

## ARTICLE OPEN



# Latent class analysis of psychotic-affective disorders with data-driven plasma proteomics

Sang Jin Rhee<sup>1,14</sup>, Dongyoon Shin<sup>2,14</sup>, Daun Shin<sup>3,4</sup>, Yoojin Song<sup>3,4</sup>, Eun-Jeong Joo<sup>5,6</sup>, Hee Yeon Jung<sup>3,7,8</sup>, Sungwon Roh<sup>9</sup>, Sang-Hyuk Lee<sup>10</sup>, Hyeyoung Kim<sup>11</sup>, Minji Bang<sup>10</sup>, Kyu Young Lee<sup>5,12</sup>, Se Hyun Kim<sup>4</sup>, Minah Kim<sup>3,4</sup>, Jihyeon Lee<sup>2</sup>, Jaeyeon Kim<sup>2,13</sup>, Yeongshin Kim<sup>2</sup>, Jun Soo Kwon<sup>3,4,8</sup>, Kyooseob Ha<sup>3,4,8</sup>, Youngsoo Kim<sup>2,13</sup> and Yong Min Ahn<sup>3,4,8</sup>✉

© The Author(s) 2023

Data-driven approaches to subtype transdiagnostic samples are important for understanding heterogeneity within disorders and overlap between disorders. Thus, this study was conducted to determine whether plasma proteomics-based clustering could subtype patients with transdiagnostic psychotic-affective disorder diagnoses. The study population included 504 patients with schizophrenia, bipolar disorder, and major depressive disorder and 160 healthy controls, aged 19 to 65 years. Multiple reaction monitoring was performed using plasma samples from each individual. Pathologic peptides were determined by linear regression between patients and healthy controls. Latent class analysis was conducted in patients after peptide values were stratified by sex and divided into tertile values. Significant demographic and clinical characteristics were determined for the latent clusters. The latent class analysis was repeated when healthy controls were included. Twelve peptides were significantly different between the patients and healthy controls after controlling for significant covariates. Latent class analysis based on these peptides after stratification by sex revealed two distinct classes of patients. The negative symptom factor of the Brief Psychiatric Rating Scale was significantly different between the classes ( $t = -2.070$ ,  $p = 0.039$ ). When healthy controls were included, two latent classes were identified, and the negative symptom factor of the Brief Psychiatric Rating Scale was still significant ( $t = -2.372$ ,  $p = 0.018$ ). In conclusion, negative symptoms should be considered a significant biological aspect for understanding the heterogeneity and overlap of psychotic-affective disorders.

*Translational Psychiatry* (2023)13:44; <https://doi.org/10.1038/s41398-023-02321-9>

## INTRODUCTION

Psychiatric disorders, including schizophrenia (SCZ), bipolar disorder (BD), and major depressive disorder (MDD), are known to exhibit within-disorder heterogeneity and between-disorder overlap [1]. This is probably due to the diagnostic procedure for psychiatric disorders, as they are based on subjective symptoms and behavioral observations, with a lack of biological validity. Thus, to discover a more homogenous biological subgroup, recent studies have focused on identifying subtypes using data-driven approaches [2]. Finding a biological subtype in a transdiagnostic sample can contribute to deepening our understanding of the pathophysiology and the heterogeneity and overlap between disorders. Research projects, such as the Roadmap for Mental Health Research in Europe (ROAMER) [3] and the National Institute of Mental Health (NIMH) Research Domain Criteria (RDoC) [4], have emphasized the need for these approaches.

Focusing on proteomics data using a data-driven approach, efforts have been made to discover subtypes of depression [5, 6]. A previous study using the Netherlands Study of Depression and Anxiety (NESDA), clustered patients with depression and anxiety based on biological features, including blood proteomics, and discovered three classes based on metabolic health [6]. However, to the best of our knowledge, no study has focused on proteomics-based cluster analysis in a transdiagnostic psychotic-affective disorder spectrum to date. This study was based on a previous study in which we differentiated SCZ, BD, and MDD using plasma proteins [7]. In this study, we aimed to identify clusters based on peripheral plasma proteomic data. Demographic and clinical characteristics were compared across the identified classes to understand the differences between them.

<sup>1</sup>Biomedical Research Institute, Seoul National University Hospital, Seoul, Republic of Korea. <sup>2</sup>Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea. <sup>3</sup>Department of Psychiatry, Seoul National University College of Medicine, Seoul, Republic of Korea. <sup>4</sup>Department of Neuropsychiatry, Seoul National University Hospital, Seoul, Republic of Korea. <sup>5</sup>Department of Neuropsychiatry, School of Medicine, Eulji University, Daejeon, Republic of Korea. <sup>6</sup>Department of Psychiatry, Uijeongbu Eulji Medical Center, Eulji University, Uijeongbu, Republic of Korea. <sup>7</sup>Department of Psychiatry, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea. <sup>8</sup>Institute of Human Behavioral Medicine, Seoul National University Medical Research Center, Seoul, Republic of Korea. <sup>9</sup>Department of Psychiatry, Hanyang University Hospital and Hanyang University College of Medicine, Seoul, Republic of Korea. <sup>10</sup>Department of Psychiatry, CHA Bundang Medical Center, CHA University School of Medicine, Seongnam, Republic of Korea. <sup>11</sup>Department of Psychiatry, Inha University Hospital, Incheon, Republic of Korea. <sup>12</sup>Department of Psychiatry, Nowon Eulji University Hospital, Seoul, Republic of Korea. <sup>13</sup>Institute of Medical and Biological Engineering Medical Research Center, Seoul National University College of Medicine, Seoul, Republic of Korea. <sup>14</sup>These authors contributed equally: Sang Jin Rhee, Dongyoon Shin. ✉email: [biolab@snu.ac.kr](mailto:biolab@snu.ac.kr); [aym@snu.ac.kr](mailto:aym@snu.ac.kr)

Received: 22 August 2022 Revised: 9 January 2023 Accepted: 13 January 2023

Published online: 06 February 2023

## MATERIALS AND METHODS

### Clinical samples

The study population comprised 515 patients (171 SCZ, 170 BD, and 174 MDD) and 160 healthy controls (HC) aged 19–65 years who were enrolled between August 2018 and December 2020. Statistical analysis was performed, excluding 11 patients who had missing covariate values. Patients were enrolled from Seoul National University Hospital; Nowon Eulji Medical Center, Eulji University; Seoul Metropolitan Government, Seoul National University Boramae Medical Center; Hanyang University Hospital; Inha University Hospital; and Cha University Bundang Medical Center. HC were recruited from Seoul National University Hospital via advertisements. The diagnosis was based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth edition (DSM-5), and confirmed using the Mini-International Neuropsychiatric Interview (MINI). HC had no psychiatric diagnosis based on the MINI and no known psychiatric family history among first- and second-degree relatives.

Patients and HC were excluded based on the following criteria: use of any anti-inflammatory analgesics (including nonsteroidal anti-inflammatory drugs and steroids, with the exception of acetaminophen) for the previous 2 weeks before participation; a history of neuromodulation or neurosurgery; central nervous system (CNS) diseases (including epilepsy, meningitis, Parkinson's disease, and stroke); cancer; tuberculosis; current lactation or pregnancy; a history of substance abuse other than alcohol, caffeine, and nicotine; intensive psychotherapy for the previous 2 months before participation; predicted intellectual disability; and difficulty in interpreting Korean. Current psychotropic medication use was not an exclusion criterion. Most studies were based on previous reports on the association between these conditions and altered protein expression [7]. Those who had recently received neuromodulation and psychotherapy were excluded to confine the treatment effects to psychotropic medications.

Plasma samples were collected in a 6-mL ethylenediaminetetraacetic acid (EDTA) tube (ref 367863; Becton, Dickinson and Company, Trenton, NJ) and centrifuged at 1100–1300×g for 10–15 min at room temperature or 4 °C. The supernatant was collected and stored in an Eppendorf tube at ≤−70 °C.

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human subjects/patients were approved by the Institutional Review Boards of Seoul National University Hospital (IRB No. 1806-1065-951) and all other participating hospitals. Written informed consent was obtained from each participant after the procedure was fully explained.

### Demographics and clinical features

The demographics considered were age, sex, body mass index (BMI), blood collection time, fasting time, current alcohol use, current exercise status, and current smoking status. Age and BMI were analyzed as continuous variables, and sex (men/women), blood collection time (AM, PM), fasting time (<8, ≥8 h), current alcohol use (yes/no), current exercise status (yes/no), and current smoking status (yes/no) were analyzed as dichotomous variables. Current alcohol use was defined as having at least one drink once per week. Current exercise status was defined using the World Health Organization's recommendation of moderate-intensity physical activity for at least 30 min once per week [8].

For patients, medication use was analyzed as a dichotomous variable for antipsychotics, lithium/anticonvulsants, antidepressants, and benzodiazepines/hypnotics. The chronicity of the disease or medication was assessed using the continuous parameters of the duration from first onset (years) and duration from first medication (years).

The primary symptoms considered for analysis were clinician rater scales. Symptom severity was assessed using the Brief Psychiatric Rating Scale (BPRS) [9], Young Mania Rating Scale [10], Montgomery–Åsberg Depression Rating Scale (MADRS) [11], and Hamilton Anxiety Scale [12]. Four factors of the 24-item BPRS were also considered based on a previous study [13].

The following scales were used for subjective reports. The Symptom Checklist-90-Revised [14] was considered for subjective symptoms. A brief form of the World Health Organization Quality of Life Assessment Instrument [15], Childhood Trauma Questionnaire [16], short form of the Wender Utah Rating Scale [17], Composite Scale of Morningness [18], and Seasonal Pattern Assessment Questionnaire [19] were also used.

### Plasma proteomic quantification

More specific methods for targeted proteomic analysis have been described previously [7], and are described in Supplementary methods.

For each plasma sample, the six most abundant proteins were depleted using a MARS-6 column (Agilent Technologies, Santa Clara, CA, USA). A total of 100 µg of protein was reduced with 0.2% RapiGest and 20 mM dithiothreitol at 60 °C for 1 h and alkylated with 100 mM iodoacetamide in the dark at room temperature for 30 min. The samples were then digested with trypsin solution at 37 °C for 4 h. Digestion was completed by adding 10% formic acid. The sample was centrifuged at 4 °C for 1 h, and 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 (<sup>13</sup>C<sub>6</sub><sup>15</sup>N<sub>2</sub> or <sup>13</sup>C<sub>6</sub><sup>15</sup>N<sub>4</sub>) [purity >70%].

Liquid chromatography–multiple reaction monitoring–mass spectrometry was performed using a 1260 Infinity HPLC system coupled to an Agilent 6490 triple quadrupole MS (Agilent Technologies, Santa Clara, CA, USA). 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 (formic acid/ acetonitrile (v/v)) at 50 µL/min for 10 min. After the valve position 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. The bound peptides were separated on the column and eluted with a linear gradient of 3–35% solvent B at 40 µL/min for 50 min.

Mass spectra were generated in positive ion mode. The collision energy was optimized by adding the intensities of the individual transitions that resulted in the largest peak areas. Only SIS peptides corresponding to the 642 target peptides were initially analyzed to evaluate their retention times. The retention times were then compared with those of endogenous target peptides by analyzing the matrix of endogenous peptides with SIS peptides of the targets and a heavy β-galactosidase peptide [purity >99%]. Subsequently, the final targets were quantified in the individual blood samples.

The raw data from the liquid chromatography–multiple reaction monitoring–mass spectrometry analysis were processed using Skyline (version 19.1.0) (MacCoss Lab, Seattle, WA, USA). Peptide quantification was calculated with the peak area ratio (PAR), which is the ratio of the endogenous to SIS peptide peak area. From the 642 target peptides, 54 unstable peptides with low intensities (intensity <1000), unequal retention times between light and heavy peptides, and skewed peaks were excluded. Subsequently, the final PAR values of 588 target peptides across 675 samples were normalized by the area of the heavy β-galactosidase peptide to reduce technical variability.

### Demographic differences between patients and HC

After excluding 11 patients with missing covariates, demographic differences were compared between patients with psychosis-affective disorders and HC. Categorical data were analyzed using the chi-square test, and continuous data were analyzed using independent *t*-tests.

### Determination of peptide markers differentiating patients and HC

As from our previous study, 588 stable peptide markers were found to be eligible as proteomic candidates [7]. Log<sub>2</sub> transformation followed by batch correction for sample preparation batches using the Combat algorithm was performed using the R package proBatch [20]. Peptides with PAR values ≤0.01 or ≥100 for at least 5% of the total study population were initially excluded, in line with our previous study [7]. First, peptide markers were chosen when they were statistically significant in univariate analysis with the peptide values as the dependent variable and patients versus HC as the dichotomous independent variable. Next, peptides that were still significant when controlling for age, sex, BMI, blood collection time, fasting time, alcohol use, exercise, and smoking behavior were selected. Additional control of hospital type was conducted, and the peptides that were still statistically significant were selected as initial candidates. To eliminate any residual hospital batch effects for patient clustering, peptides were further excluded when they were statistically different between at least two hospital types after post hoc analysis (Tukey's method) of analysis of variance (ANOVA), which was conducted in patients only.

### Latent class analysis

As there are no definite norms for these peptides, each peptide was divided into three groups, based on the first and second tertiles, and was labeled as low (−1), medium (0), and high (1), in line with a previous

study [6]. The procedure was stratified by sex, as it is known to affect certain markers [21]. Latent class analysis (LCA) was conducted, with the start values increased to 2000 random sets of starting values for the initial stage and 50 for the optimization stage, and the number of iterations increased to 50. The analysis started with one class, as additional classes were added, the optimal number of latent classes was confirmed by comparing the Akaike information criteria, Bayesian information criterion (BIC), sample-size adjusted BIC (ssaBIC), entropy, Lo–Mendell–Rubin likelihood ratio, and bootstrapped likelihood ratio test. Then, the study participants were assigned to each class based on posterior class probability.

After LCA, the demographic and clinical characteristics were compared between the latent classes. Categorical and continuous data were analyzed using the chi-square test and independent t-test, respectively. Additionally, a LCA was performed, including HC. In the additional analysis, as HC differed between classes, it was controlled as a covariate when performing linear regression to compare latent classes.

### Bioinformatics analysis

The peptides that were significantly different between clusters were subjected to ingenuity pathway analysis (QIAGEN, Hilden, Germany) for network analysis. The *p* values of the chi-squared test between latent classes were subjected when performing ingenuity pathway analysis. Subsequently, the diseases/functions and canonical pathways associated with the network were predicted using Fisher's exact test. Among the top ten diseases/functions and canonical pathways ranked according to statistical significance, diseases/functions and canonical pathways consisting of at least half of the peptides were determined.

### Statistical analysis

Statistical analyses for demographical/clinical and targeted proteomic data were performed using SPSS version 21.0 (IBM Corporation, Armonk, NY, USA) and R version 4.1.2 (<https://www.R-project.org>). LCA was performed using Mplus version 8.7 (Muthén & Muthén, Los Angeles, CA, USA). Statistical tests were two-tailed, and statistical significance was set at *p* value <0.05.

## RESULTS

### Demographic differences between patients and HC

Differences in demographics between patients with psychosis-affective disorder and HC were analyzed. The patients had higher BMI, exercised less frequently, drank less frequently, and smoked more frequently than HC. Fewer blood samples were collected from fasting patients (Table 1).

### Selection of proteomic candidates that significantly differ between patients and HC

Of the 588 peptide markers, 133 peptides were initially excluded, as the PAR was  $\leq 0.01$  or  $\geq 100$ , for at least 5% of the total study population. A total of 101 peptides were statistically significant in univariate analysis with the batch-corrected PAR values of the peptides as the dependent variable and patients versus HC as the independent variable. After additional control with age, sex, BMI, blood collection time, fasting time, alcohol use, exercise, and smoking behavior, only 30 peptides remained significant. Additional control of hospital type resulted in 18 peptides that were significantly different between patients and HC. Of the 18 peptides, six were excluded as they were significantly different between at least two hospitals in univariate analysis, when analyzed only in patients. The resulting 12 peptides were subjected to LCA (Supplementary Table 1 and Supplementary Fig 1).

### LCA

The fit indices of the LCA are summarized in Table 2. The Akaike information criteria decreased continuously as classes increased; however, the BIC was lowest with the two-class model, and the ssaBIC was lowest with the three-class model. The Lo–Mendell–Rubin likelihood ratio value was only significant for the two-class model, whereas bootstrapped likelihood ratio test

**Table 1.** Demographics of the study population (*n* = 664).

Characteristics	Patients ( <i>n</i> = 504)	Controls ( <i>n</i> = 160)	Statistics	<i>p</i> value <sup>a</sup>
Age, mean (SD), years	36.42 (12.60)	35.94 (11.20)	<i>t</i> = 0.451	0.65
Sex (Male), <i>n</i> (%)	191 (37.9%)	48 (30.0%)	$\chi^2$ = 3.287	0.07
BMI, mean (SD), kg/m <sup>2</sup>	24.60 (4.50)	22.08 (2.72)	<i>t</i> = 8.594	<0.001
Blood collection time: AM, <i>n</i> (%)	184 (36.5%)	69 (43.1%)	$\chi^2$ = 2.255	0.13
Fasting time: at least 8 h, <i>n</i> (%)	114 (22.6%)	87 (54.4%)	$\chi^2$ = 58.022	<0.001
Alcohol drinking (at least once a week), <i>n</i> (%)	153 (30.4%)	75 (46.9%)	$\chi^2$ = 14.696	<0.001
Exercise (moderate), <i>n</i> (%)	173 (34.3%)	116 (72.5%)	$\chi^2$ = 72.001	<0.001
Current smoker, <i>n</i> (%)	152 (30.2%)	8 (5.0%)	$\chi^2$ = 42.029	<0.001

Categorical variables based on chi-squares tests and continuous variables based on *t*-tests.

SD standard distribution, BMI body mass index.

<sup>a</sup>Boldface values are statistically significant at *p* < 0.05.

**Table 2.** Comparisons of fit indices of latent class analysis (*n* = 504).

Model	AIC	BIC	ssaBIC	Entropy	LMR-LR <i>p</i> value	BLRT <i>p</i> value	N (proportion %)			
							1	2	3	4
1 class	13336.81	13438.16	13361.98	N/A	N/A	N/A	504 (100.0%)			
2 class	13187.46	13394.36	13238.83	0.604	0.0001	<0.0001	265 (52.6%)	239 (47.4%)		
3 class	13155.09	13467.56	13232.67	0.750	0.0690	<0.0001	273 (54.2%)	215 (42.7%)	16 (3.2%)	
4 class	13133.93	13551.97	13237.73	0.800	0.7630	0.0200	235 (46.6%)	216 (42.9%)	31 (6.2%)	22 (4.4%)

AIC Akaike information criteria, BIC Bayesian information criterion, ssaBIC sample-size adjusted Bayesian information criterion, LMR-LR Lo–Mendell–Rubin likelihood ratio, BLRT Bootstrapped likelihood ratio test, N/A not applicable.

**Table 3.** Clinical characteristics between latent classes ( $n = 504$ ).

Characteristics	Class 1 GELS-increased $n = 239$	Class 2 GELS-decreased $n = 265$	Statistics	$p$ value <sup>a</sup>
Total scores of				
BPRS, mean (SD)	40.90 (9.52)	41.59 (9.18)	$t = -0.810$	0.42
YMRS, mean (SD)	4.29 (6.02)	3.65 (5.07)	$t = 1.293$	0.20
MADRS, mean (SD)	18.71 (11.18)	19.51 (11.29)	$t = -0.796$	0.43
HAM-A, mean (SD)	11.18 (7.05)	11.26 (7.19)	$t = -0.133$	0.90
Significant subscales/items				
BPRS negative symptoms factor, mean (SD)	4.24 (2.13)	4.65 (2.28)	$t = -2.070$	<b>0.039</b>
BPRS13, mean (SD) (self-neglect)	1.15 (0.52)	1.26 (0.66)	$t = -2.235$	<b>0.026</b>
BPRS18, mean (SD) (motor retardation)	1.36 (0.80)	1.52 (0.92)	$t = -2.157$	<b>0.031</b>
BPRS21, mean (SD) (excitement)	1.32 (0.81)	1.17 (0.52)	$t = 2.418$	<b>0.016</b>
BPRS23, mean (SD) (motor hyperactivity)	1.15 (0.44)	1.08 (0.30)	$t = 1.981$	<b>0.048</b>

Statistical analysis based on  $t$ -tests.

BPRS Brief Psychiatric Rating Scale, YMRS Young Mania Rating Scale, MADRS Montgomery–Asberg Depression Rating Scale, HAM-A Hamilton Anxiety Scale.

<sup>a</sup>Boldface values are statistically significant at  $p < 0.05$ .

was significant for all classes. Additionally, for models with three and four classes, the proportion of the smallest group was less than 50 individuals. In conclusion, the overall indices suggested that the two-class model was accepted as the final model. The latent class classification and protein expression tertiles are presented in Supplementary Table 2. Considering the differences in peptide levels, Class 1 was named the GELS-increased class, and Class 2 was named the GELS-decreased class.

No significant differences were found between the latent classes when the demographic characteristics were considered (Supplementary Table 3). Although significant differences were not observed in the total scores of clinician rater scales, the negative symptom factor of the BPRS was higher in the GELS-decreased latent class ( $t = -2.070$ ,  $p = 0.039$ ). Considering the individual items, the self-neglect and motor retardation items of the BPRS were increased in the GELS-decreased latent class, and the excitement and motor hyperactivity items of the BPRS were increased in the GELS-increased latent class (Table 3).

Next, when including HC, two latent classes were identified. The proportion of HC and various total scores/factors/items of rating and subjective scales differed between classes. After additional adjustment for HC in a linear regression model, the negative symptom factor of the BPRS ( $t = -2.372$ ,  $p = 0.018$ ), the emotional withdrawal and motor retardation item of the BPRS, and the lassitude, inability to feel, and suicidal thoughts items of the MADRS were increased in a latent class when compared with the other class (Supplementary Tables 4, 5).

### Bioinformatics analysis

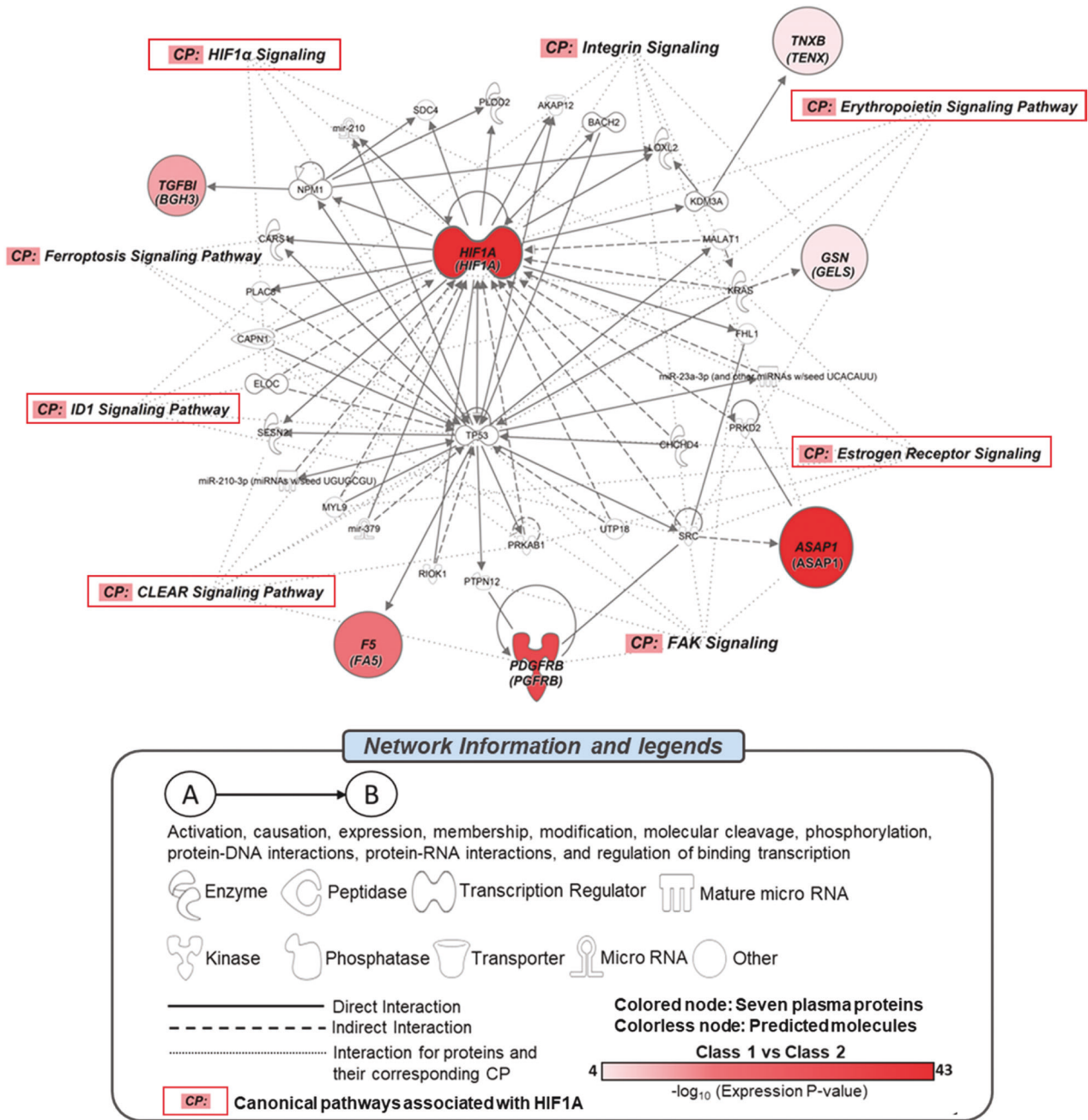
Among the eight proteins that differed between clusters (Supplementary Table 2), seven [ASAP1 (ASAP1), TGFBI (BGH3), F5 (FA5), GSN (GELS), HIF1A (HIF1A), PDGFRB (PGFRB), and TNXB (TENX) (represented by the gene symbol)] were included in the network (network score = 18), which comprised 35 molecules. Through criteria of determination for diseases/functions and canonical pathways, the network was associated with a total of eight diseases/functions, namely, dermatological diseases and conditions, inflammatory disease, inflammatory response, immunological disease, neurological disease, nervous system development and function, cellular assembly and organization, and cellular function and maintenance (Supplementary Table 6). Furthermore, the network was related to eight canonical pathways: HIF1 $\alpha$ , integrin, CLEAR, FAK, erythropoietin, estrogen receptor, ferroptosis, and ID1 signaling pathways (Fig. 1).

### DISCUSSION

This study enabled us to differentiate a group of patients within the transdiagnostic psychosis-affective spectrum into two latent classes based on 12 pathological peripheral plasma markers. The two classes differed significantly among eight peptides, and negative symptoms were significantly different between these classes. This tendency was preserved when HC was included.

The multiplex platform for quantifying proteomes in plasma enabled us to identify significant proteins (peptides) that differed between patients and HC. The number of proteins quantified was larger than that in previous studies, and the study population number was sufficient to control for multiple covariates known to be associated with proteomic expression [21] when selecting pathological proteins (significantly different proteins between patients and HC). Although these pathological proteins do not always reflect pathophysiology within patients, as shown in our previous study [7], protein selection was performed to reduce the likelihood of reflecting factors other than psychopathology [5, 6]. Additionally, although multiple covariates were controlled for when selecting pathological proteins, stratification by sex before conducting LCA enabled us to discard the effects of sex-specific protein expression. This was in line with previous studies that required stratification by both age and sex [5, 6]. Moreover, the LCA results were similar when the data were reanalyzed and HC was included. The results indicate that the classes are not specific only to psychiatric disorders but also a continuum with the normal population.

The study revealed 12 proteins that differed between patients and HC. ASAP1, BGH3, FA5, GELS, PFGRB, TENX, and VWF were increased in patients, and FCGBP, HIF1A, PAFA, S10A2, and SEM6C were increased in HC. The most significant pathological protein was PFGRB. Elevated levels of PFGRB in the cerebrospinal fluid is seen as a marker of blood–brain barrier dysfunction [22]. Interestingly, there is evidence that the levels of PGFRB in the cerebrospinal fluid and serum are positively associated [23]. As these psychiatric disorders have evidence of a leaky blood–brain barrier [24], this could have been reflected in the plasma of the study patients. The increased level of circulating VWF in patients, is in line with previous studies of SCZ, BD, and MDD [25, 26]. This implies that these psychiatric diseases might share a common mechanism of endothelium-related inflammation. Interestingly, VWF itself is known to influence blood–brain barrier permeability [27]. Another protein that has been reported in multiple psychiatric disorders was HIF1A. This molecule was also a key



**Fig. 1 Protein network and associated canonical pathways generated by IPA for eight proteins differentiating latent classes.** Seven of the eight proteins were included in the protein network. Direct and indirect interactions are represented by solid and dashed lines. Proteins are represented by a gene symbol with a protein entry name in parentheses. Shapes notify the molecular classes of proteins defined in the legend. Colored nodes notify the seven proteins, and white nodes represent predicted molecules. Canonical pathways involved in proteins in the network are represented by dotted lines. Canonical pathways in the red box are related to hub protein (HIF1A). Differences in protein expression levels of the seven features between the latent classes are represented by  $-\log_{10}$  (expression  $P$  value). IPA ingenuity pathway analysis, CP canonical pathway.

protein in the network analysis. HIF1A has been proposed to have a protective effect on depression [28]. However, its mRNA expression within peripheral white blood cells seems to be state-dependent in MDD and BD [29]. Although further studies are needed, this does imply that these diseases could be commonly involved in pathways of energy metabolism and oxidative stress.

Recent studies have started to apply transdiagnostic approaches to cluster patients with psychosis-affective disorders, considering biological correlates. Stein et al. (2021) revealed that the negative syndrome in a transdiagnostic sample like ours is

associated with the gray matter volume of the bilateral frontal opercula, and that no association exists between the diagnosis of the patients and the gray matter volume [30]. However, the study clustered the patients based on clinical data and analyzed the associations with gray matter volumes, showing multiple associations between other clinical clusters and other brain region volumes [30], which differed from our bottom-up approach using biological data to cluster patients. Another study by Chang et al. (2021) subtyped transdiagnostic patients based on frontal-posterior functional imbalance and found that the distribution

of SCZ, BD, and MDD differs between clusters, but does not differ in clinical symptoms [31]. More studies should be conducted using various biological correlates with a data-driven approach to cluster these patients and investigate their associations with clinical traits and symptoms.

Consensus has been reached on negative symptom domains that include blunted affect, anhedonia, avolition, and asociality, and two factors, namely, amotivation and diminished expression [32]. Our study revealed that the items, including emotional withdrawal, motor retardation, lassitude, and inability to feel, were associated with latent classes when analyzing patients, or patients with HC. Evaluations using scales specific for negative symptoms are needed to confirm the associations between specific domains and factors. Even though efforts have expanded, the reports on the pathological mechanisms of negative symptoms are inconsistent, since negative symptoms are also heterogeneous in its nature [32]. Most studies have been based on patients with psychosis, in which several studies have shown significant associations between negative symptoms and inflammatory biomarkers [33].

The present study revealed that specific pathways, including systemic inflammation, hypoxia, and signal transduction, were associated with latent classes. Estrogen receptor, erythropoietin, and integrin pathways have been proposed as significant pathophysiologies in previous studies of SCZ, BD, and MDD, including a recent systematic review of peripheral blood proteomes [34–39]. However, as these pathways have associations with multiple psychiatric conditions, it could be associated with a common psychiatric dimension. The present study proposes that negative symptoms have a potential association with these pathways. Especially, the estrogen receptor pathway has been proposed for its ameliorative role in negative symptoms of SCZ not only due to sex differences in its severity and prognosis, but as hormonal replacement therapy has a protective effect for negative symptoms in women [40]. Additionally, a recent report of single-cell level lymphocytes revealed that NF- $\kappa$ B p65 and Stat 3 cell signaling alterations were shared between MDD and SCZ in a transdiagnostic sample, and suggested that they could represent a shared substrate for negative symptomology [41]. Both proteins are known to be associated with integrin pathways [42, 43]. However, cautious interpretation of specific results is necessary because the proteins were not from the CNS, and these pathways all have intracellular components. Even though the seven proteins of the network are known to have secretory pathways or have the potential to be secreted from intracellular to extracellular regions [44–48], the link between plasma and the CNS is still under investigation. Nevertheless, considering that these symptoms are resistant to treatment, investigating the proposed mechanisms could expand our knowledge of their pathophysiology.

By contrast, the differential diagnosis of SCZ, BD, and MDD were not associated with latent classes. This implies that without the consideration of negative symptoms, the conventional differentiation between SCZ, BD, and MDD based on proteomics will have limitations, as it does not reflect systemic biological manifestations. Therefore, phenotypes should consider not only the ICD (International Statistical Classification of Diseases and Related Health Problems) or DSM (Diagnostic and Statistical Manual of Mental Disorders)-based symptom checklists but also the relationship between negative symptoms and associated biological correlates. This will deepen our understanding of the pathophysiology of psychotic-affective disorders and enable us to explain the heterogeneity within and the overlap between disorders.

However, the results also imply the obvious gap between circulating proteins and psychiatric manifestations, including structurally, the blood–brain barrier. These gaps are probably why a clearer separation in psychiatric manifestations between the latent classes were not seen, even though we selected proteins that differentiated psychiatric diseases and HC, and

controlled significant covariates. There could be other biological measures that are better for biological subtyping [5], that might reflect psychiatric symptoms with a stronger association.

### Strengths and limitations

This study has the following limitations. First, it was a cross-sectional study; therefore, causality could not be determined. A longitudinal study with multiple measurements of clinical symptoms and plasma proteomes would enable us to investigate the preservation of latent classes. Second, the proteins were obtained from the blood; therefore, functional analysis has limitations. Although there is evidence of blood–brain barrier dysfunction in psychiatric disorders, blood does not always reflect the CNS. Third, the differences between the scales were small. The scales themselves might not be able to capture negative symptoms sensitively; therefore, more detailed evaluations should be performed in the future. Especially the results of the individual items need to be cautiously interpreted, as they are numerically very close. Fourth, biological analysis was conducted using proteomics only. Integration with other omics or other biological correlates could capture sophisticated mechanisms that would reveal more accurate biological subtypes. Fifth, there were several preprocessing procedures to conduct LCA, including stratification by sex, and decreasing the dimension of protein values from a continuous to a discrete variable [6], which could have resulted in the loss of information, and affect the latent classes. However, LCA has its advantages, as it is considered a more statistically robust method than cluster analysis, since it is model-based, and generates fit statistics [49]. Finally, independent validation was not performed in this study.

Nevertheless, the strength of the study was that it was the first, to the knowledge of the authors, to compare a transdiagnostic psychosis-affective disorder population with numerous proteomic targets from the blood. Multiplexing proteins simultaneously enabled us to select significant pathological proteins and cluster patients into two biological subgroups. Negative symptoms need more attention because they vary between these subgroups, which tend to be neglected and remain poorly understood. Further biological studies should use longitudinal designs with detailed evaluation of negative symptoms to deepen our understanding of psychosis-affective disorders.

### DATA AVAILABILITY

The quantitation information of 12 targets included in the latent classes were deposited to the Panorama Public repository (<https://panoramaweb.org/Zzgv9f.url>). Email: panorama+reviewer162@proteinms.net, Password: TOUvWkEL. The other datasets presented in this study may be available from the corresponding authors on reasonable request.

### REFERENCES

- Chan CC, Shanahan M, Ospina LH, Larsen EM, Burdick KE. Premorbid adjustment trajectories in schizophrenia and bipolar disorder: a transdiagnostic cluster analysis. *Psychiatry Res.* 2019;272:655–62.
- Bowen EFW, Burgess JL, Granger R, Kleinman JE, Rhodes CH. DLPFC transcriptome defines two molecular subtypes of schizophrenia. *Transl Psychiatry.* 2019;9:147.
- Haro JM, Ayuso-Mateos JL, Bitter I, Demotes-Mainard J, Leboyer M, Lewis SW, et al. ROAMER: roadmap for mental health research in Europe. *Int J Methods Psychiatr Res.* 2014;23:1–14.
- Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry.* 2010;167:748–51.
- Beijers L, van Loo HM, Romeijn JW, Lamers F, Schoevers RA, Wardenaar KJ. Investigating data-driven biological subtypes of psychiatric disorders using specification-curve analysis. *Psychol Med.* 2020;52:1–12.
- Beijers L, Wardenaar KJ, Bosker FJ, Lamers F, van Grootheest G, de Boer MK, et al. Biomarker-based subtyping of depression and anxiety disorders using Latent Class Analysis. A NESDA study. *Psychol Med.* 2019;49:617–27.

7. Shin D, Rhee SJ, Shin D, Joo EJ, Jung HY, Roh S, et al. Integrating proteomic and clinical data to discriminate major psychiatric disorders: applications for major depressive disorder, bipolar disorder, and schizophrenia. *Clin Transl Med*. 2022;12:e929.
8. World Health Organization. Global recommendations on physical activity for health. Geneva: World Health Organization; 2010.
9. Hafkenscheid A. Psychometric evaluation of a standardized and expanded Brief Psychiatric Rating Scale. *Acta Psychiatr Scand*. 1991;84:294–300.
10. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry*. 1978;133:429–35.
11. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134:382–9.
12. Hamilton M. The assessment of anxiety states by rating. *Br J Med Psychol*. 1959;32:50–5.
13. Dazzi F, Shafer A, Lauriola M. Meta-analysis of the Brief Psychiatric Rating Scale - Expanded (BPRS-E) structure and arguments for a new version. *J Psychiatr Res*. 2016;81:140–51.
14. Derogatis LR. SCL-90-R: administration, scoring & procedures manual-II for the (revised) version and other instruments of the psychopathology rating scale series. *Clinical Psychometric Research*. 1992:1–16.
15. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. *Psychol Med*. 1998;28:551–8.
16. Bernstein DP, Fink L. Childhood trauma questionnaire: a retrospective self-report manual. San Antonio, TX: The Psychological Corporation; 1998.
17. Ward MF, Wender PH, Reimherr FW. The Wender Utah Rating Scale: an aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. *Am J Psychiatry*. 1993;150:885–90.
18. Smith CS, Reilly C, Midkiff K. Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. *J Appl Psychol*. 1989;74:728–38.
19. Rosenthal NE, Sack DA, Gillin JC, Lewy AJ, Goodwin FK, Davenport Y, et al. Seasonal affective disorder. A description of the syndrome and preliminary findings with light therapy. *Arch Gen Psychiatry*. 1984;41:72–80.
20. Čuklina J, Lee CH, Williams EG, Sajic T, Collins BC, Rodríguez Martínez M, et al. Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. *Mol Syst Biol*. 2021;17:e10240.
21. Preece RL, Han SYS, Bahn S. Proteomic approaches to identify blood-based biomarkers for depression and bipolar disorders. *Expert Rev Proteom*. 2018;15:325–40.
22. Montagne A, Barnes SR, Sweeney MD, Halliday MR, Sagare AP, Zhao Z, et al. Blood-brain barrier breakdown in the aging human hippocampus. *Neuron*. 2015;85:296–302.
23. Miners JS, Kehoe PG, Love S, Zetterberg H, Blennow K. CSF evidence of pericyte damage in Alzheimer's disease is associated with markers of blood-brain barrier dysfunction and disease pathology. *Alzheimers Res Ther*. 2019;11:81.
24. Futtrup J, Margolinsky R, Benros ME, Moos T, Routhé LJ, Rungby J, et al. Blood-brain barrier pathology in patients with severe mental disorders: a systematic review and meta-analysis of biomarkers in case-control studies. *Brain Behav Immun Health*. 2020;6:100102.
25. Hope S, Melle I, Aukrust P, Steen NE, Birkenaes AB, Lorentzen S, et al. Similar immune profile in bipolar disorder and schizophrenia: selective increase in soluble tumor necrosis factor receptor I and von Willebrand factor. *Bipolar Disord*. 2009;11:726–34.
26. Lopez-Vilchez I, Diaz-Ricart M, Navarro V, Torramade S, Zamorano-Leon J, Lopez-Farre A, et al. Endothelial damage in major depression patients is modulated by SSRI treatment, as demonstrated by circulating biomarkers and an in vitro cell model. *Transl Psychiatry*. 2016;6:e886.
27. Suidan GL, Brill A, De Meyer SF, Voorhees JR, Cifuni SM, Cabral JE, et al. Endothelial Von Willebrand factor promotes blood-brain barrier flexibility and provides protection from hypoxia and seizures in mice. *Arterioscler Thromb Vasc Biol*. 2013;33:2112–20.
28. Kang I, Kondo D, Kim J, Lyoo IK, Yurgelun-Todd D, Hwang J, et al. Elevating the level of hypoxia inducible factor may be a new potential target for the treatment of depression. *Med Hypotheses*. 2021;146:110398.
29. Shibata T, Yamagata H, Uchida S, Otsuki K, Hobarata T, Higuchi F, et al. The alteration of hypoxia inducible factor-1 (HIF-1) and its target genes in mood disorder patients. *Prog Neuropsychopharmacol Biol Psychiatry*. 2013;43:222–9.
30. Stein F, Meller T, Brosch K, Schmitt S, Ringwald K, Pfarr JK, et al. Psychopathological syndromes across affective and psychotic disorders correlate with gray matter volumes. *Schizophr Bull*. 2021;47:1740–50.
31. Chang M, Womer FY, Gong X, Chen X, Tang L, Feng R, et al. Identifying and validating subtypes within major psychiatric disorders based on frontal-posterior functional imbalance via deep learning. *Mol Psychiatry*. 2021;26:2991–3002.
32. van der Meer L, Kaiser S, Castelein S. Negative symptoms in schizophrenia: reconsidering evidence and focus in clinical trials. *Br J Psychiatry*. 2021;219:359–60.
33. Dunleavy C, Elsworth RJ, Upthegrove R, Wood SJ, Aldred S. Inflammation in first-episode psychosis: the contribution of inflammatory biomarkers to the emergence of negative symptoms, a systematic review and meta-analysis. *Acta Psychiatr Scand*. 2022;146:6–20.
34. Hwang WJ, Lee TY, Kim NS, Kwon JS. The role of estrogen receptors and their signaling across psychiatric disorders. *Int J Mol Sci*. 2020;22:373.
35. Jaudon F, Thalhammer A, Cingolani LA. Integrin adhesion in brain assembly: from molecular structure to neuropsychiatric disorders. *Eur J Neurosci*. 2021;53:3831–50.
36. Fan Y, Abrahamsen G, Mills R, Calderon CC, Tee JY, Leyton L, et al. Focal adhesion dynamics are altered in schizophrenia. *Biol Psychiatry*. 2013;74:418–26.
37. Miskowiak KW, Vinberg M, Harmer CJ, Ehrenreich H, Knudsen GM, Macoveanu J, et al. Effects of erythropoietin on depressive symptoms and neurocognitive deficits in depression and bipolar disorder. *Trials* 2010;11:97.
38. Ehrenreich H, Degner D, Meller J, Brines M, Behe M, Hasselblatt M, et al. Erythropoietin: a candidate compound for neuroprotection in schizophrenia. *Mol Psychiatry*. 2004;9:42–54.
39. Fernandes BS, Dai Y, Jia P, Zhao Z. Charting the proteome landscape in major psychiatric disorders: from biomarkers to biological pathways towards drug discovery. *Eur Neuropsychopharmacol*. 2022;61:43–59.
40. Rao ML, Kolsch H. Effects of estrogen on brain development and neuroprotection—implications for negative symptoms in schizophrenia. *Psychoneuroendocrinology* 2003;28:83–96.
41. Al-Kuraishy HM, Al-Gareeb AI, Waheed HJ, Al-Maiah TJ. Differential effect of metformin and/or glyburide on apelin serum levels in patients with type 2 diabetes mellitus: Concepts and clinical practice. *J Adv Pharm Technol Res*. 2018;9:80–6.
42. Scatena M, Almeida M, Chaisson ML, Fausto N, Nicosia RF, Giachelli CM. NF- $\kappa$ B mediates  $\alpha$ v $\beta$ 3 integrin-induced endothelial cell survival. *J Cell Biol*. 1998;141:1083–93.
43. Visavadiya NP, Keasey MP, Razskazovskiy V, Banerjee K, Jia C, Lovins C, et al. Integrin-FAK signaling rapidly and potently promotes mitochondrial function through STAT3. *Cell Commun Signal*. 2016;14:32.
44. Keerthikumar S, Chisanga D, Ariyaratne D, Al Saffar H, Anand S, Zhao K, et al. ExoCarta: a web-based compendium of exosomal cargo. *J Mol Biol*. 2016;428:688–92.
45. 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:D959–65.
46. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019;37:420–3.
47. Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. *Protein Eng Des Sel*. 2004;17:349–56.
48. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305:567–80.
49. Sinha P, Calfee CS, Delucchi KL. Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. *Crit Care Med*. 2021;49:e63–e79.

## ACKNOWLEDGEMENTS

We would like to thank the participants who took part in this study. This work was supported by the Industrial Strategic Technology Development Program, funded by the Ministry of Trade, Industry, and Energy (MOTIE, Korea) (No.20000134). This work was also supported by the Korea Health Industry Development Institute (No. HU22C0052).

## AUTHOR CONTRIBUTIONS

SJR and DS1 (Dongyoon Shin) contributed equally to this work and share the first authorship. SJR, DS1, DS2, YS, E-JJ, HYJ, SR, S-HL, HK, MB, YK2 (Youngsoo Kim), and YMA contributed to the study concept and design. DS1, JL, JK, YK1, and YK2 acquired the proteomic data. SJR, DS2, YS, E-JJ, HYJ, SR, S-HL, HK, MB, KYL, SHK, MK, JSK, KH, and YMA obtained the clinical data. SJR and DS1 performed the statistical analysis. SJR, DS1, YS2, and YMA drafted this manuscript. All authors critically reviewed the manuscript and approved the final version.

## COMPETING INTERESTS

YMA receives research support from or serves as a speaker for Janssen Korea Ltd., Lundbeck Korea Co. Ltd, and Korea Otsuka Pharmaceutical. The remaining authors declare no competing interests.

**ADDITIONAL INFORMATION**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41398-023-02321-9>.

**Correspondence** and requests for materials should be addressed to Youngsoo Kim or Yong Min Ahn.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023