	0.105	SCZ	**0.014**	-0.113	HC	**0.009**	-0.008	HC	0.975
P10646	TFPI1	IAYEEIFVK	TFPI	6.455	**0.002**	-0.289	BD	**0.005**	0.004	BD	0.999	-0.285	HC	**0.007**
P07911	UROM	VLNLGPITR	UMOD	5.754	**0.003**	-0.258	BD	**0.017**	-0.035	HC	0.929	-0.292	HC	**0.006**

[a] UniProt accession number
[b] Overlapping proteomic features between MPM models are presented as different colors; blue for MDD versus BD and MDD versus SCZ, green for MDD versus BD and BD versus SCZ, and orange for MDD versus SCZ and BD versus SCZ.
[c] Bold font denotes statistical significance at *P*-value < 0.05. ANOVA for 3 groups comparison, Tukey's HSD for post-hoc analysis
[d] Fold change calculated with logarithmic$_{(2)}$ transformation
Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, HC = healthy controls

Table 2-10. Clinical features used in symptom checklist-based models and clinician rater score-based models

SCLB models

MDD vs BD Clinical variable	MDD vs SCZ Clinical variable	BD vs SCZ Clinical variable
SCL-90-R_SOM	SCL-90-R_SOM	SCL-90-R_OCD
SCL-90-R_IPS	SCL-90-R_IPS	SCL-90-R_IPS
SCL-90-R_PSY	SCL-90-R_PSY	SCL-90-R_PSY
SCL-90-R_60	SCL-90-R_PAR	SCL-90-R_PHO
SCL-90-R_DEP	SCL-90-R_DEP	SCL-90-R_DEP
		SCL-90-R_ANG

CRSB models

MDD vs BD Clinical variable	MDD vs SCZ Clinical variable	BD vs SCZ Clinical variable
BPRS	BPRS	BPRS
YMRS	YMRS	YMRS
MADRS	MADRS	MADRS
HAM-A	HAM-A	HAM-A

SCLB models constructed with the highest differentiation performance combination by binary logistic regression. CRSB models constructed with the total score of the scales by binary logistic regression.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, SCLB = symptom checklist-based, SCL-90-R = Symptom Checklist-90-Revised, SOM = somatization dimension, OCD = obsessive-compulsive dimension, IPS = interpersonal sensitivity dimension, PSY = psychoticism dimension, 60 = overeating item, PAR = paranoid ideation dimension, PHO = phobic anxiety dimension, DEP = depression dimension, ANG = hostility dimension, CRSB = clinician rater score-based, BPRS = Brief Psychiatric Rating Scale, YMRS = Young Mania Rating Scale, MADRS = Montgomery-Asberg Depression Rating Scale, HAM-A = Hamilton Anxiety Scale

Table 2-11. Predicted networks generated by ingenuity pathway analysis with proteomic features from the final multiprotein marker models

ID	Molecules in Network	Total molecules	Network Score	Focus molecules
1	**Akt, ALPL, ATP1A1, BPIFB1, C1RL, CDH13, CFB, CG, CTNND1, DBH, DDR1, ERK1/2, FSH, Growth hormone, Histone h3, IGFBP3, IGFBP5, IL1RAP, Lh, MAP4K4, MBL2, NFkB (complex), NPC2, P38, MAPK, PLIN3, PROC, PROS1, PSMD1, PTGS1, SAA1, SERPINA3, TCF, TCF21, TF, TYRO3**	**35**	**41**	**19**
2	**ANPEP, APLP2, APOB, CETP, CLDN3, COL10A1, Collagen type II, CPB2, CTSS, EHF, ETV5, F11, GLI1, GLI2, GPR37, HNF1A, HSPA1A/HSPA1B, IL1RN, ITIH2, KLK2, KLK6, LDL, LRP1, MMP2, MSI2, NOS3, NR5A2, PCSK9, PLIN2, SAA4, SERPIND1, SERPING1, STUB1, TFPI, VTN**	**35**	**23**	**12**
3	ALDOC, APC, BMP2, CCL3, CDK5, CTNNB1, CXCL1, EPAS1, ESR2, GPX3, GSK3B, HNRNPA1, IL13, IL22, IPO5, IRS1, JINK1/2, miR-483-3p (miRNAs w/seed CACUCCU), MTOR, MYB, MYOC, OGA, PARP, PARP1, PDIA3, PIK3R1, PKN1, PLXNC1, PPARG, RAN, RASSF1, RPTOR, STK11, SUZ12, TERT	35	6	4
4	PADI2, SERPINA6	2	2	1
5	CRYM, KDM1A, SBDS	3	2	1
6	AR, Hedgehog, PGC, SFTPB	4	2	1
7	GABARAP, GABARAPL1, GABARAPL2, NAGLU, TFEB	5	2	1

The networks satisfying network score≥ 20 are denoted by bold font.

Table 2-12. Proteomic profiling data of the differentially expressed proteins which overlapped with the proteomic features from the final multiprotein marker models

MDD vs BD

Protein	Gene name	ANOVA significance (MDD vs BD vs HC)	Clusters	Post-hoc analysis significance (MDD vs BD)	Fold-change[a]	Upregulated	*P*-value[b]	Consistency of statistical significance and expression pattern[c]
ALDOC	ALDOC	+	cluster 4		0.46	BD	0.727	N
ITIH2	**ITIH2**	**+**	**cluster 1**	**+**	**-0.19**	**MDD**	**0.001**	**Y**
TRFE	TF	+	cluster 2	+	0.26	BD	0.002	N
SAA1	SAA1	+	cluster 4	+	-1.03	MDD	0.004	N

MDD vs SCZ

Protein	Gene name	ANOVA significance (MDD vs SCZ vs HC)	Clusters	Post-hoc analysis significance (MDD vs SCZ)	Fold-change[a]	Upregulated	*P*-value[b]	Consistency of statistical significance and expression pattern[c]
ITIH2	**ITIH2**	**+**	**cluster 2**	**+**	**-0.13**	**MDD**	**0.003**	**Y**
ALDOC	ALDOC	+	cluster 4		0.05	SCZ	0.993	N
PROS	PROS1	+	cluster 1	+	-0.14	MDD	0.006	N
CBG	SERPINA6	+	cluster 2		-0.59	MDD	0.236	N
TFPI1	**TFPI**	**+**	**cluster 1**	**+**	**-0.86**	**MDD**	**0.043**	**Y**
SAA1	SAA1	+	cluster 4	+	-0.23	MDD	0.024	N

BD vs SCZ

Protein	Gene name	ANOVA significance (BD vs SCZ vs HC)	Clusters	Post-hoc analysis significance (BD vs SCZ)	Fold-change[a]	Upregulated	*P*-value[b]	Consistency of statistical significance and expression pattern[c]
C1RL	**C1RL**	**+**	**cluster 1**	**+**	**-0.21**	**BD**	**0.003**	**Y**

[a]Fold change calculated with logarithmic$_{(2)}$ transformation
[b]*P*-value < 0.05 is considered statistically significant, based on Tukey's HSD
[c]Consistency of statistical significance and expression pattern in both proteomic platforms
Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, HC = healthy controls

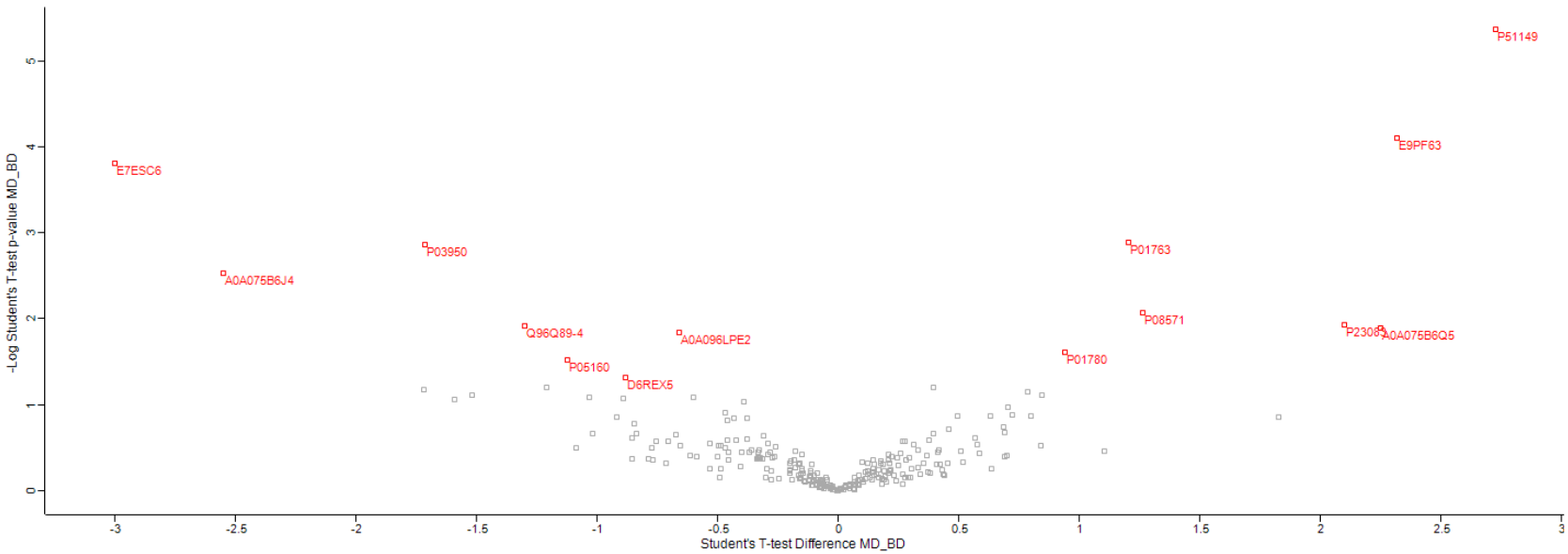


Figure 1-1. Volcano plot of differentially expressed proteins between drug free major depressive disorder and bipolar disorder. Representative protein IDs are statistically significant DEPs based on a *P*-value < 0.05 (red-color).

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, DEP = differentially expressed protein,

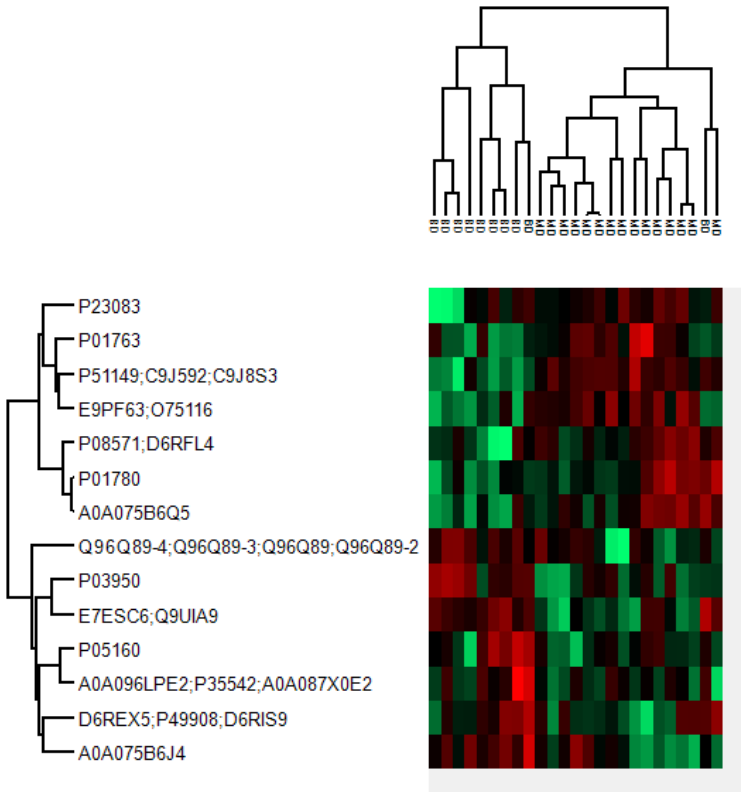


Figure 1-2. Hierarchical clustering of differentially expressed proteins between drug free major depressive disorder and bipolar disorder

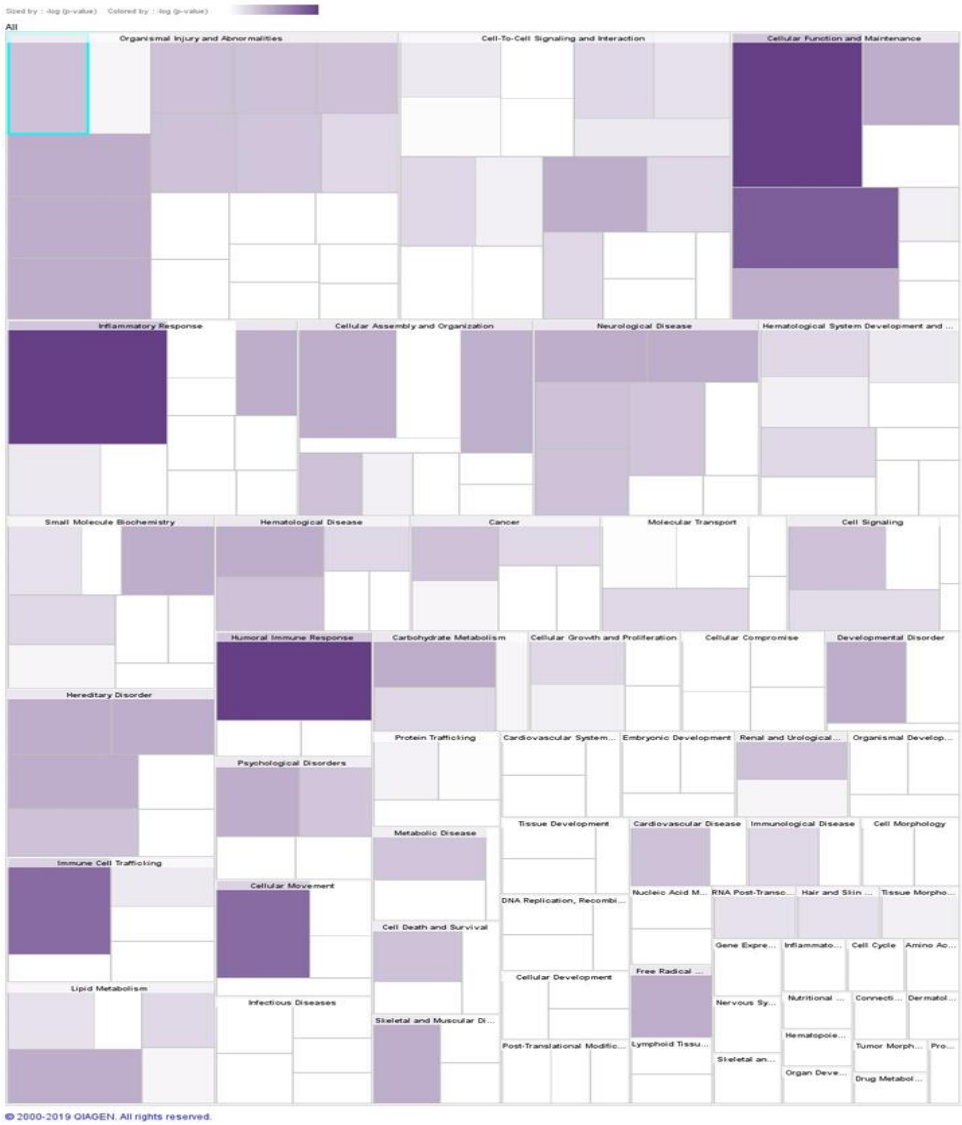


Figure 1-3. Tree-map of the diseases and functions associated with the differentially expressed proteins between drug free major depressive disorder and bipolar disorder. Major boxes represent categories of biological diseases/functions and individual rectangles represent an individual biological disease/function. The rectangle size correlates with increasing overlap significance and darker colors represent lower *P*-values.

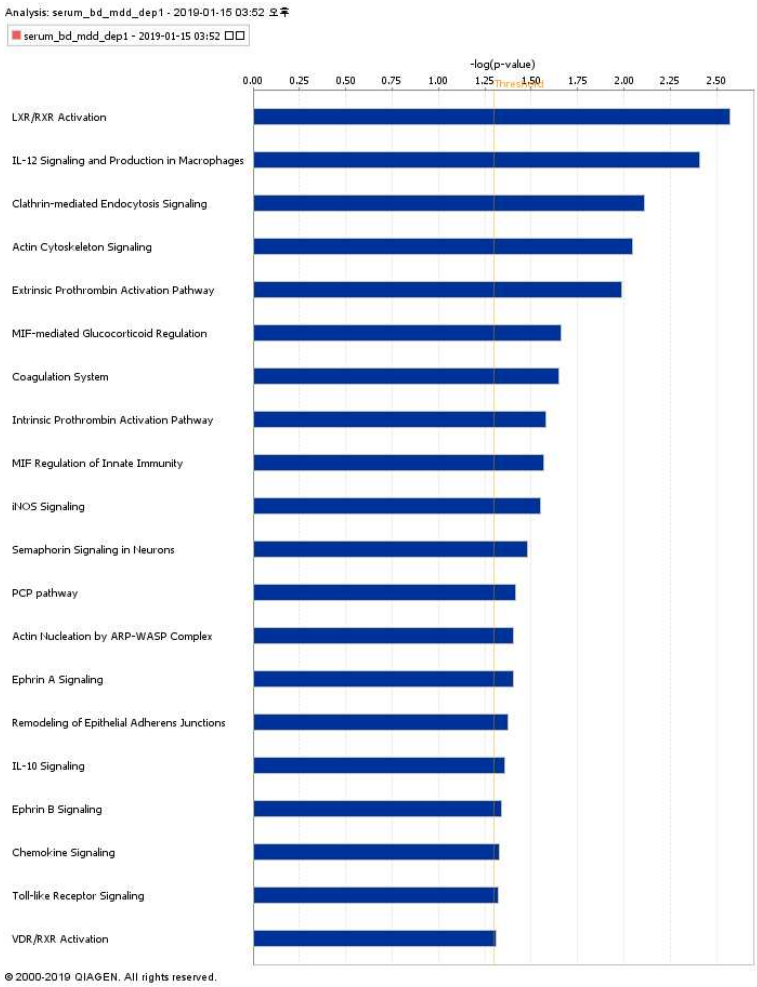
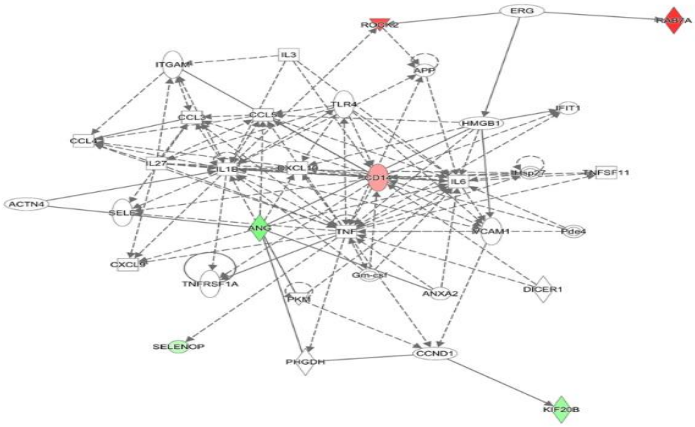


Figure 1-4. Canonical pathway analysis of differentially expressed proteins between drug free major depressive disorder and bipolar disorder



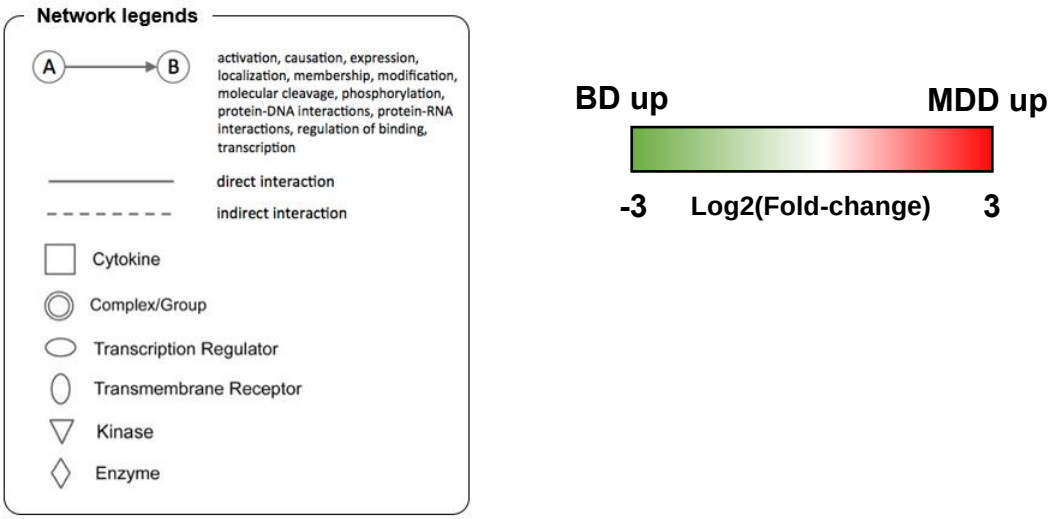


Figure 1-5. Top protein network generated by ingenuity pathway analysis of differentially expressed proteins between drug free major depressive disorder and bipolar disorder. Direct and indirect interactions are represented by the solid and dashed lines, respectively. The shapes represent the molecular classes of the proteins defined in the legend. The protein interaction networks were generated through the use of IPA. MDD, major depressive disorder; BP, bipolar disorder; IPA, Ingenuity Pathway Analysis.

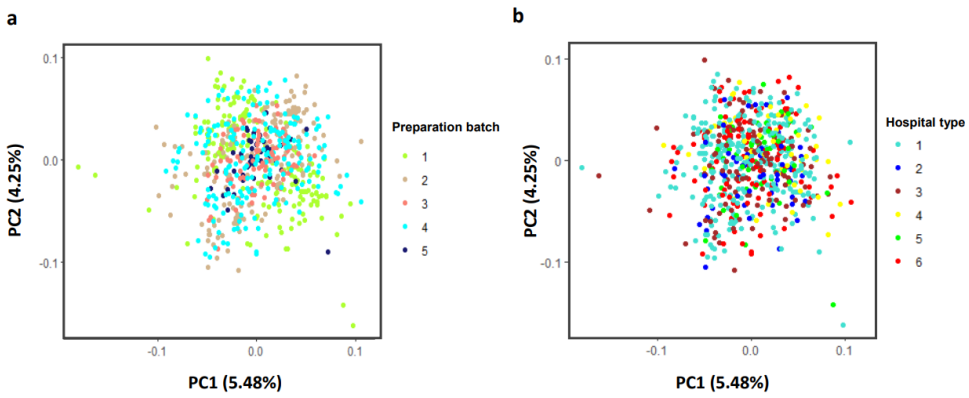


Figure 2-1. Principal component analysis plot of adjusted peak area ratio of LC-MRM-MS targets after batch effect correction. Principal component analysis was performed with the clinical samples after batch effect correction for sample preparation batches. (a) Sample preparation batches and (b) hosptial types are colored in the principal component analysis plot (1 = Seoul National Unviersity Hospital, 2 = Seoul Metropolitan Government Seoul National University Boramae Medical Center, 3 = Nowon Eulji Medical Center, Eulji University; 4 = Cha University Bundang Medical Center, 5 = Inha University Hospital, and 6 = Hanyang University Seoul Hospital)

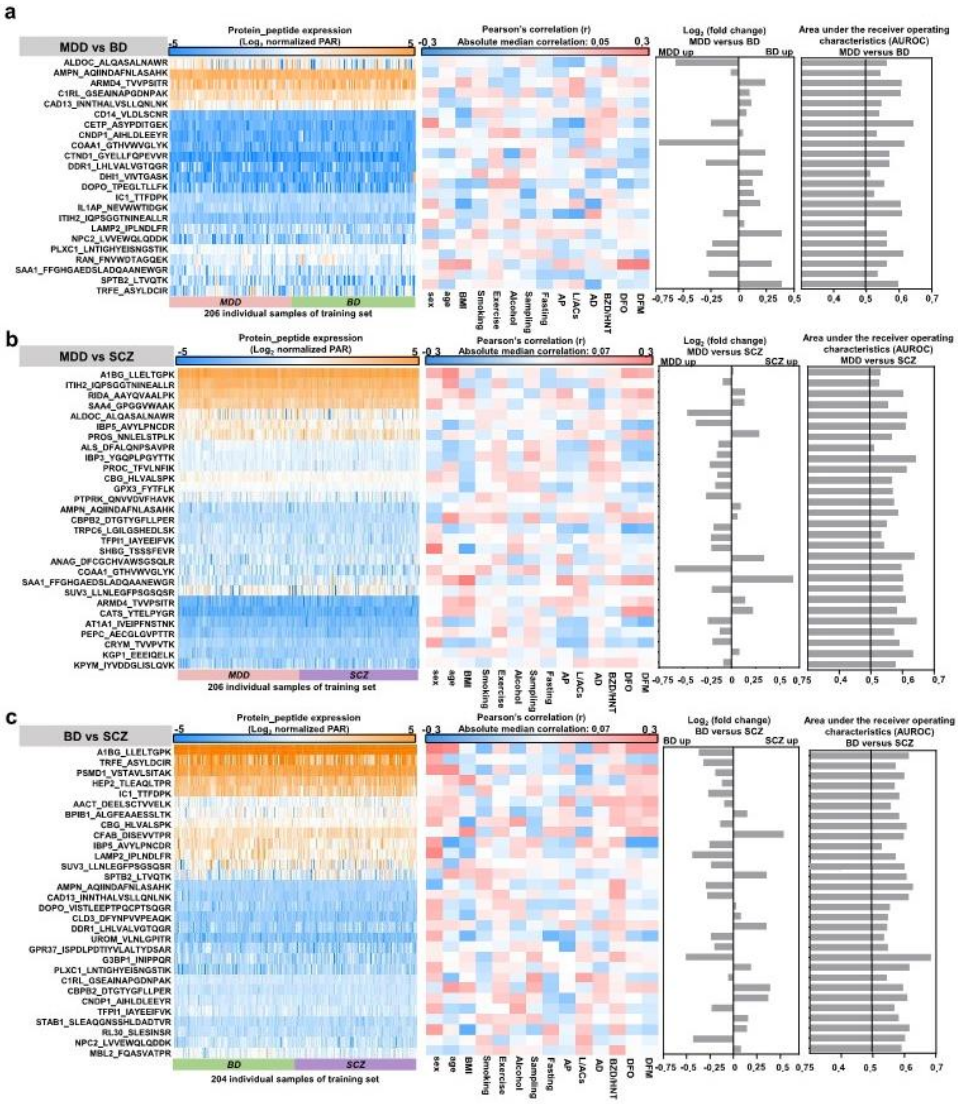


Figure 2-2. Univariate analysis of proteomic candiate features. Heatmaps of the proteomic candidate feature expression, and correlation plots of proteomic candidate feaures with demographics and clinical characteristics are presented. Alterations in fold change, and AUROCs of the individual proteomic candidate features are further plotted. Univariate analysis for (a) MDD versus BD, (b) MDD versus SCZ, and (c) BD versus SCZ.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, PAR = peak area ratio, AUROC = area under the reciever operating characteristics, BMI = body mass index, AP = antipsychotics, L/AC = lithium/anticonvulsant, AD = antidepressant, BDZ/HNT = benzodiazepine/hypnotic, DFO = duration from first onset, DFM = duration from first medication

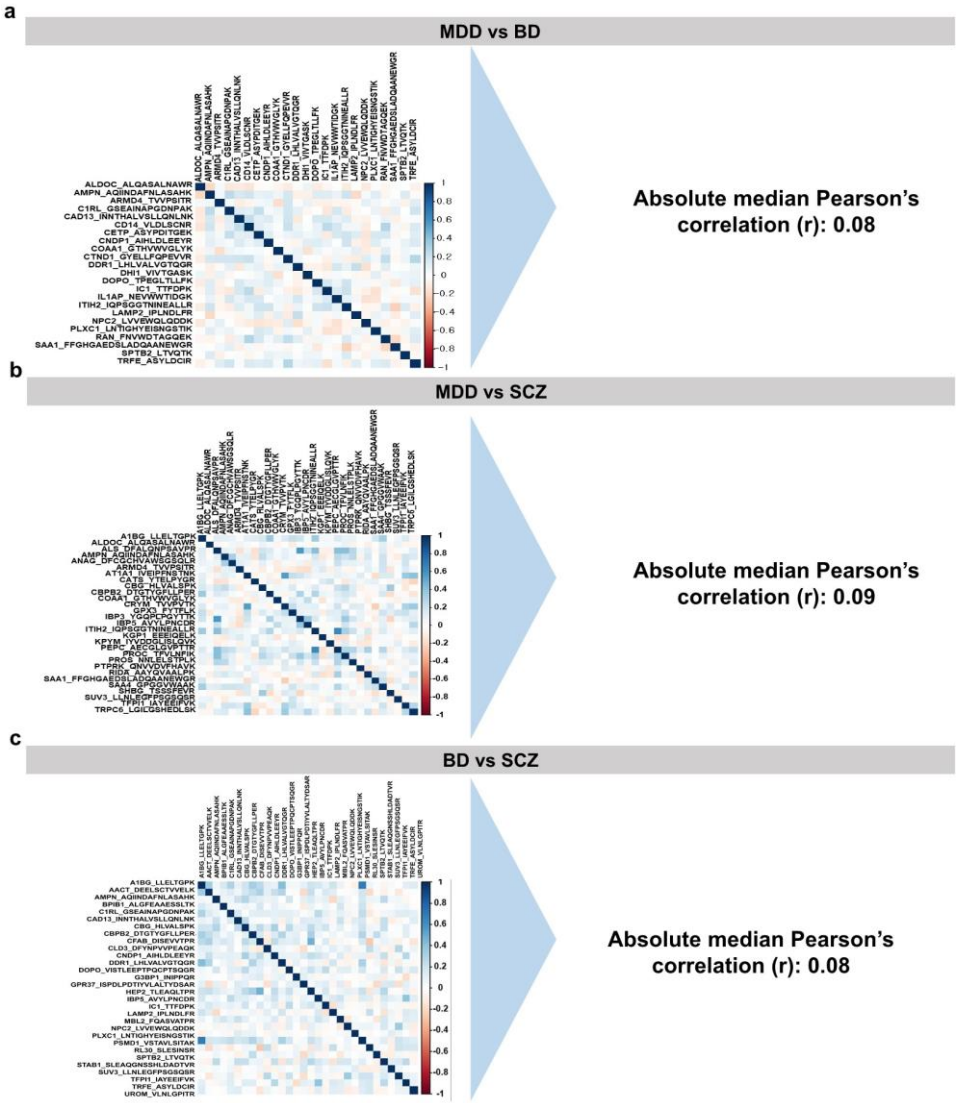


Figure 2-3. Correlation analysis between proteomic candidate features. The correlation matrix of the proteomic candidate features based on Pearson's correlation is presented for (a) MDD versus BD, (b) MDD versus SCZ, and (c) BD versus SCZ.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia

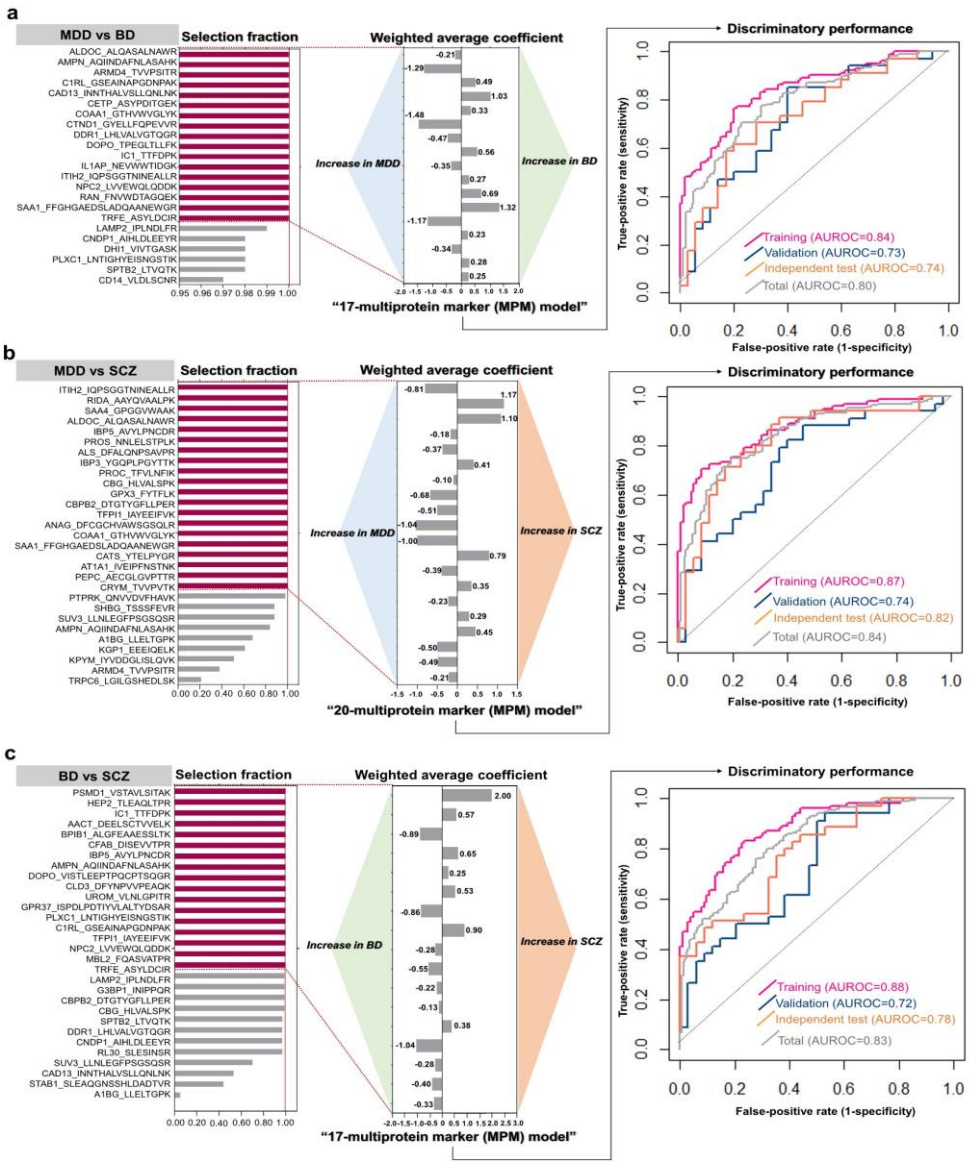


Figure 2-4. Development of multiprotein marker models based on selection fraction = 1. The selected features (selection fraction=1) in the MPM models are shown as pink bars. Weighted average coefficients and its directions for disease types are presented. Results of MPM models for (a) MDD versus BD, (b) MDD versus SCZ, and (c) BD versus SCZ.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, AUROC = area under the receiver operating characteristics, MPM = multiprotein marker

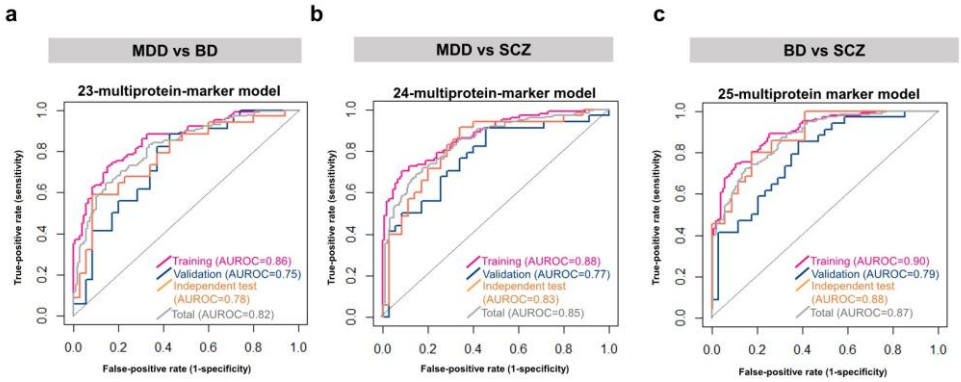


Figure 2-5. Differentiation performance of multiprotein marker models based on selection fraction ≥ 0.8.

Results of MPM models for (a) MDD versus BD, (b) MDD versus SCZ, and (c) BD versus SCZ.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, AUROC = area under the receiver operating characteristics.

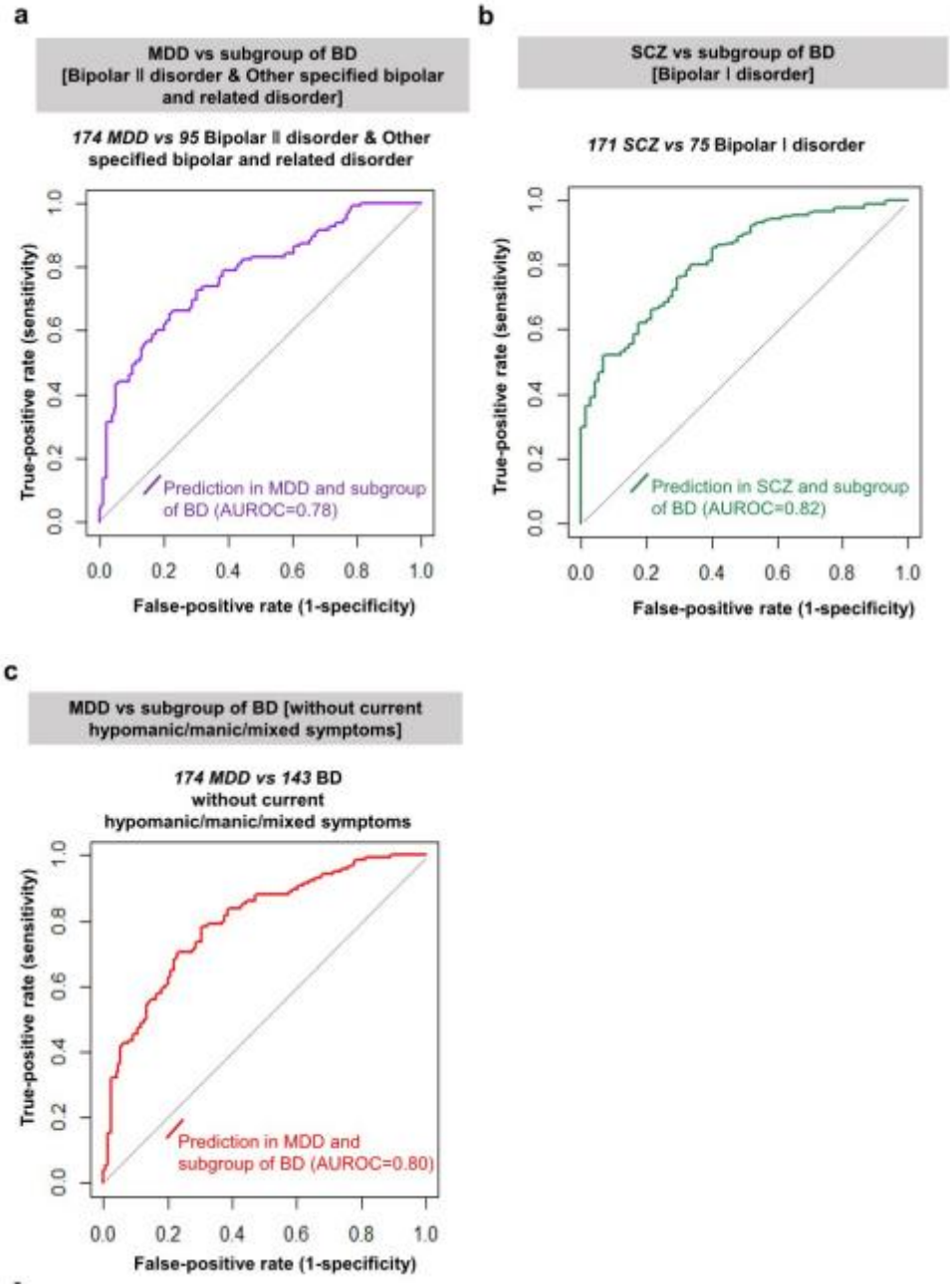


Figure 2-6. Differentiation performance of multiprotein marker models with bipolar disorder subgroups.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia.

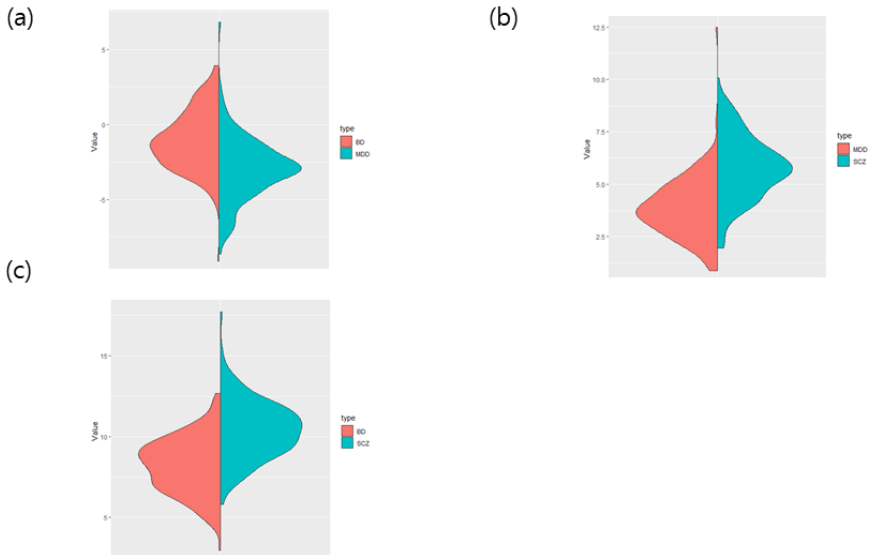


Figure 2-7. Violin plots for the final multiprotein marker model values

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia.

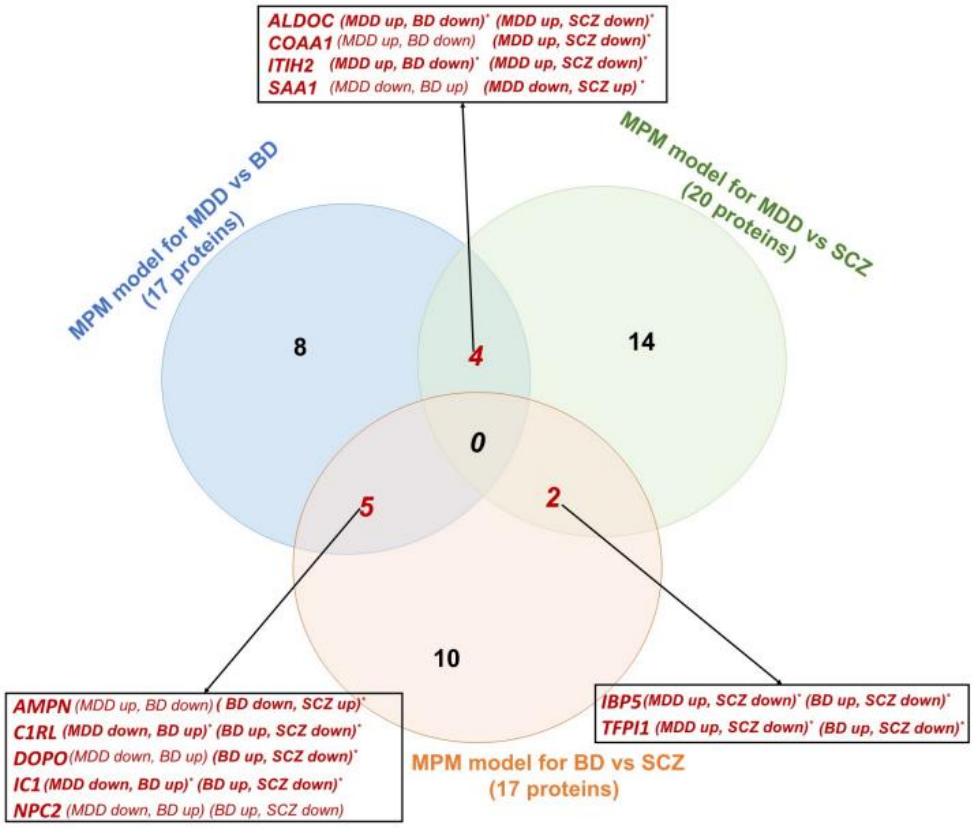


Figure 2-8. Overlapping proteomic features for the final multiprotein marker models and its expression levels. Overlapping proteomic features represented as protein entries between the final MPM models are shown, and their expression pattern is indicated as up for upregulation and down for downregulation. Proteomic features with significance differences between disease types are signified by bold font and asterisk.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, MPM = multiprotein marker

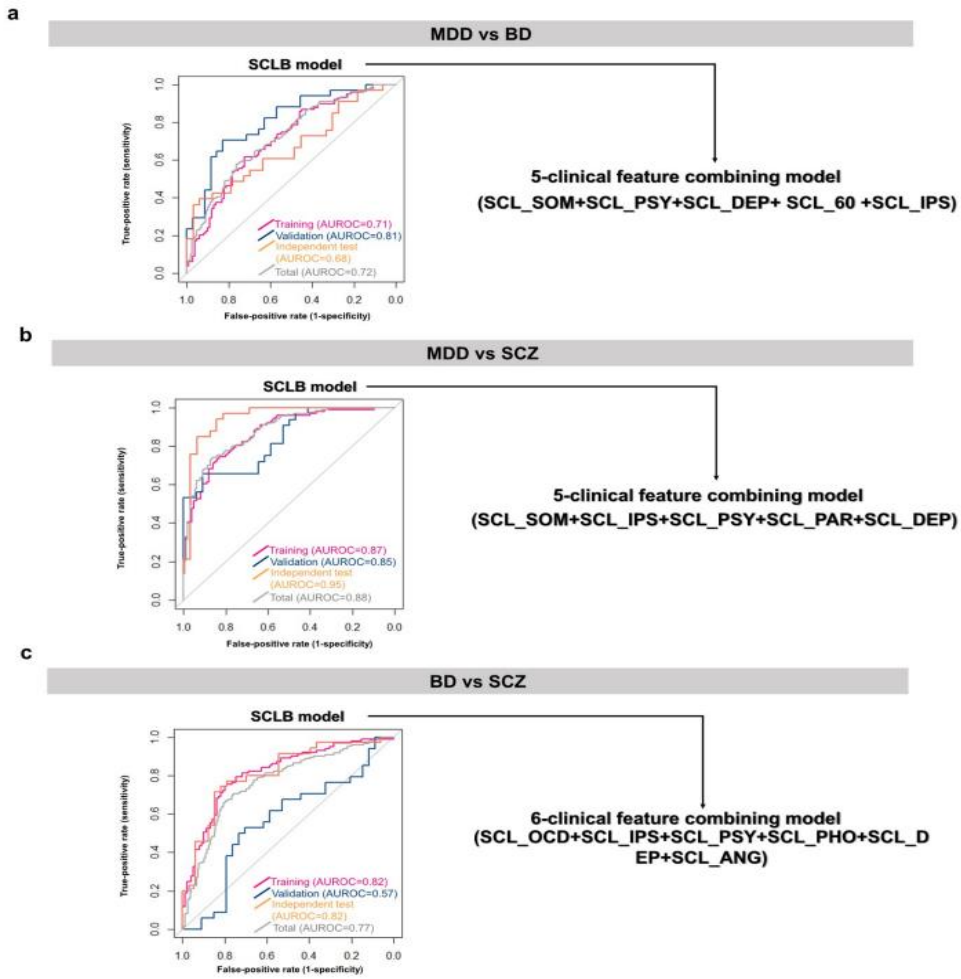
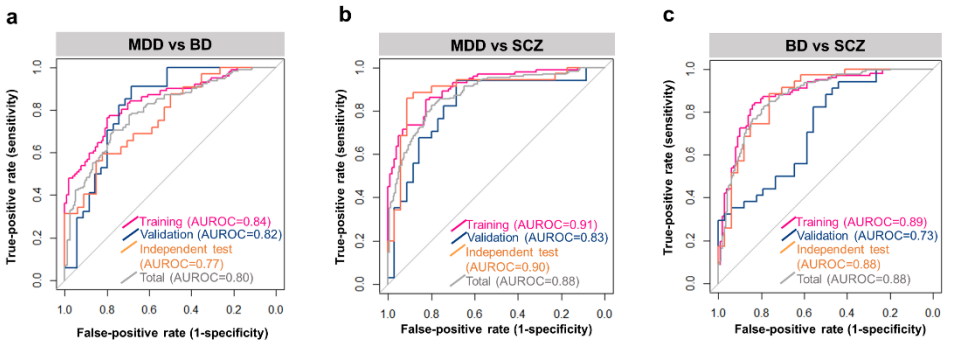


Figure 2-9. Differentiation performance of symptom checklist-based models. Results of SCLB models for (a) MDD versus BD, (b) MDD versus SCZ, and (c) BD versus SCZ. Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, SCLB = symptom checklist-based, SCL = Symptom Checklist-90-Revised, SOM = somatization dimension, OCD = obsessive-compulsive dimension, IPS = interpersonal sensitivity dimension, PSY = psychoticism dimension, 60 = overeating item, PAR = paranoid ideation dimension, PHO = phobic anxiety dimension, DEP = depression dimension, ANG = hostility dimension, AUROC = area under the receiver operating characteristics.



Performance results (Independent test set)

	Value
AUROC	*0.77*
Optimal cutoff	*0.48*
Accuracy	*0.76*
Sensitivity	*0.71*
Specificity	*0.82*
PPV	*0.77*
NPV	*0.75*

Performance results (Independent test set)

	Value
AUROC	*0.90*
Optimal cutoff	*0.50*
Accuracy	*0.89*
Sensitivity	*0.86*
Specificity	*0.92*
PPV	*0.87*
NPV	*0.91*

Performance results (Independent test set)

	Value
AUROC	*0.88*
Optimal cutoff	*0.62*
Accuracy	*0.83*
Sensitivity	*0.89*
Specificity	*0.77*
PPV	*0.80*
NPV	*0.87*

Figure 2-10. Differentiation and diagnostic performance of ensemble models. Results of ES models for (a) MDD versus BD, (b) MDD versus SCZ, and (c) BD versus SCZ. Optimal cutoff based on Youden index.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, ES = ensemble, BPRS = Brief Psychiatric Rating Scale, YMRS = Young Mania Rating Scale, MADRS = Montgomery-Asberg Depression Rating Scale, HAM-A = Hamilton Anxiety Scale, AUROC = area under the receiver operating characteristics, PPV = positive predictive value, NPV = negative predictive value

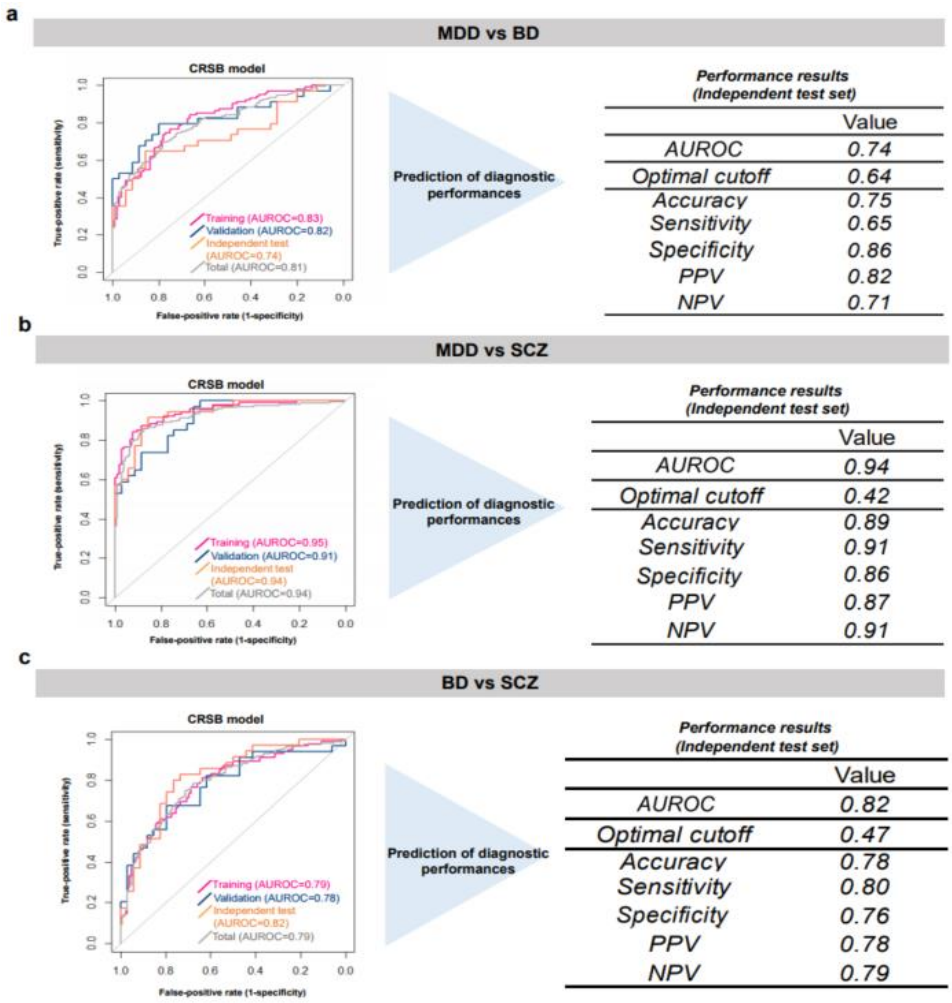


Figure 2-11. Differentiation and diagnostic performance of clinician rater score-based models. Results of CRSB models for (a) MDD versus BD, (b) MDD versus SCZ, and (c) BD versus SCZ. Optimal cutoff based on Youden index.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, CRSB = clinician rater score-based, BPRS = Brief Psychiatric Rating Scale, YMRS = Young Mania Rating Scale, MADRS = Montgomery-Asberg Depression Rating Scale, HAM-A = Hamilton Anxiety Scale, AUROC = area under the receiver operating characteristics, PPV = positive predictive value, NPV = negative predictive value

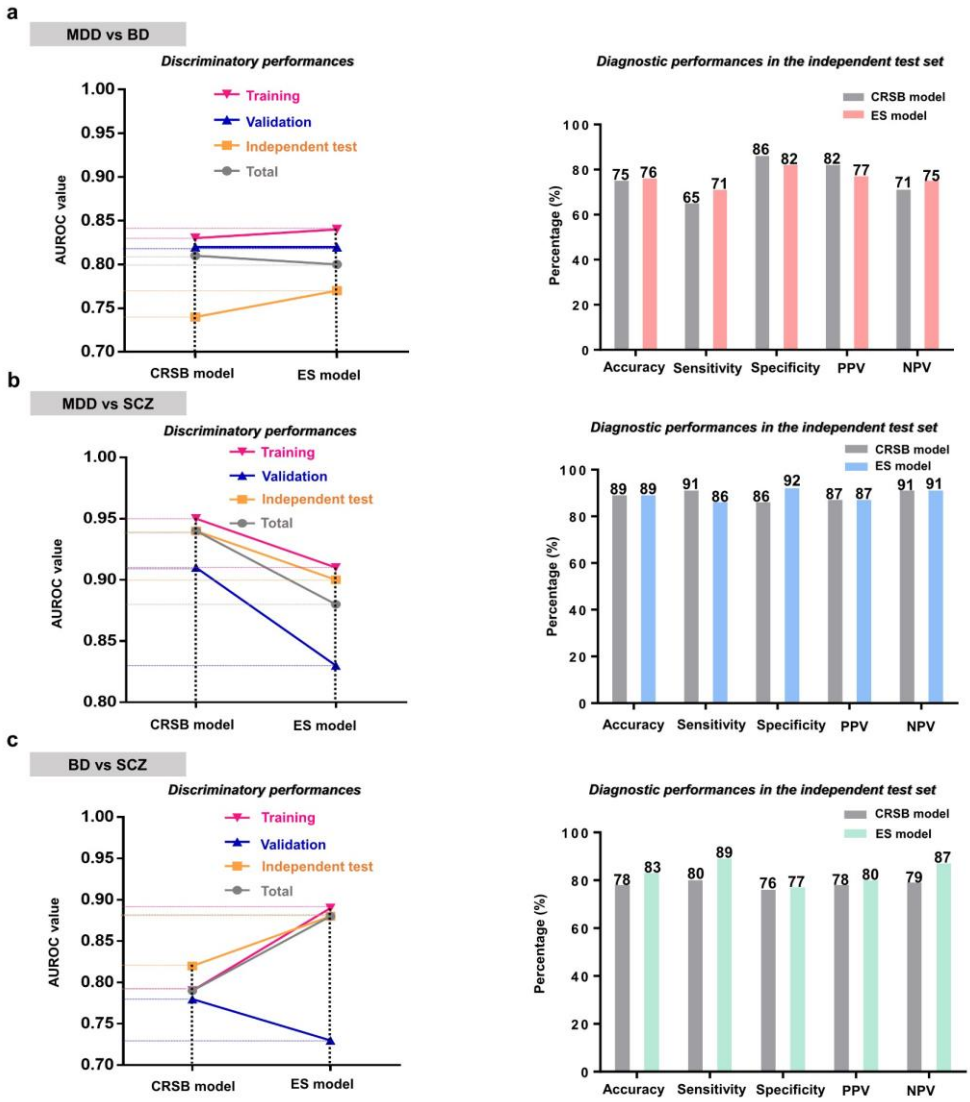


Figure 2-12. Performance comparison of ensemble models and clinician rater score-based models. Differentiation performances in each set, and diagnostic performances in the independent test set are compared for (a) MDD versus BD, (b) MDD versus SCZ, and (c) BD versus SCZ.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, CRSB = clinician rater score-based, ES = ensemble, AUROC = area under the receiver operating characteristics, PPV = positive predictive value, NPV = negative predictive value

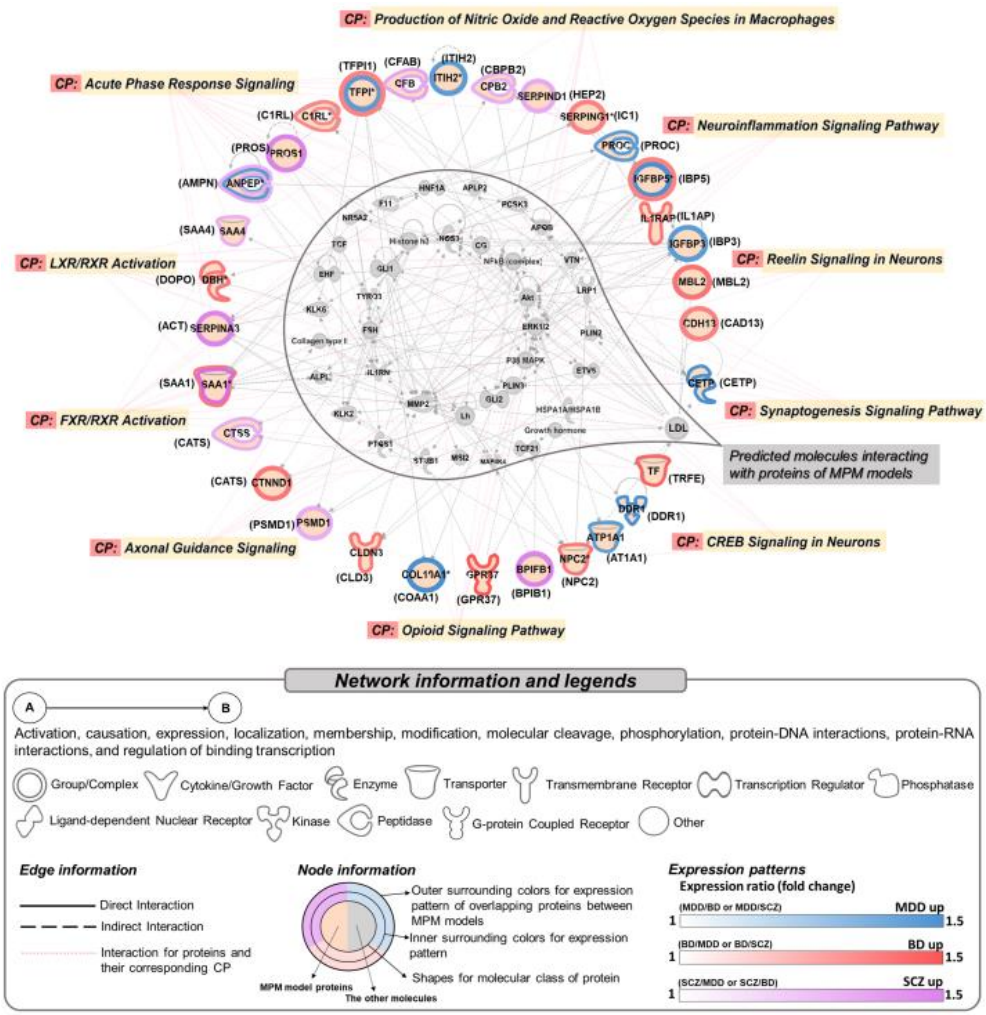


Figure 2-13. Merged protein networks and associated canonical pathways from proteomic features of the final multiprotein marker models. Merged protein networks with network score ≥ 20 and the corresponding canonical pathways were generated. Canonical pathways associated with proteins in the network are presented as light pink dotted lines. Overlapping proteomic features between the final MPM models are denoted by an asterisk. Each protein is presented as its gene name, and the corresponding protein entry is in parentheses.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, CP = canonical pathway, MPM = multiprotein marker.

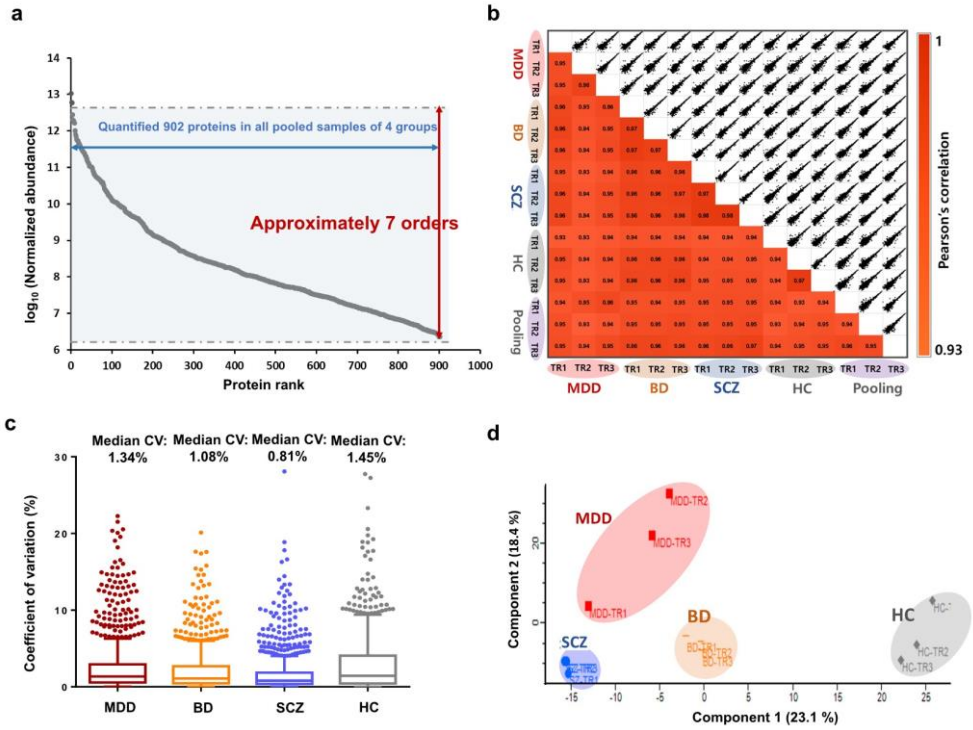


Figure 2-14. Proteomic profiling data of pooled plasma samples. (a) Dynamic range of the 902 quantified proteins. (b) Technical variances in MS analysis. (c) Coefficient of variation (CV) values between technical replicates of MS analysis. (d) Principal component analysis of distinct clusters.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, HC = healthy controls, TR = technical replicate, CV = coefficient of variation, MS = mass spectrometry

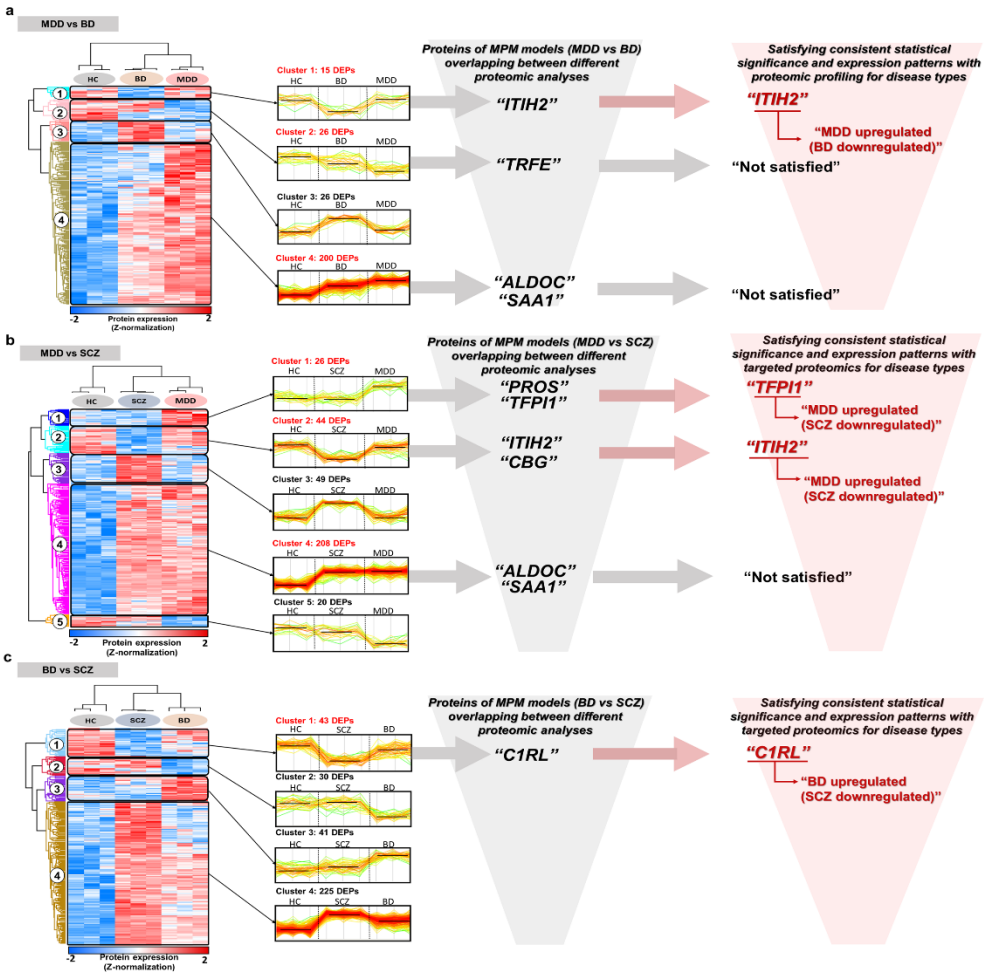


Figure 2-15. Consistency of proteomic features from the final multiprotein marker models between targeted proteomics and proteomic profiling data. Clusters from DEPs of proteomic profiling analysis and their corresponding expression levels are presented as heatmaps. For each cluster, the number of DEPs and expression patterns are presented. Overlapping proteins of the final MPM models with consistent statistical significance and expression pattern for both proteomic platforms are in wine-colored font. Results for (a) MDD versus BD, (b) MDD versus SCZ, and (c) BD versus SCZ.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, HC = healthy controls, MPM = multiprotein marker, DEP = differentially expressed protein

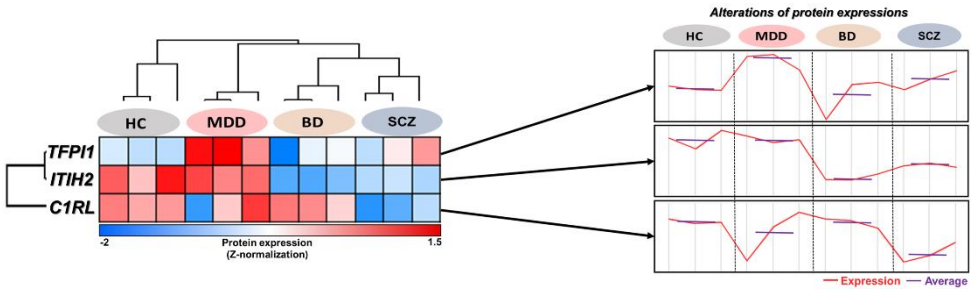


Figure 2-16. Proteomic profiling data of the consistent proteomic features in differentiating major psychiatric disorders. Alterations in expressions of the consistent proteomic features, which satisfied consistent statistical significance and expression pattern between targeted proteomics and proteomic profiling, are presented as heatmaps and line graphs. Alterations in protein expression are indicated by a red line, and average protein expression for each group is indicated by a purple line.

Abbreviations: MDD = major depressive disorder, BD = bipolar disorder, SCZ = schizophrenia, HC = healthy controls

Bibliography

1. Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. Lancet Psychiatry. 2016;3(2):171-8.

2. Baughman KR, Bonfine N, Dugan SE, Adams R, Gallagher M, Olds RS, et al. Disease Burden Among Individuals with Severe Mental Illness in a Community Setting. Community Ment Health J. 2016;52(4):424-32.

3. Chong HY, Teoh SL, Wu DB, Kotirum S, Chiou CF, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. Neuropsychiatr Dis Treat. 2016;12:357-73.

4. Ferrari AJ, Stockings E, Khoo JP, Erskine HE, Degenhardt L, Vos T, et al. The prevalence and burden of bipolar disorder: findings from the Global Burden of Disease Study 2013. Bipolar Disord. 2016;18(5):440-50.

5. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med. 2013;10(11):e1001547.

6. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005;62(6):593-602.

7. Moreno-Küstner B, Martín C, Pastor L. Prevalence of psychotic disorders and its association with methodological issues. A systematic review

and meta-analyses. PloS one. 2018;13(4):e0195687.

8. World Health Organization. Depression and other common mental disorders: global health estimates. World Health Organization; 2017.

9. GBD 2019 Mental Disorders Collaborators. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Psychiatry. 2022 Feb;9(2): 137-150.

10. Kim W, Jang SY, Chun SY, Lee TH, Han KT, Park EC. Mortality in Schizophrenia and Other Psychoses: Data from the South Korea National Health Insurance Cohort, 2002-2013. J Korean Med Sci. 2017;32(5):835-42.

11. 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. J Affect Disord. 2005;84(2-3):149-57.

12. Hirschfeld RM, Lewis L, Vornik LA. Perceptions and impact of bipolar disorder: how far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. J Clin Psychiatry. 2003;64(2):161-74.

13. Ghaemi SN, Boiman EE, Goodwin FK. Diagnosing bipolar disorder and the effect of antidepressants: a naturalistic study. J Clin Psychiatry. 2000;61(10):804-8.

14. Gonzalez-Pinto A, Gutierrez M, Mosquera F, Ballesteros J, Lopez P, Ezcurra J, et al. First episode in bipolar disorder: misdiagnosis and psychotic

symptoms. J Affect Disord. 1998;50(1):41-4.

15. Perlis RH. Misdiagnosis of bipolar disorder. Am J Manag Care. 2005;11(9 Suppl):S271-4.

16. Altamura AC, Goikolea JM. Differential diagnoses and management strategies in patients with schizophrenia and bipolar disorder. Neuropsychiatr Dis Treat. 2008;4(1):311-7.

17. Preece RL, Han SYS, Bahn S. Proteomic approaches to identify blood-based biomarkers for depression and bipolar disorders. Expert Rev Proteomics. 2018;15(4):325-40.

18. Patel S. Role of proteomics in biomarker discovery and psychiatric disorders: current status, potentials, limitations and future challenges. Expert Rev Proteomics. 2012;9(3):249-65.

19. Lakhan SE. Schizophrenia proteomics: biomarkers on the path to laboratory medicine? Diagn Pathol. 2006;1:11.

20. Hoehe MR, Morris-Rosendahl DJ. The role of genetics and genomics in clinical psychiatry. Dialogues Clin Neurosci. 2018;20(3):169-77.

21. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. Nature. 2016;537(7620):347-55.

22. Choudhary C, Mann M. Decoding signalling networks by mass spectrometry-based proteomics. Nat Rev Mol Cell Biol. 2010;11(6):427-39.

23. Frantzi M, Bhat A, Latosinska A. Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker

development. Clin Transl Med. 2014;3(1):7.

24. Domenici E, Muglia P. The search for peripheral disease markers in psychiatry by genomic and proteomic approaches. Expert Opin Med Diagn. 2007;1(2):235-51.

25. Haenisch F, Cooper JD, Reif A, Kittel-Schneider S, Steiner J, Leweke FM, et al. Towards a blood-based diagnostic panel for bipolar disorder. Brain Behav Immun. 2016;52:49-57.

26. Li Y, Zhou K, Zhang Z, Sun L, Yang J, Zhang M, et al. Label-free quantitative proteomic analysis reveals dysfunction of complement pathway in peripheral blood of schizophrenia patients: evidence for the immune hypothesis of schizophrenia. Mol Biosyst. 2012;8(10):2664-71.

27. Comes AL, Papiol S, Mueller T, Geyer PE, Mann M, Schulze TG. Proteomics for blood biomarker exploration of severe mental illness: pitfalls of the past and potential for the future. Transl Psychiatry. 2018;8(1):160.

28. Shi Y, Song R, Wang L, Qi Y, Zhang H, Zhu J, et al. Identifying Plasma Biomarkers with high specificity for major depressive disorder: A multi-level proteomics study. J Affect Disord. 2020;277:620-30.

29. Chen J, Huang C, Song Y, Shi H, Wu D, Yang Y, et al. Comparative proteomic analysis of plasma from bipolar depression and depressive disorder: identification of proteins associated with immune regulatory. Protein Cell. 2015;6(12):908-11.

30. Ren J, Zhao G, Sun X, Liu H, Jiang P, Chen J, et al. Identification of

plasma biomarkers for distinguishing bipolar depression from major depressive disorder by iTRAQ-coupled LC-MS/MS and bioinformatics analysis. Psychoneuroendocrinology. 2017;86:17-24.

31. Kittel-Schneider S, Hahn T, Haenisch F, McNeill R, Reif A, Bahn S. Proteomic Profiling as a Diagnostic Biomarker for Discriminating Between Bipolar and Unipolar Depression. Front Psychiatry. 2020;11:189.

32. Idemoto K, Ishima T, Niitsu T, Hata T, Yoshida S, Hattori K, et al. Platelet-derived growth factor BB: A potential diagnostic blood biomarker for differentiating bipolar disorder from major depressive disorder. J Psychiatr Res. 2021;134:48-56.

33. Tomasik J, Han SYS, Barton-Owen G, Mirea DM, Martin-Key NA, Rustogi N, et al. A machine learning algorithm to differentiate bipolar disorder from major depressive disorder using an online mental health questionnaire and blood biomarker data. Transl Psychiatry. 2021;11(1):41.

34. Smirnova L, Seregin A, Boksha I, Dmitrieva E, Simutkin G, Kornetova E, et al. The difference in serum proteomes in schizophrenia and bipolar disorder. BMC Genomics. 2019;20(Suppl 7):535.

35. Knöchel C, Kniep J, Cooper JD, Stäblein M, Wenzler S, Sarlon J, et al. Altered apolipoprotein C expression in association with cognition impairments and hippocampus volume in schizophrenia and bipolar disorder. Eur Arch Psychiatry Clin Neurosci. 2017;267(3):199-212.

36. Santa Cruz EC, Zandonadi FDS, Fontes W, Sussulini A. A pilot study

indicating the dysregulation of the complement and coagulation cascades in treated schizophrenia and bipolar disorder patients. Biochim Biophys Acta Proteins Proteom. 2021;1869(8):140657.

37. de Jesus JR, Galazzi RM, de Lima TB, Banzato CEM, de Almeida Lima ESLF, de Rosalmeida Dantas C, et al. Simplifying the human serum proteome for discriminating patients with bipolar disorder of other psychiatry conditions. Clin Biochem. 2017;50(18):1118-25.

38. Schulz SC, Overgaard S, Bond DJ, Kaldate R. Assessment of Proteomic Measures Across Serious Psychiatric Illness. Clin Schizophr Relat Psychoses. 2017;11(2):103-12.

39. Ding YH, Guo JH, Hu QY, Jiang W, Wang KZ. Protein Biomarkers in Serum of Patients with Schizophrenia. Cell Biochem Biophys. 2015;72(3):799-805.

40. Coppens V, De Wachter O, Goossens J, Hendrix J, Maudsley S, Azmi A, et al. Profiling of the Peripheral Blood Mononuclear Cell Proteome in Schizophrenia and Mood Disorders for the Discovery of Discriminatory Biomarkers: A Proof-of-Concept Study. Neuropsychobiology. 2020;79(4-5):324-34.

41. Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol. 1967;6(4):278-96.

42. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma Proteome Profiling to Assess Human Health and Disease. Cell Syst.

2016;2(3):185-95.

43. Han D, Jin J, Woo J, Min H, Kim Y. Proteomic analysis of mouse astrocytes and their secretome by a combination of FASP and StageTip-based, high pH, reversed-phase fractionation. Proteomics. 2014;14(13-14):1604-9.

44. Woo J, Han D, Park J, Kim SJ, Kim Y. In-depth characterization of the secretome of mouse CNS cell lines by LC-MS/MS without prefractionation. Proteomics. 2015;15(21):3617-22.

45. Nagaraj N, Kulak NA, Cox J, Neuhauser N, Mayr K, Hoerning O, et al. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. Mol Cell Proteomics. 2012;11(3):M111.013722.

46. Lee H, Kim K, Woo J, Park J, Kim H, Lee KE, et al. Quantitative Proteomic Analysis Identifies AHNAK (Neuroblast Differentiation-associated Protein AHNAK) as a Novel Candidate Biomarker for Bladder Urothelial Carcinoma Diagnosis by Liquid-based Cytology. Mol Cell Proteomics. 2018;17(9):1788-802.

47. Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc. 2016;11(12):2301-19.

48. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011;10(4):1794-805.

49. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016;13(9):731-40.

50. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014;30(4):523-30.

51. Alsaif M, Guest PC, Schwarz E, Reif A, Kittel-Schneider S, Spain M, et al. Analysis of serum and plasma identifies differences in molecular coverage, measurement variability, and candidate biomarker selection. Proteomics Clin Appl. 2012;6(5-6):297-303.

52. Rodriguez L, Mohamed NV, Desjardins A, Lippe R, Fon EA, Leclerc N. Rab7A regulates tau secretion. J Neurochem. 2017;141(4):592-605.

53. Ma G, Liu M, Du K, Zhong X, Gong S, Jiao L, et al. Differential Expression of mRNAs in the Brain Tissues of Patients with Alzheimer's Disease Based on GEO Expression Profile and Its Clinical Significance. Biomed Res Int. 2019;2019:8179145.

54. Armstrong A, Mattsson N, Appelqvist H, Janefjord C, Sandin L, Agholme L, et al. Lysosomal network proteins as potential novel CSF biomarkers for Alzheimer's disease. Neuromolecular Med. 2014;16(1):150-60.

55. Zhurov V, Stead JD, Merali Z, Palkovits M, Faludi G, Schild-Poulter C, et al. Molecular pathway reconstruction and analysis of disturbed gene

expression in depressed individuals who died by suicide. PloS One. 2012;7(10):e47581.

56. Mateus D, Marini ES, Progida C, Bakke O. Rab7a modulates ER stress and ER morphology. Biochim Biophys Acta Mol Cell Res. 2018;1865(5):781-93.

57. Lizaso A, Tan KT, Lee YH. beta-adrenergic receptor-stimulated lipolysis requires the RAB7-mediated autolysosomal lipid degradation. Autophagy. 2013;9(8):1228-43.

58. Gold PW, Licinio J, Pavlatou MG. Pathological parainflammation and endoplasmic reticulum stress in depression: potential translational targets through the CNS insulin, klotho and PPAR-gamma systems. Mol Psychiatry. 2013;18(2):154-65.

59. Pfaffenseller B, Wollenhaupt-Aguiar B, Fries GR, Colpo GD, Burque RK, Bristot G, et al. Impaired endoplasmic reticulum stress response in bipolar disorder: cellular evidence of illness progression. Int J Neuropsychopharmacol. 2014;17(9):1453-63.

60. Mondal AC, Fatima M. Direct and indirect evidences of BDNF and NGF as key modulators in depression: role of antidepressants treatment. Int J Neurosci. 2019 Mar;129(3):283-.

61. Segoviano-Mendoza M, Cardenas-de la Cruz M, Salas-Pacheco J, Vazquez-Alaniz F, La Llave-Leon O, Castellanos-Juarez F, et al. Hypocholesterolemia is an independent risk factor for depression disorder

and suicide attempt in Northern Mexican population. BMC Psychiatry. 2018;18(1):7.

62. Chong CM, Ai N, Lee SM. ROCK in CNS: Different Roles of Isoforms and Therapeutic Target for Neurodegenerative Disorders. Curr Drug Targets. 2017;18(4):455-62.

63. Nakagawa O, Fujisawa K, Ishizaki T, Saito Y, Nakao K, Narumiya S. ROCK-I and ROCK-II, two isoforms of Rho-associated coiled-coil forming protein serine/threonine kinase in mice. FEBS Lett. 1996;392(2):189-93.

64. Olivier JD, Akerud H, Skalkidou A, Kaihola H, Sundstrom-Poromaa I. The effects of antenatal depression and antidepressant treatment on placental gene expression. Front Cell Neurosci. 2014;8:465.

65. Lech K, Ackermann K, Revell VL, Lao O, Skene DJ, Kayser M. Dissecting Daily and Circadian Expression Rhythms of Clock-Controlled Genes in Human Blood. J Biol Rhythms. 2016;31(1):68-81.

66. Saito T, Hirano M, Ide T, Ichiki T, Koibuchi N, Sunagawa K, et al. Pivotal role of Rho-associated kinase 2 in generating the intrinsic circadian rhythm of vascular contractility. Circulation. 2013;127(1):104-14.

67. Krane-Gartiser K, Vaaler AE, Fasmer OB, Sorensen K, Morken G, Scott J. Variability of activity patterns across mood disorders and time of day. BMC Psychiatry. 2017;17(1):404.

68. Aksu M, Pleiner T, Karaca S, Kappert C, Dehne HJ, Seibel K, et al. Xpo7 is a broad-spectrum exportin and a nuclear import receptor. J Cell Biol.

2018;217(7):2329-40.

69. Johnson C, Drgon T, Liu QR, Walther D, Edenberg H, Rice J, et al. Pooled association genome scanning for alcohol dependence using 104,268 SNPs: validation and use to identify alcoholism vulnerability loci in unrelated individuals from the collaborative study on the genetics of alcoholism. Am J Med Genet B Neuropsychiatr Gen. 2006;141b(8):844-53.

70. Lydall GJ, Bass NJ, McQuillin A, Lawrence J, Anjorin A, Kandaswamy R, et al. Confirmation of prior evidence of genetic susceptibility to alcoholism in a genome-wide association study of comorbid alcoholism and bipolar disorder. Psychiatr Genet. 2011;21(6):294-306.

71. Edwards PA, Kennedy MA, Mak PA. LXRs; oxysterol-activated nuclear receptors that regulate genes controlling lipid homeostasis. Vascul Pharmacol. 2002;38(4):249-56.

72. Yang C, Zhou C, Li J, Chen Z, Shi H, Yang W, et al. Quantitative proteomic study of the plasma reveals acute phase response and LXR/RXR and FXR/RXR activation in the chronic unpredictable mild stress mouse model of depression. Mol Med Rep. 2018;17(1):93-102.

73. Bhattacharya A, Derecki NC, Lovenberg TW, Drevets WC. Role of neuro-immunological factors in the pathophysiology of mood disorders. Psychopharmacology. 2016;233(9):1623-36.

74. Wu Y, Tang J, Zhou C, Zhao L, Chen J, Zeng L, et al. Quantitative proteomics analysis of the liver reveals immune regulation and lipid

metabolism dysregulation in a mouse model of depression. Behav Brain Res. 2016;311:330-9.

75. Hamdani N, Bengoufa D, Godin O, Doukhan R, Le Guen E, Daban-Huard C, et al. Immunoglobulin sub-class distribution in bipolar disorder and schizophrenia: potential relationship with latent Toxoplasma Gondii infection. BMC Psychiatry. 2018;18(1):239.

76. Jabbi M, Arasappan D, Eickhoff SB, Strakowski SM, Nemeroff CB, Hofmann HA. Neuro-transcriptomic signatures for mood disorder morbidity and suicide mortality. J Psychiatr Res. 2020;127:62-74.

77. Najjar S, Pearlman DM, Alper K, Najjar A, Devinsky O. Neuroinflammation and psychiatric illness. J Neuroinflammation. 2013;10:43.

78. Garay-Baquero DJ, White CH, Walker NF, Tebruegge M, Schiff HF, Ugarte-Gil C, et al. Comprehensive plasma proteomic profiling reveals biomarkers for active tuberculosis. JCI Insight. 2020 Sep 17;5(18):e137427.

79. Vora N, Kalagiri R, Mallett LH, Oh JH, Wajid U, Munir S, et al. Proteomics and Metabolomics in Pregnancy-An Overview. Obstet Gynecol Surv. 2019;74(2):111-25.

80. Kim Y, Kang UB, Kim S, Lee HB, Moon HG, Han W, et al. A Validation Study of a Multiple Reaction Monitoring-Based Proteomic Assay to Diagnose Breast Cancer. J Breast Cancer. 2019;22(4):579-86.

81. Dong W, Qiu C, Gong D, Jiang X, Liu W, Liu W, et al. Proteomics

and bioinformatics approaches for the identification of plasma biomarkers to detect Parkinson's disease. Exp Ther Med. 2019;18(4):2833-42.

82. Ryan KM, Glaviano A, O'Donovan SM, Kolshus E, Dunne R, Kavanagh A, et al. Electroconvulsive therapy modulates plasma pigment epithelium-derived factor in depression: a proteomics study. Transl Psychiatry. 2017;7(3):e1073.

83. Noorbakhsh F, Aminian A, Power C. Application of "Omics" Technologies for Diagnosis and Pathogenesis of Neurological Infections. Curr Neurol Neurosci Rep. 2015;15(9):58.

84. Jayanthi S, Buie S, Moore S, Herning RI, Better W, Wilson NM, et al. Heavy marijuana users show increased serum apolipoprotein C-III levels: evidence from proteomic analyses. Mol Psychiatry. 2010;15(1):101-12.

85. World Health Organization. Global Recommendations on Physical Activity for Health: World Health Organization; 2010.

86. Hafkenscheid A. Psychometric evaluation of a standardized and expanded Brief Psychiatric Rating Scale. Acta Psychiatr Scand. 1991;84(3):294-300.

87. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry. 1978;133:429-35.

88. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry. 1979;134:382-9.

89. Hamilton M. The assessment of anxiety states by rating. Br J Med

Psychol. 1959;32(1):50-5.

90. Derogatis LR. SCL-90-R : Administration, scoring & procedures manual-II for the (revised) version and other instruments of the psychopathology rating scale series (2nd ed). Towson, MD: Clinical Psychometric Research. 1992:1-16.

91. Sussman N, Mullen J, Paulsson B, Vågerö M. Rates of remission/euthymia with quetiapine in combination with lithium/divalproex for the treatment of acute mania. J Affect Disord. 2007;100 Suppl 1:S55-63.

92. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. 2007;2(8):1896-906.

93. Han D, Moon S, Kim Y, Kim J, Jin J, Kim Y. In-depth proteomic analysis of mouse microglia using a combination of FASP and StageTip-based, high pH, reversed-phase fractionation. Proteomics. 2013;13(20):2984-8.

94. Jia P, Han G, Zhao J, Lu P, Zhao Z. SZGR 2.0: a one-stop shop of schizophrenia candidate genes. Nucleic Acids Res. 2017;45(D1):D915-D24.

95. Wu Y, Li X, Liu J, Luo XJ, Yao YG. SZDB2.0: an updated comprehensive resource for schizophrenia research. Hum Genet. 2020;139(10):1285-97.

96. Kim S, Webster MJ. The Stanley Neuropathology Consortium Integrative Database (SNCID) for Psychiatric Disorders. Neurosci Bull.

2019;35(2):277-82.

97. Gutierrez-Sacristan A, Grosdidier S, Valverde O, Torrens M, Bravo A, Pinero J, et al. PsyGeNET: a knowledge platform on psychiatric disorders and their genes. Bioinformatics. 2015;31(18):3075-7.

98. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, et al. Plasma Proteome Database as a resource for proteomics research: 2014 update. Nucleic Acids Res. 2014;42(Database issue):D959-65.

99. Ponten F, Schwenk JM, Asplund A, Edqvist PH. The Human Protein Atlas as a proteomic resource for biomarker discovery. J Intern Med. 2011;270(5):428-46.

100. Shin D, Rhee SJ, Lee J, Yeo I, Do M, Joo EJ, et al. Quantitative Proteomic Approach for Discriminating Major Depressive Disorder and Bipolar Disorder by Multiple Reaction Monitoring-Mass Spectrometry. J Proteome Res. 2021;20(6):3188-203.

101. Rhee SJ, Han D, Lee Y, Kim H, Lee J, Lee K, et al. Comparison of serum protein profiles between major depressive disorder and bipolar disorder. BMC Psychiatry. 2020;20(1):145.

102. Abbatiello SE, Mani DR, Keshishian H, Carr SA. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. Clin Chem. 2010;56(2):291-305.

103. Cuklina J, Lee CH, Williams EG, Collins B, Sajic T, Pedrioli P, Rodriguez-Martinez M, Aebersold R. Computational challenges in biomarker discovery from high-throughput proteomic data. Ph.D. thesis, ETH Zurich. 2018.

104. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1-22.

105. Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res. 2004;33(2):261-304.

106. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. E1071: Misc Functions of the Department of Statistics, Probability Therory Group, TU Wien. 2019.

107. Džeroski S, Ženko B. Is combining classifiers with stacking better than selecting the best one? Machine learning. 2004;54(3):255-73.

108. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.

109. Youden WJ. Index for rating diagnostic tests. Cancer. 1950;3(1):32-5.

110. Rutledge RB, Chekroud AM, Huys QJ. Machine learning and big data in psychiatry: toward clinical applications. Curr Opin Neurobiol. 2019;55:152-9.

111. Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci. 2004;44(1):1-12.

112. Mast AE. Tissue Factor Pathway Inhibitor: Multiple Anticoagulant Activities for a Single Protein. Arterioscler Thromb Vasc Biol. 2016;36(1):9-14.

113. Fei X, Wang H, Yuan W, Wo M, Jiang L. Tissue Factor Pathway Inhibitor-1 Is a Valuable Marker for the Prediction of Deep Venous Thrombosis and Tumor Metastasis in Patients with Lung Cancer. Biomed Res Int. 2017;2017:8983763.

114. Zhuo L, Hascall VC, Kimata K. Inter-alpha-trypsin inhibitor, a covalent protein-glycosaminoglycan-protein complex. J Biol Chem. 2004;279(37):38079-82.

115. Wang Q, Su X, Jiang X, Dong X, Fan Y, Zhang J, et al. iTRAQ technology-based identification of human peripheral serum proteins associated with depression. Neuroscience. 2016;330:291-325.

116. Focking M, Sabherwal S, Cates HM, Scaife C, Dicker P, Hryniewiecka M, et al. Complement pathway changes at age 12 are associated with psychotic experiences at age 18 in a longitudinal population-based study: evidence for a role of stress. Mol Psychiatry. 2021;26(2):524-33.

117. Jossin Y. Reelin Functions, Mechanisms of Action and Signaling Pathways During Brain Development and Maturation. Biomolecules. 2020 Jun 26;10(6):964.

118. Hye A, Lynham S, Thambisetty M, Causevic M, Campbell J, Byers HL, et al. Proteome-based plasma biomarkers for Alzheimer's disease. Brain. 2006;129(Pt 11):3042-50.

119. Chung YC, Ko HW, Bok E, Park ES, Huh SH, Nam JH, et al. The role of neuroinflammation on the pathogenesis of Parkinson's disease. BMB Rep. 2010;43(4):225-32.

120. Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. Am J Med Genet B Neuropsychiatr Genet. 2006;141b(3):261-8.

121. Nam E, Lee YB, Moon C, Chang KA. Serum Tau Proteins as Potential Biomarkers for the Assessment of Alzheimer's Disease Progression. Int J Mol Sci. 2020;21(14).

122. Sjöstedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. Science. 2020;367(6482).

123. Shweiki A, Rami M, Oeckl P, Steinacker P, Barschke P, Dorner-Ciossek C, et al. Proteomic analysis reveals a biosignature of decreased synaptic protein in cerebrospinal fluid of major depressive disorder. Transl Psychiatry. 2020;10(1):1-12.

Abstract in Korean

배경: 주요우울장애, 양극성장애, 조현병은 대표적인 주요정신질환으로 이들은 일생동안 지속되는 장해, 그리고 증가된 사망율과 연관되어 있다고 알려져 있다. 주관적 증상과 행동 관찰을 통한 진단 방법은 이들의 감별진단을 어렵게 할 때가 있다. 단백체 프로파일링과 표적 단백체 정량이 이러한 질환들을 객관적으로 감별하는데 도움이 될 수 있다는 최신 연구들이 나오고 있다. 그래서 이번 연구는, 말초 혈액의 단백체 정량을 바탕으로 질환 간을 비교, 감별하고자 하였다.

방법: 약물을 2주 이상 복용하지 않은 주요우울장애 환자 15명과 양극성장애 환자 10명의 혈청 시료로 질량 분석기 기반 단백체 프로파일링 분석을 시행하였다. *T*-검정을 통해 각 군 간 발현양에 유의하게 차이나는 단백질들을 찾아내고자 하였다 (연구 1). 이 연구는 174명의 주요우울장애, 171명의 조현병, 170명의 양극성장애, 그리고 160명의 정상대조군 혈장을 분석함으로써, 확장하고자 하였다. 질환들을 감별할 수 있는 표적 단백체를 정량하고, 단백체 프로파일링과 비교함으로써 단백체 발현 변화의 일관성을 확인고자 하였다. LASSO 회귀분석, 표적 단백체 변수 추출, 모델 평균화의 과정을 거쳐 다중단백체마커 모델을 만들어서 두

질환군을 짝지어서 감별하고자 하였다. 이러한 다중단백체마커 모델과 간이정신진단검사(SCL-90-R)를 결합하여 만든 앙상블 모델과, 임상가 척도 기반 모델의 성능을 비교하고자 하였다 (연구 2). 두 연구 모두 생물정보학 분석을 통해, 감별력이 있는 단백질들과 연관된 생물학적 기능을 예측하고자 하였다.

결과: 약물을 복용하지 않은 주요우울장애와 양극성장애 사이에서 발현양에 유의하게 차이나는 14개의 단백체를 발굴하였다. RAB7A, ROCK2는 주요우울장애에서, EPO7은 양극성장애에서 발현양이 증가되어 있었다 (연구 1). 두 질환군을 비교하는 다중단백체마커 모델은, 독립된 검정 데이터 셋에서 양호한 감별 성능을 보였다 (AUROC=0.74~0.82). 게다가, 앙상블 모델의 성능은 (AUROC=0.77~0.90) 전반적으로 임상가 척도 기반 모델의 성능과 (AUROC=0.74~0.94) 동등하였다 (연구 2). 두 연구 모두, 감별력이 있는 단백체들은 세포 기능과 면역/염증 경로와 연관되어 있었다.

결론: 이번 연구에서는 단백체 정량, 그리고 임상 데이터와의 통합을 통해 주요정신질환을 비교, 감별하는 방법의 가능성을 확인할 수 있었다. 이후 연구들은 종적인 방법으로 분석할 필요가 있다.

주요어: 주요우울장애, 양극성장애, 조현병, 단백체학,

다중반응탐색법, 단백체 프로파일링

학번: 2013-21686