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

2 Sang Jin Rhee[1,#], Dongyoon Shin[2,#], Daun Shin[3,4], Yoojin Song[3,4], Eun-Jeong Joo[5,6], Hee Yeon

3 Jung[3,7,8] Sungwon Roh[9], Sang-Hyuk Lee[10], Hyeyoung Kim[11], Minji Bang[10], Kyu Young Lee[5,12],

4 Se Hyun Kim[4], Minah Kim[3,4], Jihyeon Lee[2], Jaenyeon Kim[2,13], Yeongshin Kim[2,13], Jun Soo

5 Kwon[3,4,8], Kyooseob Ha[3,4,8], and Youngsoo Kim[2,13*], Yong Min Ahn[3,4,8*]

6

7 Running title: Plasma proteome-based clustering of psychotic-affective disorders

8

9 [1]Biomedical Research Institute, Seoul National University Hospital, Seoul, Republic of Korea

10 [2]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul,

11 Republic of Korea

12 [3]Department of Psychiatry, Seoul National University College of Medicine, Seoul, Republic of

13 Korea

14 [4]Department of Neuropsychiatry, Seoul National University Hospital, Seoul, Republic of Korea

15 [5]Department of Neuropsychiatry, School of Medicine, Eulji University, Daejeon, Republic of

16 Korea

17 [6]Department of Psychiatry, Uijeongbu Eulji Medical Center, Eulji University, Uijeongbu,

18 Republic of Korea

19 [7]Department of Psychiatry, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea

20 [8]Institute of Human Behavioral Medicine, Seoul National University Medical Research Center,

21 Seoul, Republic of Korea

22   [9]Department of Psychiatry, Hanyang University Hospital and Hanyang University College of

23   Medicine, Seoul, Republic of Korea

24   [10]Department of Psychiatry, CHA Bundang Medical Center, CHA University School of

25   Medicine, Seongnam, Republic of Korea

26   [11]Department of Psychiatry, Inha University Hospital, Incheon, Republic of Korea

27   [12]Department of Psychiatry, Nowon Eulji University Hospital, Seoul, Republic of Korea

28   [13]Institute of Medical and Biological Engineering Medical Research Center, Seoul National

29   University College of Medicine, Seoul, Republic of Korea

30   [#] These authors contributed equally to this work as joint 1st authors

31

32   **Corresponding authors**

33   *Youngsoo Kim, Ph.D.

34   Department of Biomedical Sciences & Institute of Medical and Biological Engineering

35   Medical Research Center, Seoul National University College of Medicine, 103 Daehakro Seoul

36   03080, Republic of Korea. (Tel) +82-2-740-8073, (Fax) +82-2-741-0253, (E-mail)

37   biolab@snu.ac.kr

38   and

39   *Yong Min Ahn, M.D., Ph.D.

40   Department of Psychiatry, Seoul National University College of Medicine/Department of

41   Neuropsychiatry, Seoul National University Hospital/Institute of Human Behavioral

42   Medicine, Seoul National University Medical Research Center, 101 Daehakro Seoul 03080,

43   Republic of Korea. (Tel) +82-2-2072-0710, (Fax) +82-2-744-7241, (E-mail) aym@snu.ac.kr

Abstract

44

45   Data-driven approaches to subtype transdiagnostic samples are important for understanding

46   heterogeneity within disorders and overlap between disorders. Thus, this study was conducted

47   to determine whether plasma proteomics-based clustering could subtype patients with

48   transdiagnostic psychotic-affective disorder diagnoses. The study population included 504

49   patients with schizophrenia, bipolar disorder, and major depressive disorder and 160 healthy

50   controls, aged 19 to 65 years. Multiple reaction monitoring was performed using plasma

51   samples from each individual. Pathologic peptides were determined by linear regression

52   between patients and healthy controls. Latent class analysis was conducted in patients after

53   peptide values were stratified by sex and divided into tertile values. Significant demographic

54   and clinical characteristics were determined for the latent clusters. The latent class analysis was

55   repeated when healthy controls were included. Twelve peptides were significantly different

56   between the patients and healthy controls after controlling for significant covariates. Latent

57   class analysis based on these peptides after stratification by sex revealed two distinct classes

58   of patients. The negative symptom factor of the Brief Psychiatric Rating Scale was significantly

59   different between the classes (t=-2.070, p=0.039). When healthy controls were included, two

60   latent classes were identified, and the negative symptom factor of the Brief Psychiatric Rating

61   Scale was still significant (t=-2.372, p=0.018). In conclusion, negative symptoms should be

62   considered a significant biological aspect for understanding the heterogeneity and overlap of

63   psychotic-affective disorders.

64

65

66

67   **Abbreviations**

68   SCZ: schizophrenia

69   BD: bipolar disorder

70   MDD: major depressive disorder

71   HC: healthy controls

72   MINI: Mini-International Neuropsychiatric Interview

73   CNS: central nervous system

74   BMI: body mass index

75   BPRS: Brief Psychiatric Rating Scale

76   MADRS: Montgomery–Asberg Depression Rating Scale

77   SIS: stable isotope-labeled internal standard

78   PAR: peak area ratio

79   LCA: latent class analysis

80   BIC: Bayesian information criterion

81   ssaBIC: sample-size adjusted BIC

82

83

84

85

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.

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.

134   Plasma samples were collected in a 6-mL ethylenediaminetetraacetic acid (EDTA) tube (ref

135   367863; Becton, Dickinson and Company, Trenton, NJ) and centrifuged at 1100–1300 g for

136   10–15 min at room temperature or 4 °C. The supernatant was collected and stored in an

137   Eppendorf tube at ≤-70 °C.

138   The authors assert that all procedures contributing to this work comply with the ethical

139   standards of the relevant national and institutional committees on human experimentation and

140   the Helsinki Declaration of 1975, as revised in 2008. All procedures involving human

141   subjects/patients were approved by the Institutional Review Boards of Seoul National

142   University Hospital (IRB No. 1806-1065-951) and all other participating hospitals. Written

143   informed consent was obtained from each participant after the procedure was fully explained.

144

145   **Demographics and clinical features**

146   The demographics considered were age, sex, body mass index (BMI), blood collection time,

147   fasting time, current alcohol use, current exercise status, and current smoking status. Age and

148   BMI were analyzed as continuous variables, and sex (men/women), blood collection time (AM,

149   PM), fasting time (<8 h, ≥8 h), current alcohol use (yes/no), current exercise status (yes/no),

150   and current smoking status (yes/no) were analyzed as dichotomous variables. Current alcohol

151   use was defined as having at least one drink once per week. Current exercise status was defined

152   using the World Health Organization recommendation of moderate-intensity physical activity

153   for at least 30 min once per week (8).

154   For patients, medication use was analyzed as a dichotomous variable for antipsychotics,

155   lithium/anticonvulsants, antidepressants, and benzodiazepines/hypnotics. The chronicity of the

156   disease or medication was assessed using the continuous parameters of the duration from first

157   onset (years) and duration from first medication (years).

158 The primary symptoms considered for analysis were clinician rater scales. Symptom severity

159 was assessed using the Brief Psychiatric Rating Scale (BPRS) (9), Young Mania Rating Scale

160 (10), Montgomery–Asberg Depression Rating Scale (MADRS) (11), and Hamilton Anxiety

161 Scale (12). Four factors of the 24-item BRPS were also considered based on a previous study

162 (13).

163 The following scales were used for subjective reports. The Symptom Checklist-90-Revised (14)

164 was considered for subjective symptoms. A brief form of the World Health Organization

165 Quality of Life Assessment Instrument (15), Childhood Trauma Questionnaire (16), short form

166 of the Wender Utah Rating Scale (17), Composite Scale of Morningness (18), and Seasonal

167 Pattern Assessment Questionnaire (19) were also used.

168

169 **Plasma proteomic quantification**

170 More specific methods for targeted proteomic analysis have been described previously (7), and

171 are described in Supplementary methods. For each plasma sample, the six most abundant

172 proteins were depleted using a MARS-6 column (Agilent Technologies, Santa Clara, CA, USA).

173 A total of 100 μg of protein was reduced with 0.2% RapiGest and 20 mM dithiothreitol at 60 °C

174 for 1 h and alkylated with 100 mM iodoacetamide in the dark at room temperature for 30 min.

175 The samples were then digested with trypsin solution at 37 °C for 4 h. Digestion was completed

176 by adding 10% formic acid. The sample was centrifuged at 4 °C for 1 h, and the supernatants

177 were spiked with crude stable isotope-labeled internal standard (SIS) peptides, in which a C-

178 terminal lysine or arginine was heavy isotope-labeled ($^{13}C_6^{15}N_2$ or $^{13}C_6^{15}N_4$) [purity > 70%].

179 Liquid chromatography–multiple reaction monitoring–mass spectrometry was performed

180 using a 1260 Infinity HPLC system coupled to an Agilent 6490 triple quadrupole MS (Agilent

181 Technologies, Santa Clara, CA, USA). For each digested sample, 40 μL was injected into a

182   guard column (2.1 × 15.0 mm, 1.8 μm, 80 Å) (Agilent Technologies, Santa Clara, CA, USA),

183   and online desalting was conducted in 3% solvent B (formic acid/acetonitrile (v/v)) at 50

184   μL/min for 10 min. After the valve position was switched, the sample was transferred to the

185   analytical column (0.5 × 35.0 mm, 3.5 μm, 80 Å) (Agilent Technologies, Santa Clara, CA,

186   USA) in 3% solvent B at 40 μL/min for 5 min. The bound peptides were separated on the

187   column and eluted with a linear gradient of 3–35% solvent B at 40 μL/min for 50 min.

188   Mass spectra were generated in positive ion mode. The collision energy was optimized by

189   adding the intensities of the individual transitions that resulted in the largest peak areas. Only

190   SIS peptides corresponding to the 642 target peptides were initially analyzed to evaluate their

191   retention times. The retention times were then compared with those of endogenous target

192   peptides by analyzing the matrix of endogenous peptides with SIS peptides of the targets and

193   a heavy β-galactosidase peptide [purity > 99%]. Subsequently, the final targets were quantified

194   in the individual blood samples.

195   The raw data from the liquid chromatography–multiple reaction monitoring–mass

196   spectrometry analysis were processed using Skyline (version 19.1.0) (MacCoss Lab, Seattle,

197   WA, USA). Peptide quantification was calculated with the peak area ratio (PAR), which is the

198   ratio of the endogenous to SIS peptide peak area. From the 642 target peptides, 54 unstable

199   peptides with low intensities (intensity < 1000), unequal retention times between light and

200   heavy peptides, and skewed peaks were excluded. Subsequently, the final PAR values of 588

201   target peptides across 675 samples were normalized by the area of the heavy β-galactosidase

202   peptide to reduce technical variability.

203

204   **Demographic differences between patients and HC**

205 After excluding 11 patients with missing covariates, demographic differences were compared

206 between patients with psychosis-affective disorders and HC. Categorical data were analyzed

207 using the chi-square test, and continuous data were analyzed using independent t-tests.

208

Determination of peptide markers differentiating patients and HC

210 As from our previous study, 588 stable peptide markers were found to be eligible as proteomic

211 candidates (7). Log2 transformation followed by batch correction for sample preparation

212 batches using the Combat algorithm was performed using the R package proBatch (20).

213 Peptides with PAR values ≤0.01 or ≥100 for at least 5% of the total study population were

214 initially excluded, in line with our previous study (7). First, peptide markers were chosen when

215 they were statistically significant in a univariate analysis with the peptide values as the

216 dependent variable and patients versus HC as the dichotomous independent variable. Next,

217 peptides that were still significant when controlling for age, sex, BMI, blood collection time,

218 fasting time, alcohol use, exercise, and smoking behavior were selected. Additional control of

219 hospital type was conducted, and the peptides that were still statistically significant were

220 selected as initial candidates. To eliminate any residual hospital batch effects for patient

221 clustering, peptides were further excluded when they were statistically different between at

222 least two hospital types after post-hoc analysis (Tukey's method) of analysis of variance

223 (ANOVA), which was conducted in patients only.

224

Latent class analysis

226 As there are no definite norms for these peptides, each peptide was divided into three groups,

227 based on the first and second tertiles, and was labeled as low (−1), medium (0), and high (1),

228   in line with a previous study (6). The procedure was stratified by sex, as it is known to affect

229   certain markers (21). Latent class analysis (LCA) was conducted, with the start values

230   increased to 2000 random sets of starting values for the initial stage and 50 for the optimization

231   stage, and the number of iterations was increased to 50. The analysis started with one class, as

232   additional classes were added, the optimal number of latent classes was confirmed comparing

233   the Akaike information criteria, Bayesian information criterion (BIC), sample-size adjusted

234   BIC (ssaBIC), entropy, Lo–Mendell–Rubin likelihood ratio, and bootstrapped likelihood ratio

235   test. Then, the study participants were assigned to each class based on posterior class

236   probability.

237   After LCA, the demographic and clinical characteristics were compared between the latent

238   classes. Categorical and continuous data were analyzed using the chi-square test and

239   independent t-test, respectively. Additionally, an LCA was performed including HC. In the

240   additional analysis, as HC differed between classes, it was controlled as a covariate when

241   performing linear regression to compare latent classes.

242

243   **Bioinformatics analysis**

244   The peptides that were significantly different between clusters were subjected to ingenuity

245   pathway analysis (QIAGEN, Hilden, Germany) for network analysis. The p-values of the chi-

246   squared test between latent classes were subjected when performing ingenuity pathway

247   analysis. Subsequently, the diseases/functions and canonical pathways associated with the

248   network were predicted using Fisher's exact test. Among the top 10 diseases/functions and

249   canonical pathways ranked according to statistical significance, diseases/functions and

250   canonical pathways consisting of at least half of the peptides were determined.

251

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.

258

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).

(Insert 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 ≤0.01 or ≥100, for at least 5% of the total study population. A total of 101 peptides were statistically significant in a 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

273   18 peptides that were significantly different between patients and HC. Of the 18 peptides, six

274   were excluded as they were significantly different between at least two hospitals in a univariate

275   analysis, when analyzed only in patients. The resulting 12 peptides were subjected to LCA

276   (Supplementary Table 1 and Supplementary Fig 1).

277

278   **LCA**

279   The fit indices of the LCA are summarized in Table 2. The Akaike information criteria

280   decreased continuously as classes increased; however, the BIC was lowest with the two-class

281   model, and the ssaBIC was lowest with the three-class model. The Lo–Mendell–Rubin

282   likelihood ratio value was only significant for the two-class model, whereas bootstrapped

283   likelihood ratio test was significant for all classes. Additionally, for models with three and four

284   classes, the proportion of the smallest group was less than 50 individuals. In conclusion, the

285   overall indices suggested that the two-class model was accepted as the final model. The latent

286   class classification and protein expression tertiles are presented in Supplementary Table 2.

287   Considering the differences in peptide levels, Class 1 was named the GELS-increased class,

288   and Class 2 was named the GELS-decreased class.

289   *(Insert Table 2)*

290   No significant differences were found between the latent classes when the demographic

291   characteristics were considered (Supplementary Table 3). Although significant differences

292   were not observed in the total scores of clinician rater scales, the negative symptom factor of

293   the BPRS was higher in the GELS-decreased latent class (t=−2.070, p=0.039). Considering the

294   individual items, the self-neglect and motor retardation items of the BPRS were increased in

295   the GELS-decreased latent class, and the excitement and motor hyperactivity items of the

296   BPRS were increased in the GELS-increased latent class (Table 3).

297   *(Insert Table 3)*

298   Next, when including HC, two latent classes were identified. The proportion of HC and various

299   total scores/factors/items of rating and subjective scales differed between classes. After

300   additional adjustment for HC in a linear regression model, the negative symptom factor of the

301   BPRS (t=-2.372, p=0.018), the emotional withdrawal and motor retardation item of the BPRS,

302   and the lassitude, inability to feel, and suicidal thoughts items of the Montgomery–Asberg

303   depression rating scale were increased in a latent class when compared with the other class

304   (Supplementary Tables 4 and 5).

305

306   **Bioinformatics analysis**

307   Among the eight proteins that differed between clusters (Supplementary Table 2), seven

308   [ASAP1 (ASAP1), TGFBI (BGH3), F5 (FA5), GSN (GELS), HIF1A (HIF1A), PDGFRB

309   (PGFRB), and TNXB (TENX) (represented by the gene symbol)] were included in the network

310   (network score=18), which comprised 35 molecules. Through criteria of determination for

311   diseases/functions and canonical pathways, the network was associated with a total of eight

312   diseases/functions, namely, dermatological diseases and conditions, inflammatory disease,

313   inflammatory response, immunological disease, neurological disease, nervous system

314   development and function, cellular assembly and organization, and cellular function and

315   maintenance (Supplementary Table 6). Furthermore, the network was related to eight canonical

316   pathways: HIF1α, integrin, CLEAR, FAK, erythropoietin, estrogen receptor, ferroptosis, and

317   ID1 signaling pathways (Fig 1).

318   *(Insert Fig 1)*

319

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, VWF were increased in patients, and FCGBP, HIF1A, PAFA, S10A2, and SEM6C were increased in HC. The most significant pathological protein was PFGRB.

343   Elevated levels of PFGRB in the cerebrospinal fluid is seen as a marker of blood-brain barrier

344   dysfunction (22). Interestingly, there is evidence that the levels of PGFRB in the cerebrospinal

345   fluid and serum are positively associated (23). As these psychiatric disorders have evidence of

346   a leaky blood-brain barrier (24), this could have been reflected in the plasma of the study

347   patients. The increased level of circulating VWF in patients, is in line with previous studies of

348   SCZ, BD, and MDD (25, 26). This implies that these psychiatric diseases might share a

349   common mechanism of endothelium-related inflammation. Interestingly, VWF itself is known

350   to influence blood brain barrier permeability (27). Another protein that has been reported in

351   multiple psychiatric disorders was HIF1A. This molecule was also a key protein in the network

352   analysis. HIF1A has been proposed to have a protective effect in depression (28). However, its

353   mRNA expression within peripheral white blood cells seems to be state-dependent in MDD

354   and BD (29). Although further studies are needed, this does imply that these diseases could be

355   commonly involved in pathways of energy metabolism and oxidative stress.

356   Recent studies have started to apply transdiagnostic approaches to cluster patients with

357   psychosis-affective disorders, considering biological correlates. Stein et al. (2021) revealed that

358   the negative syndrome in a transdiagnostic sample like ours is associated with the gray matter

359   volume of the bilateral frontal opercula, and that no association exists between the diagnosis

360   of the patients and the gray matter volume (30). However, the study clustered the patients based

361   on clinical data and analyzed the associations with gray matter volumes, showing multiple

362   associations between other clinical clusters and other brain region volumes (30), which differed

363   from our bottom-up approach using biological data to cluster patients. Another study by Chang

364   et al. (2021) subtyped transdiagnostic patients based on frontal-posterior functional imbalance

365   and found that the distribution of SCZ, BD, and MDD differs between clusters, but does not

366   differ in clinical symptoms (31). More studies should be conducted using various biological

367   correlates with a data-driven approach to cluster these patients and investigate their

368   associations with clinical traits and symptoms.

369   Consensus has been reached on negative symptom domains that include blunted affect,

370   anhedonia, alogia, avolition, and asociality, and two factors, namely, amotivation and

371   diminished expression (32). Our study revealed that the items including emotional withdrawal,

372   motor retardation, lassitude, and inability to feel were associated with latent classes when

373   analyzing patients, or patients with HC. Evaluations using scales specific for negative

374   symptoms are needed to confirm the associations between specific domains and factors. Even

375   though efforts have been expanded, the reports on the pathological mechanisms of negative

376   symptoms are inconsistent, since negative symptoms are also heterogeneous in its nature (32).

377   Most studies have been based on patients with psychosis, in which several studies have shown

378   significant associations between negative symptoms and inflammatory biomarkers (33).

379   The present study revealed that specific pathways, including systemic inflammation, hypoxia,

380   and signal transduction, were associated with latent classes. Estrogen receptor, erythropoietin

381   and integrin pathways have been proposed as significant pathophysiologies in previous studies

382   of SCZ, BD, and MDD, including a recent systematic review of peripheral blood proteomes

383   (34-39). However, as these pathways have associations with multiple psychiatric conditions, it

384   could be associated with a common psychiatric dimension. The present study proposes that

385   negative symptoms have a potential association with these pathways. Especially, the estrogen

386   receptor pathway has been proposed for its ameliorative role in negative symptoms of SCZ not

387   only due to sex differences in its severity and prognosis, but as hormonal replacement therapy

388   has a protective effect for negative symptoms in women (40). Additionally, a recent report of

389   single-cell level lymphocytes revealed that NF-kB p65, and Stat 3 cell signaling alterations

390   were shared between MDD and SCZ in a transdiagnostic sample, and suggested that they could

391   represent as a shared substrate for negative symptomology (41). Both proteins are known to be

392 associated with integrin pathways (42, 43). However, cautious interpretation of specific results

393 is necessary because the proteins were not from the CNS, and as these pathways all have

394 intracellular components. Even though the seven proteins of the network are known to have

395 secretory pathways or have potential to be secreted from intracellular to extracellular regions

396 (44-48), the link between plasma and the CNS is still in investigation. Nevertheless,

397 considering that these symptoms are resistant to treatment, investigating the proposed

398 mechanisms could expand our knowledge of their pathophysiology.

399 By contrast, the differential diagnosis of SCZ, BD, and MDD were not associated with latent

400 classes. This implies that without the consideration of negative symptoms, the conventional

401 differentiation between SCZ, BD, and MDD based on proteomics will have limitations, as it

402 does not reflect systemic biological manifestations. Therefore, phenotypes should consider not

403 only the ICD (International Statistical Classification of Diseases and Related Health Problems)

404 or DSM (Diagnostic and Statistical Manual of Mental Disorders)-based symptom checklists

405 but also the relationship between negative symptoms and associated biological correlates. This

406 will deepen our understanding of the pathophysiology of psychotic-affective disorders and

407 enable us to explain the heterogeneity within and the overlap between disorders.

408 However, the results also imply the obvious gap between circulating proteins and psychiatric

409 manifestations, including structurally, the blood-brain barrier. These gaps are probably why a

410 clearer separation in psychiatric manifestations between the latent classes were not seen, even

411 though we selected proteins that differentiated psychiatric diseases and HC, and controlled

412 significant covariates. There could be other biological measures that are better for biological

413 subtyping (5), that might reflect psychiatric symptoms with a stronger association.

414

415 **Strengths and limitations**

416   This study has the following limitations. First, it was a cross-sectional study; therefore,

417   causality could not be determined. A longitudinal study with multiple measurements of clinical

418   symptoms and plasma proteomes would enable us to investigate the preservation of latent

419   classes. Second, the proteins were obtained from the blood; therefore, functional analysis has

420   limitations. Although there is evidence of blood–brain barrier dysfunction in psychiatric

421   disorders, blood does not always reflect the CNS. Third, the differences between the scales

422   were small. The scales themselves might not be able to capture negative symptoms sensitively;

423   therefore, more detailed evaluations should be performed in the future. Especially the results

424   of the individual items need to be cautiously interpreted, as they are numerically very close.

425   Fourth, biological analysis was conducted using proteomics only. Integration with other omics

426   or other biological correlates could capture sophisticated mechanisms that would reveal more

427   accurate biological subtypes. Fifth, there were several preprocessing procedures to conduct

428   LCA, including stratification by sex, and decreasing the dimension of protein values from a

429   continuous to discrete variable (6), which could have resulted in loss of information, and affect

430   the latent classes. However, LCA has its advantages, as it is considered a more statistically

431   robust method than cluster analysis, since it is model-based, and generates fit statistics (49).

432   Finally, independent validation was not performed in this study.

433   Nevertheless, the strength of the study was that it was the first, to the knowledge of the authors,

434   to compare a transdiagnostic psychosis-affective disorder population with numerous proteomic

435   targets from the blood. Multiplexing proteins simultaneously enabled us to select significant

436   pathological proteins and cluster patients into two biological subgroups. Negative symptoms

437   need more attention because they vary between these subgroups, which tend to be neglected

438   and remain poorly understood. Further biological studies should use longitudinal designs with

439   detailed evaluation of negative symptoms to deepen our understanding of psychosis-affective

440   disorders.

441

Data availability

The quantitation information of 12 targets included in the latent classes were deposited to

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.

448

Author contributions

S.J.R (Sang Jin Rhee) and D.S.1 (Dongyoon Shin) contributed equally to this work and share

first authorship. S.J.R., D.S.1., D.S.2., Y.S., E.J.J., H.Y.J., S.R., S.H.L., H.K., M.B., Y.K.2

(Youngsoo Kim) and Y.M.A. contributed to the study concept and design. D.S.1, J.L., J.K.,

Y.K.1, Y.K.2. acquired the proteomic data. S.J.R., D.S.2, Y.S., E.J.J., H.Y.J., S.R., S.H.L., H.K.,

M.B., K.Y.L., S.H.K., M.K., J.S.K., K.H. and Y.M.A. obtained the clinical data. S.J.R. and

D.S.1 performed the statistical analysis. S.J.R, D.S.1, Y.S.2., Y.M.A. drafted this manuscript.

All authors critically reviewed the manuscript and approved the final version.

457

Acknowledgements

460

Funding

465

Conflict of Interest

466

467   Y.M.A receives research support from or serves as a speaker for Janssen Korea Ltd., Lundbeck

468   Korea Co. Ltd, and Korea Otsuka Pharmaceutical. All other authors have no conflict of interest

469   to declare.

470

471

472

473

474

475

476

477

478

479

480

481

References

482

483   1.   Chan CC, Shanahan M, Ospina LH, Larsen EM, Burdick KE. Premorbid adjustment

484   trajectories in schizophrenia and bipolar disorder: A transdiagnostic cluster analysis. Psychiatry

485   Res. 2019;272:655-62.

486   2.   Bowen EFW, Burgess JL, Granger R, Kleinman JE, Rhodes CH. DLPFC transcriptome

487   defines two molecular subtypes of schizophrenia. Transl Psychiatry. 2019;9(1):147.

488   3.   Haro JM, Ayuso-Mateos JL, Bitter I, Demotes-Mainard J, Leboyer M, Lewis SW, et

489   al. ROAMER: roadmap for mental health research in Europe. Int J Methods Psychiatr Res.

490   2014;23 Suppl 1(Suppl 1):1-14.

491   4.   Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain

492   criteria (RDoC): toward a new classification framework for research on mental disorders. Am

493   J Psychiatry. 2010;167(7):748-51.

494   5.   Beijers L, van Loo HM, Romeijn JW, Lamers F, Schoevers RA, Wardenaar KJ.

495   Investigating data-driven biological subtypes of sychiatric disorders using specification-curve

496   analysis. Psychol Med. 2020;52(6):1-12.

497   6.   Beijers L, Wardenaar KJ, Bosker FJ, Lamers F, van Grootheest G, de Boer MK, et al.

498   Biomarker-based subtyping of depression and anxiety disorders using Latent Class Analysis.

499   A NESDA study. Psychol Med. 2019;49(4):617-27.

500   7.   Shin D, Rhee SJ, Shin D, Joo EJ, Jung HY, Roh S, et al. Integrating proteomic and

501   clinical data to discriminate major psychiatric disorders: Applications for major depressive

502   disorder, bipolar disorder, and schizophrenia. Clin Transl Med. 2022;12(6):e929.

503   8.   World Health Organization. Global Recommendations on Physical Activity for Health:

504   World Health Organization; 2010.

505   9.   Hafkenscheid A. Psychometric evaluation of a standardized and expanded Brief

Psychiatric Rating Scale. Acta Psychiatr Scand. 1991;84(3):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(1):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(3):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(6):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(5):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

531   therapy. Arch Gen Psychiatry. 1984;41(1):72-80.

532   20.   Čuklina J, Lee CH, Williams EG, Sajic T, Collins BC, Rodríguez Martínez M, et al.

533   Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. Mol Syst

534   Biol. 2021;17(8):e10240.

535   21.   Preece RL, Han SYS, Bahn S. Proteomic approaches to identify blood-based

536   biomarkers for depression and bipolar disorders. Expert Rev Proteomics. 2018;15(4):325-40.

537   22.   Montagne A, Barnes SR, Sweeney MD, Halliday MR, Sagare AP, Zhao Z, et al. Blood-

538   brain barrier breakdown in the aging human hippocampus. Neuron. 2015;85(2):296-302.

539   23.   Miners JS, Kehoe PG, Love S, Zetterberg H, Blennow K. CSF evidence of pericyte

540   damage in Alzheimer's disease is associated with markers of blood-brain barrier dysfunction

541   and disease pathology. Alzheimers Res Ther. 2019;11(1):81.

542   24.   Futtrup J, Margolinsky R, Benros ME, Moos T, Routhe LJ, Rungby J, et al. Blood-

543   brain barrier pathology in patients with severe mental disorders: a systematic review and meta-

544   analysis of biomarkers in case-control studies. Brain Behav Immun Health. 2020;6:100102.

545   25.   Hope S, Melle I, Aukrust P, Steen NE, Birkenaes AB, Lorentzen S, et al. Similar

546   immune profile in bipolar disorder and schizophrenia: selective increase in soluble tumor

547   necrosis factor receptor I and von Willebrand factor. Bipolar Disord. 2009;11(7):726-34.

548   26.   Lopez-Vilchez I, Diaz-Ricart M, Navarro V, Torramade S, Zamorano-Leon J, Lopez-

549   Farre A, et al. Endothelial damage in major depression patients is modulated by SSRI treatment,

550   as demonstrated by circulating biomarkers and an in vitro cell model. Transl Psychiatry.

551   2016;6(9):e886.

552   27.   Suidan GL, Brill A, De Meyer SF, Voorhees JR, Cifuni SM, Cabral JE, et al.

553   Endothelial Von Willebrand factor promotes blood-brain barrier flexibility and provides

554   protection from hypoxia and seizures in mice. Arterioscler Thromb Vasc Biol.

555   2013;33(9):2112-20.

556   28.   Kang I, Kondo D, Kim J, Lyoo IK, Yurgelun-Todd D, Hwang J, et al. Elevating the

557   level of hypoxia inducible factor may be a new potential target for the treatment of depression.

558   Med Hypotheses. 2021;146:110398.

559   29.   Shibata T, Yamagata H, Uchida S, Otsuki K, Hobara T, Higuchi F, et al. The alteration

560   of hypoxia inducible factor-1 (HIF-1) and its target genes in mood disorder patients. Prog

561   Neuropsychopharmacol Biol Psychiatry. 2013;43:222-9.

562   30.   Stein F, Meller T, Brosch K, Schmitt S, Ringwald K, Pfarr JK, et al.

563   Psychopathological Syndromes Across Affective and Psychotic Disorders Correlate With Gray

564   Matter Volumes. Schizophr Bull. 2021;47(6):1740-50.

565   31.   Chang M, Womer FY, Gong X, Chen X, Tang L, Feng R, et al. Identifying and

566   validating subtypes within major psychiatric disorders based on frontal-posterior functional

567   imbalance via deep learning. Mol Psychiatry. 2021;26(7):2991-3002.

568   32.   van der Meer L, Kaiser S, Castelein S. Negative symptoms in schizophrenia:

569   reconsidering evidence and focus in clinical trials. Br J Psychiatry. 2021;219(1):359-60.

570   33.   Dunleavy C, Elsworthy RJ, Upthegrove R, Wood SJ, Aldred S. Inflammation in first-

571   episode psychosis: The contribution of inflammatory biomarkers to the emergence of negative

572   symptoms, a systematic review and meta-analysis. Acta Psychiatr Scand. 2022;146(1):6-20.

573   34.   Hwang WJ, Lee TY, Kim NS, Kwon JS. The Role of Estrogen Receptors and Their

574   Signaling across Psychiatric Disorders. Int J Mol Sci. 2020;22(1):373.

575   35.   Jaudon F, Thalhammer A, Cingolani LA. Integrin adhesion in brain assembly: From

576   molecular structure to neuropsychiatric disorders. Eur J Neurosci. 2021;53(12):3831-50.

577   36.   Fan Y, Abrahamsen G, Mills R, Calderon CC, Tee JY, Leyton L, et al. Focal adhesion

578   dynamics are altered in schizophrenia. Biol Psychiatry. 2013;74(6):418-26.

579   37.   Miskowiak KW, Vinberg M, Harmer CJ, Ehrenreich H, Knudsen GM, Macoveanu J,

580   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(1):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 Suppl 2:83-96.

41.   Al-Kuraishy HM, Al-Gareeb AI, Waheed HJ, Al-Maiahy 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(3):80-6.

42.   Scatena M, Almeida M, Chaisson ML, Fausto N, Nicosia RF, Giachelli CM. NF-κB mediates αvβ3 integrin-induced endothelial cell survival. J Cell Biol. 1998;141(4):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(1):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(4):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(Database issue):D959-65.

46.   Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak

605   S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat

606   Biotechnol. 2019;37(4):420-3.

607   47.   Bendtsen JD, Jensen LJ, Blom N, Von Heijne G, Brunak S. Feature-based prediction

608   of non-classical and leaderless protein secretion. Protein Eng Des Sel. 2004;17(4):349-56.

609   48.   Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane

610   protein topology with a hidden Markov model: application to complete genomes. J Mol Biol.

611   2001;305(3):567-80.

612   49.   Sinha P, Calfee CS, Delucchi KL. Practitioner's Guide to Latent Class Analysis:

613   Methodological Considerations and Common Pitfalls. Crit Care Med. 2021;49(1):e63-e79.

614

615

616

617

618

619

620

621

622

623

625   **Figure 1. Protein network and associated canonical pathways generated by IPA for 8**

626   **proteins differentiating latent classes.** Seven of the 8 proteins were included in the protein

627   network. Direct and indirect interactions are represented by solid and dashed lines. Proteins are

628   represented by gene symbol with protein entry name in parentheses. Shapes notify the

629   molecular classes of proteins defined in the legend. Colored nodes notify the seven proteins,

630   and white nodes represent predicted molecules. Canonical pathways involved in proteins in the

631   network are represented by dotted lines. Canonical pathways in the red box are related to hub

632   protein (HIF1A). Differences in protein expression levels of the seven features between the

633   latent classes are represented by -$\log_{10}$ (expression P-value). IPA, ingenuity pathway analysis;

634   CP, canonical pathway.
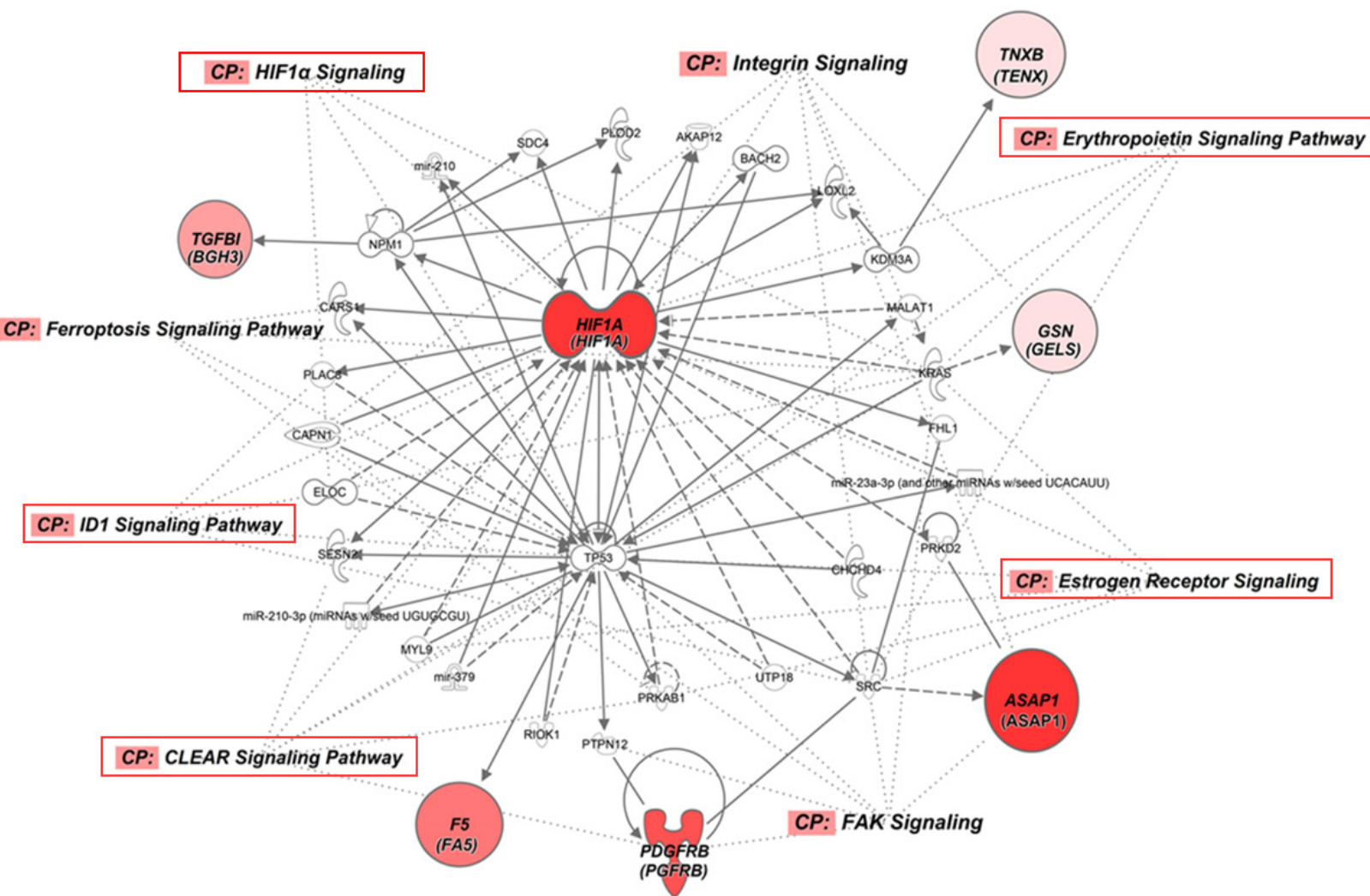
Network Information and legends

A → B
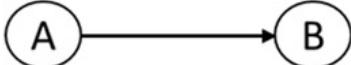
Activation, causation, expression, membership, modification, molecular cleavage, phosphorylation, protein-DNA interactions, protein-RNA interactions, and regulation of binding transcription

Enzyme   Peptidase   Transcription Regulator   Mature micro RNA

Kinase   Phosphatase   Transporter   Micro RNA   Other

——— Direct Interaction
- - - Indirect Interaction
········· Interaction for proteins and their corresponding CP

Colored node: Seven plasma proteins
Colorless node: Predicted molecules

Class 1 vs Class 2

4 ────────── 43

-log₁₀ (Expression P-value)
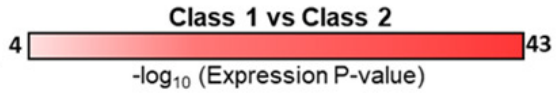
CP: **Canonical pathways associated with HIF1A**

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

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

Abbreviations : SD = standard distribution, BMI = body mass index

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

[a]Boldface values are statistically significant at p<0.05

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

Model	AIC	BIC	saBIC	Entropy	LMR-LR p-value	BLRT p-value	1
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%)
3 class	13155.09	13467.56	13232.67	0.750	0.0690	< 0.0001	273 (54.2%)
4 class	13133.93	13551.97	13237.73	0.800	0.7630	0.0200	235 (46.6%)

Abbreviations : AIC = Akaike information crtieria, BIC = Baysian information criterion, ssɛ
Baysian information criterion, LMR-LR = Lo-Mendell-Rubin likelihood ratio, BLRT = Boo
= not applicable.

N (proportion %)		
2	3	4
239 (47.4%)		
215 (42.7%)	16 (3.2%)	
216 (42.9%)	31 (6.2%)	22 (4.4%)

aBIC = sample-size adjusted
tstrapped likelihood ratio test, N/A

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[a]
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
HAMA, mean (SD)	11.18 (7.05)	11.26 (7.19)	t = -0.133	0.90
Significant subscales/items				
BPRS negative symptoms factor, mean (!	4.24 (2.13)	4.65 (2.28)	t = -2.070	**0.039**
BPRS13, mean (SD) (self-neglect)	1.15 (0.52)	1.26 (0.66)	t = -2.235	**0.026**
BPRS18, mean (SD) (motor retardation)	1.36 (0.80)	1.52 (0.92)	t = -2.157	**0.031**
BPRS21, mean (SD) (excitement)	1.32 (0.81)	1.17 (0.52)	t = 2.418	**0.016**
BPRS23, mean (SD) (motor hyperactivit	1.15 (0.44)	1.08 (0.30)	t = 1.981	**0.048**

Abbreviations : BPRS = Brief Psychiatric Rating Scale, YMRS = Young Mania Rating Scale, MADRS = Montgomery-Asberg Depression Rating Scale, HAM-A = Hamilton Anxiety Scale

Statistical analysis based on *t* -tests

[a]Boldface values are statistically significant at p<0.05