creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

의학박사 학위논문

“주요정신질환 고위험군의 진단, 중증도, 경과와 관련된 임상 변인과 혈액 단백체 분석 연구”

2022년 2월

서울대학교 대학원

임상의과학과 임상의과학 전공

이 현 주

“주요정신질환 고위험군의 진단, 중증도, 경과와 관련된 임상 변인과 혈액 단백체 분석 연구”

지도 교수 안 용 민

이 논문을 의학박사 학위논문으로 제출함
2021년 10월

서울대학교 대학원
임상의과학과 임상의과학 전공
이 현 주

이현주의 의학박사 학위논문을 인준함
2022년 1월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

초 록

서론: 주요정신질환의 조기진단과 치료는 젊은 연령의 고위험군 환자들의 경과 호전 및 삶의 질 향상에 매우 중요하다. 기존진단분류체계로는 증상이 비특이적인 고위험군 환자를 진단하는 것에 한계가 있으며, 고위험군의 질병 상태를 반영할 수 있는 생물학적 표지자의 발굴이 필요하다. 이에 본 연구에서는 주요정신질환의 고위험군을 대상으로 혈액 단백체를 정량 분석하여, 고위험군 상태와 질병 경과를 반영하는 단백체 마커를 다양한 방법으로 탐색하고자 하였다.

방법: 서울지역 7개의 의료기관으로부터 의뢰된 90명의 고위험군을 모집하였고, 임상평가와 채혈을 같은 날에 시행하였다. 혈중 단백질은 158개의 표적 펩타이드를 사용하여 다중 검지법으로 정량화하였다. 분석된 고위험군의 혈중 단백질 발현량을 환자군과 정상 대조군의 단백질 발현량과 비교하였고 이를 통해 고위험군 상태를 반영하는 단백체 마커를 탐색하였다(연구 1), 또한, 잠재프로파일 분석을 통해 고위험군의 이질적인 임상특성을 유형화하고, 증상 중증도 반영하는 단백체를 탐색하기 위해 상관분석을 수행하였다(연구 2). 마지막으로, 초기 평가 시 수집된 혈액 단백질 중 주요정신질환의 경과를 예측하는 임상, 단백체 마커를 탐색하기 위하여 LASSO 회귀분석을 통해 통계적으로 의미 있는 변수를 선별하였다. 결과 변수로 주요정신질환으로의 발병(Model 1), 임상 경과의 호전(Model 2)으로 두 개의 모델을 구성하였으며, 각 모델에서 3개의 하위 모델(Model a -임상 변인 모델, Model b-단백체 변인 모델, Model c-혼합 모델)로 나누어, 선별된 변수들이 포함된 모델의 성능을 AUROC로 평가하였다 (연구3).

결과: 연구 1을 통해 고위험 상태를 반영하는 27종류의 단백체 마커를

확인하였고, 이 중 대부분의 마커들이 선행 환자군 연구에서 보고된 마커임 확인하였다. 연구 2에서는 잠재프로파일 분석을 통하여, 3개의 하위군으로 고위험군을 분류하였으며, 증상으로 유형화된 분류는 연구 1의 범주와 다름을 확인하였다. 또한, 19개의 임상지표 중 2개 이상의 지표와 유의한 상관관계를 보인 단백체 마커 29개를 선별할 수 있었다. 연구3에서는 고위험군의 발병 여부를 예측하는 임상 변인(Model 1a 9개, Model 1c 5개) 및 단백체(Model 1b 15개, Model 1c 2개) 변인을 선별하였고, 선별된 단백체 변인으로 분석한 모델의 성능을 평가하였다. 혼합 모형인 Model 1c의 성능이 테스트 데이터에서 안정적으로 우수하였다(AUC 0.90-0.97). 반면 고위험군의 증상 호전을 예측하는 임상 변인(Model 2a 8개, Model 2c 9개), 단백체 변인(Model 2b 8개, Model 2c 5개)을 선별하고, 예측모델의 성능은 테스트 데이터에서 Model 2c AUC 0.70-0.75로 가장 높았다. 연구 1-3을 통해 발굴된 단백체 마커의 대부분은 단백질들이 선행 환자군 대상연구에서 보고된 것이었으며, 관련된 생물학적 경로로는 혈액 응고과정, 면역반응, 지질대사, 항산화 반응 경로였다.

결론: 본 연구를 통해 주요정신질환의 발병 전 고위험군 상태에서도 환자군에서 보고되었던 의미 있는 생물학적 변이가 관찰됨을 확인할 수 있었다. 뿐만 아니라, 고위험군의 증상 중증도를 반영하고, 고위험군의 발병 여부 및 증상 호전경과를 예측할 수 있는 공통된 단백체 마커를 확인할 수 있었다. 그러나 발굴된 단백체 마커의 정확도 안정성의 확보와, 예측 모델의 정확도를 향상시키기 위해서 향후 단백체 마커의 종단적 변화량을 평가하는 것이 필요 할 것이다. 그럼에도 본 연구에서 발굴된 단백질 및 임상지표는 향후 주요정신질환 고위험군의 상태를 생물학적으로 조기진단하고, 질병의 경과를 예측하는 유용한 마커로 활용될 수 있을 것으로 전망한다.

*본 내용의 일부는 아래 논문 출판 완료된 내용 및 관련된 내용을 일부 포함하고 있음.

1. Lee TY, Lee J, Lee HJ, Lee Y, Rhee SJ, Park DY, et al. Study Protocol for a Prospective Longitudinal Cohort Study to Identify Proteomic Predictors of Pluripotent Risk for Mental Illness: The Seoul Pluripotent Risk for Mental Illness Study. Front Psychiatry. 2020;11:340.
주요정신질환 고위험군의 선별 프로토콜 내용을 포함함.

2. Lee H, Rhee SJ, Kim J, Lee Y, Kim H, Lee J, et al. Predictive protein markers for depression severity in mood disorders: A preliminary trans-diagnostic approach study. J Psychiatr Res. 2021;142:63-72.
본 저자는 기분장애 환자의 우울 증상 중증도를 예측하는 혈장 단백체 마커 (LC-MS/MS 방법으로 분석)를 보고한 위 논문의 1저자로 논문을 출판한 바 있음.

주요어 : 주요정신질환, 고위험군, 발병, 경과, 단백체, 예측모델
학 번 : 2020-32637

목 차

초록 i

표 목차 vi

그림 목차 ix

약어 목록 xi

제 1 장 서 론 1

제 2 장 연구 1 7

제 1 절 연구대상 및 방법 8

제 2 절 결과 18

제 3 절 고찰 24

제 3 장 연구 2 28

제 1 절 연구대상 및 방법 29

제 2 절 결과 39

제 3 절 고찰 43

제 4 장 연구 3 47

제 1 절 연구대상 및 방법 48

제 2 절 결과 54

제 3 절 고찰 63

제 5 장 종합요약 66

참고문헌 127

Abstract 144

표 목차

[표 1-1]Inclusion and exclusion criteria of clinical high risk group.....68
[표 1-2]Demographic and clinical characteristics of the study subjects related to HC, CHR-BD, and BD 70
[표 1-3]Classification of subgroup of CHR-BD according to BARS criteria.....71
[표 1-4]Differentially expressed peptides(ANOVA) identified by LC-MRM analysis between HC, CHR-BP, and BD.....72
[표 1-5]Differentially expressed peptides(ANCOVA) identified by LC-MRM analysis between HC, CHR-BD, and BD 74
[표 1-6]Demographic and clinical characteristics of the HC, CHR-PRS and FEP 76
[표 1-7]Differentially expressed peptides(ANOVA) identified by LC-MRM analysis between HC, CHR-PRS, and FEP.....77
[표 1-8]Differentially expressed peptides(ANCOVA) identified by LC-MRM analysis between HC, CHR-PRS, and FEP.....78
[표 1-9]Comparison of baseline demographics and clinical characteristics between CHR-PRS vs. CHR-BD.....79
[표 1-10]Differentially expressed peptides identified by LC-MRM analysis between CHR-PRS and CHR-BD 80
[표 2-1]Fit indices of LPA 81
[표 2-2]Comparison of baseline demographics between LS,

DPD, and MD group.....82
[표 2-3]Comparison of clinical characteristics between Low LS, DPD, and MD group.....83
[표 2-4]Differentially expressed peptides(ANOVA) identified by LC-MRM analysis between LS, DPD, and MD 85
[표 2-5]Pearson's correlation coefficients(r) between clinical variables and 29 peptides.....88
[표 3-1]List of baseline 110 clinical features included in models 89
[표 3-2]List of baseline 158 proteomic features included in models.....90
[표 3-3]Summary of prediction model. 91
[표 3-4]Comparison of baseline demographical characteristics between CHR-T vs.CHR-NT 92
[표 3-5]Comparison of baseline symptom characteristics between CHR-T vs. CHR-NT 93
[표 3-6]Comparison of baseline proteomic data between CHR-T vs. CHR-NT 95
[표 3-7A]Unique models of Model 1a based on the combinations of features selected.....96
[표 3-7B]Unique models of Model 1b based on the combinations of features selected.....97
[표 3-7C]Unique models of Model 1c based on the combinations

of features selected.....98
[표 3-8]Features included in prediction models(Model 1) ..99
[표 3-9]Comparison of baseline demographical characteristics between Recovered vs. Not-recovered group.....100
[표 3-10]Comparison of baseline symptom characteristics between Recovered vs. Not-recovered group.....101
[표 3-11]Comparison of baseline proteomic data between Recovered vs. Not-recovered group 104
[표 3-12A]Unique models of Model 2a based on the combinations of features selected.....105
[표 3-12B]Unique models of Model 2b based on the combinations of features selected.....106
[표 3-12C]Unique models of Model 2c based on the combinations of features selected.....108
[표 3-13]Features included in prediction models(Model 2)110

그림 목차

[그림 1-1]Determination of quantifiable targets for CHR of

psychiatric disorders 111
[그림 1-2]Summary of DEPs in CHR-BD and CHR-PRS112
[그림 1-3]Correlation plot showing Pearson's correlation analysis between clinical variable(PANSS) and DEPs.....113
[그림 1-4]Correlation plot showing Pearson's correlation analysis between clinical variable(HAM-D, HAM-A) and DEPs 114
[그림 1-5]Correlation plot showing Pearson's correlation analysis between clinical variable(A-YMRS, B-GAF, CGI) and DEPs 115
[그림 1-6]Bioinformatics analysis of the proteins that differentially expressed in CHR 116
[그림 2-1]Mean and SD of clinical variables between Class 1 (Low severity group), Class 2 (Depressive-psychosis dominant group), and Class 3 (Manic-dominant group).....117
[그림 2-2]Correlation plot showing Pearson's correlation analysis between 19 clinical variable and peptides.....118
[그림 2-3]Bioinformatics analysis of the proteins significantly reflect severity of symptoms.....119
[그림 3-1]Volcano plot of p-values for DEPs between CHR-T and CHR-NT 120
[그림 3-2]Feature selection among 100 models analyzed from repeated application of LASSO regression with 10-fold CV on

the training data(Dependent variable : Transition)..... 121
[그림 3-3]ROC curves showing model performance in predicting the probability of Transition(CHR-T).....122
[그림 3-4]Volcano plot of p-values for DEPs between Recovered and Not-recovered group 123
[그림 3-5]Feature selection among 100 models analyzed from repeated application of LASSO regression with 10-fold CV on the training data(Dependent variable: Symptom recover) ..124
[그림 3-6]ROC curves showing model performance in predicting the probability of disease recover.....125
[그림 4] Summary of the results of study 1-3.....126

약어목록

AIC: Akaike information criterion

ANCOVA: Analysis of covariance

ANOVA: Analysis of variance

AUC: Area under the curve

AUROC: Area under the receiver operating characteristics

BARS: Bipolar at-risk criteria

BD: Bipolar disorder

BIC: Bayesian information criteria

BIS: Barratt Impulsiveness Scale

BLRT: Bootstrap likelihood ratio test

BMI: Body mass index

BPSS-AP: Bipolar prodrome symptom scale-abbreviated prospective

BRIAN: Biological rhythms interview of assessment in neuropsychiatry

CD-RISC: Connor-Davidson resilience scale

CGI-S: Clinical global impression-severity scale

CHR: Clinical high risk state

CHR-BD: Clinical high risk group –Bipolar disorder

CHR-MDD: Clinical high risk group – Major depressive disorder

CHR-PL: Clinical high risk group – Pluripotent

CHR-PRS: Clinical high risk group – Psychosis risk syndromes

CI: Confidence interval

CSM: Composite scale of morningness

CTQ-SF: Childhood trauma questionnaire-short form

CV: Cross validation

DAVID: Database for annotation, visualization and integrated discovery

DEP: Differentially expressed protein

DSM-5: Diagnostic and statistical manual of mental disorders, 5th edition

EDTA: Ethylenediaminetetraacetic acid

FDR: False discovery rate

FEP: First episode psychosis

GAF: Global assessment of functioning

GF: Global functioning: social and role scales

GO: Gene ontology

GSS: Global seasonality score

HAM-A: Hamilton rating scale for anxiety

HAM-D: Hamilton depression rating scale

HSD: Honestly significant difference

LASSO: Least absolute shirinkage and selection operator

LC: Liquid-chromatography;

LC-MS/MS: Liquid-chromatography tandem mass spectrometry

LMR: Lo-Mendell-Rubin likelihood adjusted test

LPA: Latent profile analysis

MDD: Major depressive disorder

MRM/MS: Multiple reaction monitoring/mass spectrometry

PANSS: Positive and negative syndrome scale

PQ-B: Prodromal questionnaire-brief

QIDS: Quick inventory of depressive symptomatology

ROC: Receiver-operating characteristic

saBIC: sample adjusted Bayesian information criteria

SAD: Seasonal affective disorder

SCID: Structured clinical interview for the DSM-IV-TR axis I disorder

SD: Standard deviation

SE: Standard error

SF: Selection fraction

SIPS: The structural interview of psychosis risk syndromes

SOPS: Scale of prodromal symptoms

SPAQ: Seasonal pattern assessment questionnaire

STAI-S: State-trait anxiety inventory - state

STAI-T: State-trait anxiety inventory - trait

WHOQOL-BREF: WHO quality of life-BREF

YMRS: Young mania rating scale

제 1 장 서 론

2018년 발표된 한국인의 정신질환 질병 부담 조사연구에 따르면, 정신질환 질병 부담(장애 보정 생존연수, DALY, Disability adjusted life years)은 전 질환 질병 부담의 6.4%를 차지하였다(1). 특히 연령대별 질병 부담의 순위를 살펴보았을 때 정신 및 행동 장애로 인한 질병 부담의 경우 10~19세의 경우 2위, 20~39세의 경우 3위로 젊은 연령에서 높은 비중을 차지하였다. 또한, 2015년 정신 및 행동 장애와 관련된 사회경제적 비용 규모는 약 7조 2천억 원이었으며, 경제적 질병 부담 규모가 전체적으로 증가하는 경향을 보였다(1, 2). 2000년 WHO에서 발표한 World Health Report에서는 15세에서 44세 사이에 발생하는 장해의 질병 부담을 평가하였는데, "장해로 상실한 삶의 년수(Years Lived with Disability; YLD)" 에서는 교통사고, 관절염, 결핵 등을 제치고 주요우울장애가 1위, 조현병이 3위, 양극성장애가 5위에 해당하였다(3). 젊은 연령에서 발병한 주요정신질환은, 발병 후 다양한 사회적, 기능적 측면의 손상을 야기하며, 특히 학업능력의 저하, 학교적응의 문제, 대인관계 갈등의 증가, 자살 등 만성적인 경과를 보일 가능성이 높다(3). 따라서 주요정신질환의 조기진단과 치료가 젊은 연령의 고위험군 환자들의 경과 호전 및 삶의 질 향상에 매우 중요하고 필수적임을 알 수 있다.

기존의 주요정신질환 진단은 현상학적 진단 도구인 Diagnostic and statistical manual of mental disorders, 5th edition(DSM-5)의 기준을 따라 왔으며, 이로 인하여 증상이 완전히 발현된 뒤에 주요정신질환으로 진단하였다. 그러나 증상이 완전히 발현되기 전 대상자를 선별하고 치료적 개입 여부를 결정하는 것이 매우 중요하므로, DSM-5의

진단기준을 적용하는 것은 한계점이 있다. 주요정신질환의 경우 증상이 완전히 발현하기 전인 전구기, 즉 고위험 상태(Clinical high risk state, CHR)에 있는 환자들은 흔히 우울감, 기분 변화, 불면과 같은 비특이적인 증상을 호소하므로, DSM-5의 기존 진단체계로는 진단하는 것에 한계가 있다(4). 또한 최신 연구결과에 따르면 임상적으로 추후 조현병이 발병할 것이라고 예상한 정신증 고위험군 집단에서도 실제로는 초기 정신병적 증상 외에 우울감, 기분 변화 등 이외의 증상이 다양하게 나타나며, 시간이 지나도 정신증 외에 우울증, 양극성장애, 불안장애 등 다양한 질병이 발병하는 것으로 보고되었다(5, 6). 또한, 양극성장애로 발병 전 단계로 평가하고 추적관찰 한 대상자 중에서도, 정신병적 증상을 동반하는 양극성장애가 발병하는 빈도가 높고, 이 경우 경과가 장기화 될수록 조현병 등 정신증으로 재진단될 수 있음이 보고되었다(5).

생물학적 마커 연구들 또한, 생물학적 마커가 특정 질환에 국한되는 것이 아니라, 다양한 질환에 걸쳐 공통적으로 발견될 수 있음을 시사한다(7-10). 이러한 DSM-5 진단의 한계를 극복하고, 고위험군 상태의 대상자를 포괄하고자, 새로운 가설로, 임상 단계모델(Clinical staging model) 이론이 최근 주목받고 있다(11). 임상단계모델은 정신질환의 구별된 진단보다는, 증상의 심각도에 따라 개인을 분류하는 이론이다. 이 이론에 따르면 정신질환의 조기발견을 위해 역치하 증상을 호소하는 개인을 선별하는 동시에, 질병의 진행 정도(disease progression)를 반영하는 생물학적 마커의 필요성을 강조한다. 따라서 고위험군 상태를 반영할 수 있는 생물학적 표지자의 발굴은 고위험군의 상태를 진단하고, 치료적 개입이 필요한 시기를 결정함에 많은 도움을 줄 수 있을 것이다.

현재까지 다양한 지표들을 이용한 생물학적 연구가 정신과 영역에서 진 보고되었다(12-15). 그러나 조현병 환자들을 대상으로 하였을 때 모든 마커를(임상적 특징, 뇌파, 뇌영상, 혈액검사) 조합해서 분석해야 조현병 발병의 양성 예측도가 98%로 가장 높다는 보고가 있으며, 현재까지 명확하게 충분한 민감도와 특이도, 정확도를 만족하는 생물학적 마커는 드물다(16). 또한, 뇌의 기능과 이상을 반영하는 생물학적 이상을 발견하기 위하여 주로 신경 인지학적, 뇌파학적, 뇌영상학적 도구들이 현재까지 주로 사용되었다(17, 18). 그러나 뇌영상학 방법은 임상 환경에서 쉽게 접근하기가 어려우며, 비용이 높은 단점이 있다. 비정신과의 70% 이상의 진단은 접근성, 단순성 및 저렴한 비용의 이유로, 혈액검사를 사용하여 이루어지고 있다(8, 19, 20). 최근 연구에 따르면, 정신질환과 관련된 뇌 특정 부위에서의 변이 또한, 혈액에서 검출될 수 있으며 있다(21). 특히 신경아세포 및 신경세포에서 분비되는 단백질이 혈액에서 검출될 수 있으며, 이는 중추신경계의 이상을 말초 혈액에서 확인할 수 있음을 시사한다(21). 따라서 정신과 질환의 평가 및 진단을 위해 사용될 수 있는 혈액 마커를 발굴하는 것은 매우 필요하며, 향후 실용적으로 활용될 가능성이 매우 높을 것이다(22). 따라서 본 연구에서는 혈액의 단백체 마커를 사용하여 주요정신질환 고위험군의 진단, 증상의 심각도를 반영하는 생물학적 마커를 발굴하고자 하였다. 또한, 혈액 단백체 정량 방법으로 질량분석기반 다중 검지법(Multiple reaction monitoring-mass spectrometry, MRM-MS)을 사용하였다. MRM-MS 분석법은 극미량의 시료로 100~300여 개의 단백질 표지자를 한 번에 정량 분석할 수 있는 기술로, 대규모 단백질체에 대한 절대 정량 분석이 가능한 분석방법이다(23). 다양한 단백체 마커들이 정신과 질환에서

보고되고 있으므로, 이러한 변이들을 동시에 높은 정확도로 측정할 수 있는 본 분석법을 적용하였다.

본 연구를 위해 가장 먼저 선행연구들을 통하여 임상적 주요정신질환 고위험군의 모집기준을 마련하였다(24). 최근 보고된 정신증에 국한된 고위험군 코호트 연구결과에 따르면, 정신증에 국한된 경우 발병률이 예상보다 낮으며, 이러한 결과의 한계의 한 이유로 초기 엄격한 등록기준이 문제일 수 있음을 지적하였다(25, 26). 또한, 초기 증상은 질병의 경과에 따라 변화할 수 있고 예상한 진단명을 빗나갈 수 있음이 선행연구를 통해 보고된 바 있다. 이러한 고위험군 연구의 한계를 극복하기 위하여 최근 McGorry 외 연구진은 트랜스진단 코호트 접근법(CHARMS approach)을 제안하였다(27). CHARMS는 발병의 결과를 정신증으로 국한하지 않고 양극성장애, 우울증 등 광범위한 정신질환으로 발병의 결과를 확대한다. 본 연구에서는 CHARMS 접근법에 기반하여 가능한 다양한 증상영역을 포괄하는 주요정신질환 고위험군을 모집하고자 하였다. 이를 위해 주요정신 고위험군의 모집 범위를 정신증 고위험군(Clinical high risk state-Psychosis risk syndrome, CHR-PRS), 양극성장애 고위험군(CHR-Bipolar disorder, CHR-BD), 주요우울장애 고위험군(CHR-Major depressive disorder)으로 포괄하였다. 주요정신질환 전구기 증후군을 진단하는데 사용되는 구조화된 진단 인터뷰 도구 SIPS(the structural interview of psychosis risk syndromes)를 사용하고, SOPS(Scale of prodromal symptoms)기준을 사용하여 CHR-PRS 대상자를 모집하였다(28, 29). CHR-BD의 경우 SIPS의 구조와 매우 유사하게 개발된 BPSS-AP(Bipolar prodrome symptom scale-abbreviated prospective)와 BARS 기준(Bipolar at risk state criteria)을 사용하였다(30, 31).

주요우울장애 고위험군은 DSM-5의 우울 삽화 진단기준의 문항을 모두 충족하지 못하거나, 1주 이상-2주 미만의 기간 동안 우울한 기분이 지속하는 경우로 정의하였다. 뿐만 아니라, 위 세 진단군에 포함되지 않는 대상자를 다능성 위험군(CHR-Pluripotent, CHR-PL)로 정의하였다. 다능성 위험군이란 고위험군 상태의 대상자가 발병 전 초기 단계는 여러 질병으로 진행할 가능성을 포괄한다는 의미로 선행연구에서 먼저 제안된 바 있다(32). 마지막으로 확장군(CHR-Extended)은 25세 미만이며, 최근 5년 이내에 제 2형 양극성장애, 기분순환장애, 재발성 우울장애를 진단받은 대상자로, 기진단 된 진단명에서 다른 진단분류로 재진단 될 수 있는 가능성을 확장 및 포괄하였다. 본 연구의 고위험군 대상자 모집기준은 선행 프로토콜 논문을 통하여 발표되었다(24).

본 연구는 주요정신질환 고위험군의 진단, 중증도와 관련된 혈액 단백체 마커를 발굴하고, 경과를 예측 모델을 발견하고자 하였다.

연구 1에서는 고위험군 상태에서 특징적으로 관찰되는 단백체 마커에 대해 탐색하기 위해 고위험군의 혈장 단백질 발현량을 정상 대조군(Healthy control, HC)과 기진단 된 환자군의 혈액 단백체 발현량과 비교 하였다. 선행 프로토콜(24)의 모집기준에서 CHR-PRS군으로 분류된 피험자들의 혈장 단백질 발현량을 초기 정신증 환자군(First episode psychosis, FEP) 및 HC군과 비교하였으며, CHR-BD로 분류된 피험자들의 혈장 단백질 발현량을 양극성장애 환자군(BD, Bipolar disorder) 및 HC군과 비교하였다.

연구 2에서는 연구 1 및 선행논문(24)에서 적용한 고위험군의 분류 기준의 한계를 극복하기 위해 고위험군의 이질적인 증상 도메인을 종합하고, 이질적인 특성을 증상에 기반하여 유형화하고자 하였다. 이를 위하여 잠재프로파일 분석(Latent profile analysis, LPA)을 수행하고,

잠재계층을 분류 하였다. 또한, 증상의 심각도를 반영하는 임상척도와 단백질 발현량 사이의 상관분석을 통하여, 질환의 중증도를 반영하는 단백체 마커를 탐색하였다.

마지막으로, 초기에 측정된 단백질 중 주요정신질환의 경과를 예측하는 단백체 마커를 탐색하였다. 이를 위해 주요정신질환으로의 이환 여부 및 임상 경과 호전 여부를 결과 변수로 통계적 분석을 시행하였다.

이를 통해 고위험군 상태를 반영하는 혈액 단백체 마커를 구축하고 향후 종적 단백체 연구의 기반자료로 활용될 것을 기대하였다.

제 2 장 – 연구 1. 주요정신질환 환자군 -고위험군-정상 대조군의 혈장 단백체 발현량 비교를 통한 고위험 상태 반영 단백체 마커 탐색

주요정신질환은 질병의 진행이 만성화됨에 따라 기능 장애 또는 돌이킬 수 없는 구조적 변화의 위험을 높인다. 따라서 대부분의 정신과 의사들은 조기 개입이 환자의 질병 진행을 막는 데 도움이 될 수 있다는 데 동의한다. 심리사회적 개입 및 약물치료는 현재 정신증 및 양극성장애의 위험이 높은 개인에게 권장된다. 그러나 고위험 상태에서 정신 약리학적 개입이 필요한지는 정신과 전문의 사이의 이견이 있다(33). 그 이유는 고위험군의 임상 경과를 조기에 예측하는 것이 어렵지만, 약물 과다복용이나 부작용은 임상의가 흔히 겪고 있기 때문이다(33, 34). 따라서 기진단 된 환자들에게서 관찰된 생물학적 변이가 고위험군 상태의 환자에게서도 발견된다면 초기에 약물치료적 개입의 필요성과 타당성을 뒷받침할 것이다.

그러나 대부분의 혈액 단백체 마커 연구들은 증상이 완전히 발현된 환자들을 대상으로 진행되었고(22), 일부 정신증 고위험군, 양극성장애 환자의 자녀를 대상으로 매우 제한적으로 이루어졌다(35). 따라서 연구 1에서는 고위험군의 단백질 발현량을 정상 대조군과 환자군의 혈액 단백체 발현량과 비교 하여 고위험군에서 특징적으로 발견되는 단백체 마커에 대하여 탐색하고자 하였다. 또한, 기존에 환자군에서 보고된 단백체 마커들의 변이가, 고위험군 상태에서도 발견되는지 확인하고자 하였다.

연구 1에서는 진단의 이질성을 최소화하기 위하여, 정신증과 양극성장애 군으로 분리하여 분석을 수행하였다.

제 1 절 연구 대상 및 방법

1. 연구 대상자

1-1. 주요정신질환 고위험군

연구 대상자는 7개의 의료기관(서울대학교병원, 분당서울대학교병원, 서울특별시 보라매병원, 한양대학교병원, 국립정신건강센터, 국군수도병원, 서울대학교 보건진료소)에서 모집하였다. 각 기관의 연구 대상자는 서울대학교병원의 기분장애 클리닉, 청년클리닉에 방문하여 숙련된 평가자에게 평가받은 후 채혈하였다. 본 연구의 대상자들의 연구 참여 기준은 15-35세 사이의 주요정신질환 고위험 대상자로, 표1-1의 네 가지 범주 중 한 가지 이상에 해당하는 경우로 하였다 (24). 주요정신질환 고위험 대상자 모집 프로토콜은 기존 출간된 본 연구의 프로토콜 관련 논문인 Lee TY et al.(2020)에 기술된 기준으로 모집하였다.

2017년 6월부터 2019년 12월까지 총 90명의 주요정신질환 고위험군 대상자가 본 연구에 등록되었으며, 정신증 고위험군(CHR-PRS) 29명(32.2%), 양극성장애 고위험군(CHR-BD) 40명(44.4%), 주요우울장애 고위험군(CHR-MDD) 2명(2.2%), 다능성 기능 장애 고위험군(CHR-PL) 4명(4.4%), 확장 고위험군(CHR-Extended) 15명(BD 환자군 13명, 재발성 MDD 환자 2명)이 등록되었다.

연구 1에서는 FEP, HC와 단백질 발현량 비교를 위하여 CHR-PRS을, BD, HC와 단백질 발현량의 비교를 위하여 CHR-BD의 임상 및 혈장 단백질 정보를 이용하였다.

1-2. 주요정신질환 환자군

본 연구의 주요정신질환 환자군의 연구 참여 기준은 1) 15-35세로, 2) 최근 5년 동안 FEP 또는 BD(I형 또는 II형)를 진단받은 병력이 있는 환자들로 정의하였다. 2012년 5월부터 2018년 9월까지 서울대학교병원 방문자들 중에서 연구 참여자를 모집하였다. 총 41명의 FEP, CHR-Extended 중 BD 환자군 13명을 포함하고, 모집된 42명의 BD 환자를 합하여 55명의 BD(BD-I 7명, BD-II 48명)를 분석에 포함하였다.

1-3. 정상 대조군

주요정신질환 환자군, 고위험군과 연령-성별을 매칭한 HC 환자들을 모집하였고, (1) 이전에 주요정신질환으로 진단받은 병력이 없으며, (2) 직계가족 중 정신질환의 병력이 없는 환자들을 모집하였다. 총 75명의 HC가 최종 모집되었다.

HC와 CHR-BD, BD 단백질 발현량을 비교하는 분석에서는 75명의 HC이 모두 다 분석에 포함되었다. 반면, HC, CHR-PRS, FEP의 단백질 발현량을 비교하는 분석에서는 CHR-PRS의 경우 평균 연령이 20.7세로 나이를 매칭하여 17-30세(평균 연령 22.7세)의 45명의 HC 대상자가 분석을 위해 선택되었다.

1-4. 제외기준

연구 배제기준은 1) 신경조절치료(예, 전기경련치료, 경두개 자기자극치료, 경두개 직류자극치료)를 받은 적이 있는 경우, 2) 신경수술을 받은 경우, 3) 뇌수막염의 병력 또는, 간질이나 심각한 두부외상을 포함한 중추신경계 질환이 있는 경우, 4) 정신과적 증상을

유발할 수 있는 신체 질환(예, 갑상선질환)이 있는 경우, 5) 임신, 6) 지적 장애가 있는 경우, 7) 니코틴을 제외한 물질남용 병력이 있는 경우로 하였다.

연구에 참여한 대상자들은 모두 연구 참여 전 동의서를 작성하였으며, 18세 미만인 경우 대상자 본인, 부모/법적 보호자 모두가 동의서를 작성하였다. 본 연구는 서울대병원 기관 윤리심의위원회 승인을 받았으며, 헬싱키 선언의 원칙에 따라 수행되었다.

2. 임상증상평가도구

모든 연구 대상자들은 평가에 숙련된 간호사 또는 정신건강의학과 전문의와 대면 인터뷰를 통하여 임상 증상을 평가받았다.

2-1. 주요정신질환 환자군 및 고위험군 공통 평가도구

헤밀턴 우울증 평가 척도(Hamilton Depression Rating Scale, HAM-D).

HAM-D는 17개 항목으로 구성된 평가 척도로 면담자가 평가하는 우울 증상의 중증도 평가 척도이다(36). 한글판 신뢰도 및 타당화 연구에 따르면, 내적 일관성 신뢰도는 0.76였으며(37), 총점의 범위는 0-52점으로 점수가 높을수록 우울증이 심각하다는 것을 의미한다. 하부요인은 한국어판의 HAM-D를 사용한 국내 선행연구(38)의 하부요인을 사용하였다. 우울 증상의 중증도 반영을 위하여 HAM-D 총점과 5개의 하부요인(무쾌감증/지연, 죄책감/초조, 신체 증상, 불면, 식욕 저하) 점수를 활용하였다.

Young 조증 평가 척도(Young Mania Rating Scale, YMRS)

조증, 경조증 기분 상태의 연구 대상자를 배제 진단하기 위하여 YMRS 척도를 사용하였다(39, 40). YMRS는 조증 증상을 면담자가 평가하는 척도로, 총 11개의 문항으로 구성되어 있으며, 총점의 범위는 0-60점으로 점수가 높을수록 조증이 심각하다는 것을 의미한다. 한글판 신뢰도 및 타당화 연구에 따르면, 내적 일관성 신뢰도는 0.73였으며(40), 조증의 중증도 반영을 위하여 YMRS 총점과 3개의 하부요인(정신병적 조증, 과민형 조증, 조증) 점수를 활용하였다. 하부요인은 YMRS 요인분석 선행연구(41)의 하부요인을 사용하였다.

헤밀턴 불안 평가 척도(Hamilton Rating Scale for Anxiety, HAM-A).

HAM-A는 대상자의 지난 7일간의 불안 증상을 면담자가 평가하는 불안 증상의 중증도 평가 척도이다(42, 43). 14개 항목으로 구성된 평가 척도로, 심리적, 신체적 불안 모두를 포함한다. 각 문항은 0-4점으로 평가되며, 총점의 범위는 0-56점으로 점수가 높을수록 불안의 정도가 심각하다는 것을 의미한다.

양성 및 음성증후군 척도(Positive and Negative Syndrome Scale, PANSS)

PANSS는 정신증의 양성증상 및 음성증상에 중점을 두고 환자의 정신병리를 면담자가 평가하는 척도다(44, 45). PANSS는 양성척도 7문항, 음성척도 7문항, 일반 정신병리 척도 16개, 총 30개의 항목으로 구성되어 있다. 각 문항은 증상의 심각도에 따라 1-7점으로 평가한다. 한글판 신뢰도 및 타당화 연구에 따르면 내적 일관성 신뢰도는, 양성척도 0.73, 음성척도 0.84, 일반 정신병리 척도 0.74였으며, 각 문항의 평가 기준에 대해서는 상세한 기준이 문항마다 기재되어

있다(45).

전반적 임상 인상 척도-심각도(Clinical Global Impression-Severity, CGI-S)

CGI-S는 면담자가 환자가 진단받은 증상의 중증도를 1-7점으로 평가하는 척도이다(46). 각 점수는 '1점-정상/증상이 전혀 없음', '2점-약간의 증상, 효과적인 기능 수행은 지속할 수 있음', '3점-경도의 증상, 주관적인 고통이 있으나 기능적인 장애는 심하지 않음', '4점-중증도, 몇 가지 뚜렷한 증상, 중증도의 기능 장애', '5점-중고도, 뚜렷하고 중요한 증상, 상당한 수준의 기능 장애', '6점-고도, 매우 뚜렷하고 중요한 증상, 대부분의 영역에서 기능 수행 불가', '7점-최고도, 극도의 증상, 완전히 무능력 상태, 보살핌을 필요로 함'으로 평가한다.

총괄 기능 평가 척도(Global Assessment of functioning, GAF)

GAF는 면담자가 환자의 일정 기간 동안의 심리적 증상, 사회적, 직업적 기능에 대해서 평가하는 단일 문항이다(47, 48). 1-100점으로 평가하며, 제공된 정보가 평가하기에 불충분한 경우 0점으로 평가하며, 한글판 척도의 평가자 간 신뢰도는 0.91로 높은 신뢰도를 보였다(48). 1-100점을 10개의 구간으로 분류하여, 각 구간마다 환자의 증상 및 기능 저하가 어느 구간에 해당하는지 평가할 수 있도록 상세히 기술되어 있다. 주요정신질환 환자군은 현재의 증상 및 사회적 직업적 기능에 대하여 평가하였으며, 주요정신질환 고위험군 대상자는 현재, 지난 한 해 동안의 가장 높은 점수, 병전 상태(최초의 증상 이전)의 평균점수로 평가 기간을 분류하여 평가하였다.

2-2. 주요정신질환 고위험군 평가도구

DSM-IV축 I 장애 진단을 위한 구조화된 임상적 면담(Structured Clinical Interview for the diagnostic and statistical manual of mental disorder-IV-TR axis I disorder, SCID-I)

SCID는 DSM 진단분류를 기반으로 주요정신질환의 진단을 내리기 위하여 개발된 구조화된 면담 도구다(49, 50). 주요정신질환 고위험군이, 주요정신질환으로 진단될 수 있는 가능성을 배제하기 위해 구조화된 면담 도구를 사용하였다.

3. 혈장 단백질 분석

모든 연구 참여자는 임상 증상 평가와 동일한 일자에 혈액을 채취하였다. 채취한 혈액은 서울대학교병원 의생명연구원 융합연구 협력부의 단백체학 실험실에 전달하였고, 단백체학 실험실에서 혈장 시료의 전처리 및 질량분석을 수행하였다.

3-1. 혈장 시료의 전처리

혈액샘플은 EDTA 튜브에 채취하여, 4℃, 300rpm에서 10분 동안 원심분리 하였다. 분리된 혈장은 -70℃에 보관하였으며, Geyer et al. (2016) 방법에 따라 추가 처리되었다(19). 각 혈장 검체에 완충재(8M urea, 5mM TCEP, 20mM CAA in 0.1 M ABC)를 추가하고 샘플당 2 Ml을 사용하도록 공정을 최적화하였다. 단백체를 변성, 알킬화하기 위해서, 혼합물을60 ℃에서 25분간 열을 가하였고, 실온에서 냉각하였다. 냉각 후, 단백질 분해를 위해 트립신/LysC혼합물(100:1 비율의 단백질-단백질분해효소 비율 100:1)을 사용하여 하루 동안 처리하였다. 트립신을 이용하여(효소-기질 비율 1:1000) 다음 단백질

분해과정을 진행하였다. 모든 펩타이드는 10% trifluoroacetic acid로 산성화되었으며, Han et al. (2014)(51, 52)에서 보고된 C18-StageTips를 사용하여 desalting 하였다. 전 처리된 샘플은 진공 건조기에서 완전히 건조되어, 질량분석 전까지 -80℃에 저장하였다.

3-2. 주요정신질환 고위험군 관련 혈액 단백체 마커 후보의 선별

Mongan et al.(2021)(53) 연구에서 정신증 고위험군의 발병과 관련된 단백체 마커를 질량분석법을 사용하여 최근 보고하였으나, 주요정신질환 고위험군과 관련된 혈액 단백체 마커 연구는 환자군과 비교할 때 아직 드물다. 특히, 현재까지 보고된 연구들은 주로 정신증 고위험군 대상자로 국한되었으며, 염증과 관련된 마커(54)에 제한하여 보고하거나, 특정 신경 항체를 탐색하였으나(55) 의미 있는 발견을 보고하지 못하였다.

이에 본 연구에서는 주요정신질환 고위험군 관련 혈액 단백체 마커 후보를 선별하기 위해, 환자군의 선행연구를 이용한 데이터 마이닝과 환자군 혈액시료의 정량 단백체 분석(High-throughput single liquid chromatography-tandem mass spectrometry, LC-MS/MS)을 통해 단백체 마커 후보군을 선별하였다(그림 1-1). 기존 선행연구 및 데이터 마이닝을 통해 주요정신질환과 관련된 342개의 단백체 마커를 선별할 수 있었다. 또한, 기구축된 FEP, MDD, BD, HC의 혈장 시료를 Discovery sample set 분석을 통해 주요정신질환과 관련성이 있는 120개의 단백체 마커 후보군을 선별하였다. 본 연구의 후보 마커 선별 과정에서 발견한 양극성장애 및 주요우울장애와 관련된 단백체 마커(56) 및 우울증 중증도와 관련된 단백체 마커(57)는 본 연구자가 참여한 선행연구를 통해 보고한 바 있다.

데이터 마이닝과 환자군의 정량 단백체 분석을 통해 선별한 혈액 단백체 마커를 대상으로 In-silico trypsin digestion 과정을 통하여 발생 가능한 모든 Tryptic peptide를 확인하고, MRM에 적합한 peptide sequence를 가지는 펩타이드를 1차 filtering 하였다. 이후 자체 구축한 peptide ranking system(Peptide atlas, SRM atlas, CPTAC portal 등)으로 MRM assay로 개발된 적이 있는 펩타이드에 score를 주고, In-house spectrum library에서 발견된 펩타이드와 peak intensity가 높은 peptide에 score를 주는 방식으로 펩타이드에 rank를 부여하였다. 이를 통해 각 단백질 당 우선순위의 펩타이드를 선별, 총 225개의 heavy isotope labeled 펩타이드를 확보하였다. 225개의 확보된 TOP transition을 사용한 multiplex MRM 반복분석을 통해 표적 펩타이드에 대한 정량성을 확인하고, endogenous 펩타이드의 양과 양의 range를 확인하여 multiplex MRM으로 정량 가능한 158개의 펩타이드(92개의 단백질)를 최종 선별하였다(그림 1-1).

3-3. 다중 검지법을 이용한 혈장 단백질 정량 분석(Liquid chromatography-multiple reaction monitoring-mass spectrometry, LC-MRM-MS)

최종 선별된 158개의 표적 펩타이드를 사용, LC-MRM-MS 분석법으로 HC, CHR군, 환자군(FEP, BD)의 혈장 단백질을 정량화 하였다. 최근 Mongan et al.(2021)에서 정신증 고위험군의 발병과 관련된 단백질을 질량분석법(LC-MS/MS)을 통해 보고하였으나(53), 고위험군의 혈액 단백질을 질량분석법으로 분석한 연구는 현재까지 매우 드물며, MRM-MS 분석법을 분석한 연구는 현재까지 보고되지 않았다. 특히, 고위험군을 대상으로 한 혈액 단백체 연구는 환자군과 비교할 때

아직 보고된 단백체 마커의 수가 드물기 때문에 (54, 55), 본 연구에서는 다량의 단백질 변이를 동시에 높은 정확도로 측정할 수 있는 본 분석법을 적용하였다.

4. 통계분석

4-1. 단백체 데이터의 전처리

단백체 데이터의 원자료는 Skyline 소프트웨어를 사용하여 전처리하였다 (58). 각 펩타이드의 정량은 Abundance ratio, 즉 내인성과 SIS quantifier의 비로 산출되었다. 본 연구의 분석에 사용된 HC 및 환자군의 혈액 시료는 서울대학교병원 내 두 개의 클리닉(청년 클리닉, 기분장애 클리닉)에서 채혈되었다. 따라서 채혈 및 보관 과정에서 발생하는 배치 효과를 줄이기 위하여 Combat 알고리즘(http://genepattern.boadinstitute.org)을 사용하여 보정하였다. 보정 후 158개의 단백질의 주성분 분석 plot(principal component analysis, PCA)을 통하여 확인하였으며, 이 중 15개의 펩타이드(8개의 펩타이드)는 배치 효과가 보정 후에도 차이를 보였으므로 환자군 및 HC을 포함한 분석에서 제외하였다. 단백체 데이터에 결측값은 없었다. 연구 1 분석에 포함된 사회인구학적 변인 및 임상 변인 데이터 또한, 결측값은 없었다. 단백체 정량 값의 분포가 정규분포를 만족하지 않고, 편중되어 있어 Log2 변환을 통해 전처리하였고, 이후 Shapiro-Wilk 검정을 통해 정규성을 평가하였다.

4-2. 통계분석

범주형 변수는 피어슨 카이제곱 검정(Pearson's chi-squared test), 피셔의 정확한 검정(Fisher's exact test)을 사용하였다. 두 군의 연속형

변수는 독립 표본 t 검정, Mann-Whitney의 U 검정으로, 세 군 이상의 연속형 자료는 Kruskal-Wallis H 검정, 분산분석(analysis of variance, ANOVA)으로 분석하였다. 단백질 발현은 나이, 성별, 항정신병약물의 복용(59-61)에 의해 영향을 받을 수 있으므로 이 변수들을 공변량으로 하여, ANOVA에서 유의한 차이를 보인 단백체 마커들을 대상으로 공분산분석(analysis of covariance, ANCOVA) 및 Turkey의 사후 검정을 통해 보정하였다. 상관분석은 임상 증상과 유의한 차이를 보이는 단백질(differentially expressed protein, DEP)의 발현량 간의 상관관계를 평가하기 위해서 사용되었다. 다중 비교를 위해 보정하기 위해 Benjamini-Hochberg false discovery rate(FDR)를 보정한 유의확률 q 값을 사용하였다.

생물정보학 분석은 Database for Annotation, Visualization and Integrated Discovery(DAVID v6.8, http://david.ncifcrf.gov/)를 통해 gene ontology(GO) 분석(62)을 수행하였다. Upstream regulator 예측 분석은 Ingenuity 경로 분석(IPA, QIAZEN, Hilden, Germany)을 사용하여 분석하였다. 통계분석은 R Statistical Software 버전 4.0.2 (63) 및 IBM SPSS 25.0(IBM Inc., Armonk, NY)을 사용하여 수행되었다. 유의 수준은 $p < 0.05$, $q < 0.05$ 양측 검정을 사용하였다.

제 2 절 결과

1. 정상 대조군(HC)-양극성장애 고위험군(CHR-BD)-양극성장애 환자군(BD)의 비교

1-1. 사회인구학 및 임상특성 요인 비교

표 1-2는 HC, CHR-BD, BD의 사회인구학 및 임상특성을 비교한 결과이다. HC는 나머지 두 군과 비교하여 교육연수가 더 유의하게 길었고, CHR-BD보다 BD에게서 항정신병약물 및 기분조절제의 사용빈도가 더 높았다. YMRS의 총점은 CHR-BD보다 BD에서 더 높았다. 그 외 사회인구학적(나이와 성별)특징과 임상적 특성(항우울제 사용, 벤조디아제핀 사용, HAM-D 및 CGI-S)은 그룹 간에 유의한 차이를 보이지 않았다.

양극성장애 위험 상태 기준(Bipolar at risk state criteria, BARS, 표1-1)에 따라 분류된 CHR-BD의 하위 그룹은 다음과 같다(표1-3): 그룹 1. 역치하 조증(5명, 12.5%), 그룹 2. 우울한 기분 + 기분순환특징(28명, 70.0%), 그룹 3. 우울한 기분 + 양극성장애 유전적 위험(6명, 15.0%), 그룹 4. 기분순환특징 + 양극성장애 유전적 위험(0명, 0.0%), 그룹 5. 역치하 혼재성 삽화(0명, 0.0%), 그룹 6. 기분변동(1명, 2.5%). CHR-BD 하위 그룹 1-3의 임상 증상(HAM-D, YMRS, CGI-S) 심각도를 비교하였을 때 유의한 차이는 없었다(표1-3).

1-2. 혈중 단백질의 정량 분석결과

143개의 표적 펩타이드(85개의 단백질) 중에서 30개의

펩타이드(27개의 단백질)가 ANOVA 및 사후분석 후 FDR <0.05에서 유의하였다(표1-4). 이 중 ANCOVA를 통해 공변량(나이, 성별, 항정신병약물 사용, 기분안정제 사용, 항우울제 사용, 벤조디아제핀 사용)을 보정한 결과, PRDX2[1]를 제외한 모든 펩타이드는 여전히 유의한 차이를 보였다(표1-5). ANOCVA 분석결과 유의한 29개의 펩타이드 중 23개의 펩타이드가 CHR-BD에서 BD, HC와 유의한 차이를 보였다. 이 중 2개의 펩타이드(CMGA, QSOX1)[2]이 CHR-BD에서 과발현되었고, 나머지 21개의 펩타이드(19개의 단백질[3])가 저발현되었다(표1-5). CHR-BD에서 BD보다 3개의 펩타이드(LYSC, MBL2, CBG)[4]는 유의하게 감소하였고, TPM4은 유의하게 증가하였다. 다른 두 펩타이드(CBG, SAMP)[5]는 HC에 비해 CHR-BD에서 유의하게 감소하였다.

2. 정상 대조군(HC)-정신증 고위험군(CHR-PRS)-조현병 환자군(FEP)의 비교

2-1. 사회인구학 및 임상특성 요인 비교

표 1-6는 HC, CHR-PRS, FEP의 사회인구학 및 임상특성을 비교한

---

[1] PRDX2, peroxiredoxin 2
[2] CMGA, chromatogranin-A; QSOX1, sulfhydryl oxidase 1
[3] FA10, coagulation factor X; GELS, gelsolin; APOA1, apolipoprotein A1; APOA2, apolipoprotein A2; BDNF, brain derived neurotrophic factor; ITIH4, inter-alpha-trypsin inhibitor heavy chain family member 4; LUM, lumican; VWF, von Willebrand factor; KNG1, kininogen 1; HGFA, HGF activator; KLKB1, kallikrein B1; BTD, biotinidase; A1BG, alpha-1-B glycoprotein; B2MG, beta-2-microglobulin; NCHL1, cell adhesion molecule L1 like; MASP2, mannan binding lectin serine peptidase 2; HEMO, hemopexin; CHLE, butyrylcholinesterase; PROS, protein S (alpha)
[4] LYSC, lysozyme C; MBL2, mannan-binding lectin serine protease 2; CBG, cortisol binding globulin
[5] SAMP, amyloid P component serum

결과이다. HC는 나머지 두 군과 비교하여 교육연수가 더 유의하게 길었다($p<0.001$). CHR-PRS보다 FEP에서 항정신병약물의 빈도가 높았고($p<0.001$), 기분조절제 사용의 빈도는 CHR-PRS에서 FEP보다 더 높았다($p=0.014$). PANSS 총점($p=0.001$) 및 양성증상 총점($p<0.001$)은 CHR-PRS보다 FEP에서 더 높았다. 또한, GAF-현재 점수는 FEP가 CHR-PRS보다 유의하게 낮았다($p=0.001$) 그 외 사회인구학적(나이, 성별, 직업, 신체질량지수, 정신과 질환의 가족력) 특징과 임상적 특성(항우울제 사용, 벤조디아제핀 사용, HAM-D, HAM-A, YMRS)은 그룹 간에 유의한 차이를 보이지 않았다.

2-2. 혈중 단백질의 정량 분석결과

143개의 표적 펩타이드(85개의 단백질) 중에서 16개의 펩타이드(14개의 단백질)가 ANOVA 및 사후분석 후 FDR <0.05에서 유의하였다(표1-7). 이 중 ANCOVA를 통해 공변량(나이, 성별, 항정신병약물 사용, 기분안정제 사용, 항우울제 사용, 벤조디아제핀 사용)을 보정한 결과, 모든 펩타이드는 여전히 유의한 차이를 보였다(표1-8). 이 중 9개의 펩타이드(7개의 단백질[6])가 CHR-PRS에서 FEP, HC와 비교 하였을 때 유의하게 감소하였다. 또한, 3개의 펩타이드가(PRDX2, FETUB, NCHL1)[7]이 CHR-PRS에서 FEP보다 감소 되어 있었고, 4개의 펩타이드(GELS, LUM, KLKB1, MBL2)는 CHR-PRS에서 HC보다 유의하게 감소하였다(표 1-8).

3. 정신증 고위험군(CHR-PRS)과 양극성장애 고위험군(CHR-BD)과의

[6] FA5, coagulation factor V; KNG1, HEMO, ITIH4, FA10, BTD, A1BG
[7] FETUB, Fetuin-B

비교

3-1. 사회인구학 및 임상특성 요인 비교

표1-9는 CHR-PRS와 CHR-BD 두 군의 사회인구학 및 임상특성을 비교한 결과이다. CHR-PRS에서 남성의 비율(18명, 62.1%)이 CHR-BD(13명, 32.5%)보다 높았고(p=0.015), 그 외의 사회인구학적 요인(나이, 신체질량지수, 교육연수, 사회경제적 수준)은 유의미한 차이가 없었다. 또한, 정신질환의 가족력, 정신과 치료 병력, 약물 복용력도 두 군간 유의미한 차이가 없었다(표1-9). PANNS 총점 및 모든 하위 척도(양성, 음성, 일반증상 척도) 모두 CHR-PRS군이 유의하게 높았으며, GAF-현재, GAF-한해 모두 CHR-PRS군이 CHR-BD군보다 낮았다.

3-2. 혈중 단백질의 정량 분석결과

143개의 표적 펩타이드(85개의 단백질) 중에서 14개의 펩타이드(11개의 단백질)가 $p < 0.05$에서 유의한 차이를 보였다(표 1-10). 그러나 Benjamini-Hochberg's FDR 값인 $q < 0.05$에서는 유의하지 않았다(표1-10).

결과 요약

1-3의 그룹 간 단백체 비교 분석결과를 요약하면 그림1-2와 같다. CHR-BD군에서 BD 또는 HC군과 비교하여 유의미한 차이를 보인 27개의 단백질(29개의 펩타이드)과, CHR-PRS군에서 FEP 또는 HC군과 비교하여 유의미한 차이를 보인 14개의 단백질(16개의

펩타이드) 중 11개[8]의 단백질(12개의 펩타이드)이 양쪽 모두에서 유의미한 차이를 보이는 단백질(DEP) 이었다. CHR-PRS, CHR-BD 두 군 사이에는 q <0.05 수준에서 유의한 차이를 보인 단백질이 없었으므로, 이는 고위험군에서 나타나는 단백질의 변이는 질병-특이적이기보다 고위험군 상태를 반영하는 마커가 될 수 있음을 시사한다. 이 결과를 토대로 고위험군(CHR-PRS, CHR-BD)에서 특이적으로 증가 또는 감소한 30개의 단백질을 선택하여 이러한 단백질들이 어떠한 증상상태를 반영하는지 증상과의 상관분석을 추가로 진행하였다.

4. 고위험군 특이적 발현양상 변화 단백질과 임상 증상 심각도의 상관분석

그림 1-3는 FEP, CHR 각 군에서 PANSS 총점 및 하위요인점수와 33개의 펩타이드(30개의 단백질) 발현량 사이의 상관관계 분석결과를 plot으로 요약한 결과 이다. CMGA 단백질의 발현량은 고위험군에서는 PANSS 양성증상 총점과 음의 상관관계를 보였으나, 조현병 환자군에서는 반대로 양의 상관관계를 보였다.

그림 1-4은 FEP, CHR, BD 각 군에서 HAM-D 총점 및 하위 요인 점수, HAM-A 임상 증상 심각도와 33개의 펩타이드(30개의 단백질) 발현량 사이의 상관관계 분석을 상관분석 plot으로 요약한 결과다. NCHL1, APOA1 단백질은 HAM-D 식욕감소 하부요인과 FEP, CHR 군에서 각 군에서 유의한 음의 상관관계를 보였다.

그림 1-5은 FEP, CHR, BD각 군에서 YMRS점 및 하위 요인 점수,

[8] GELS, HEMO, ITIH4, LUM, KNG1, KLKB1, FA10, BTD, A1BG, NCHL1, MBL2

CGI-S, GAF-현재, GAF-한해 점수와 33개의 펩타이드(30개의 단백질) 발현량 사이의 상관관계 분석을 상관분석 plot으로 요약한 결과이다. TPM4는 FEP, CHR 군에서 각각 YMRS 총점과 유의한 음의 상관관계를 보였다(그림 1-5). 또한, TPM4는 FEP 군에서 YMRS-정신병적 조증 하위점수와도 음의 상관관계를 보였으며, CHR군에서는 예민한 조증 하위점수와 음의 상관관계를 보이는 것으로 볼 때 조증 증상의 심각도 또는 정신증의 심각도를 반영하는 후보 마커가 될 수 있음을 시사 한다.

5. 생물정보학(Bioinformatics) 분석결과

30개의 DEP가 생물정보학 분석에 포함되었다. 대부분의 단백질은 보체 연쇄 과정 및 응고 연쇄 과정과 관련되었고, 일부 세포 단백질 대사 과정에 관여했다. 세포 성분 분석결과는 대부분의 단백체 마커가 세포 외 공간의 구성 및 소포체와 관련 있었다. 분자 기능 분석에 따르면 세린 타입-엔도펩티다아제 활성, 칼슘 이온 결합과 관련이 있다(그림 1-6A). 30개의 DEP에 대한 전사 조절자의(Transcription regulator)의 연쇄 인자를 조사하기 위하여 upstream regulator 예측 분석을 사용하였고, 조절자-타겟 네트워크를 구성하였다(그림 1-6B). 그 결과, 30개의 DEP를 조절하는 3개의 상류 조절기(CTCF, HNF1A, SP1)[9]를 발견하였다.

[9] CTCF, CCCTC-binding factor; HNFA1, hepatocyte nuclear factor 1-alpha-A; SP1, transcription factor Sp1

제 3 절 고찰

본 연구 1에서 발견된 고위험군에서 특이적으로 증가 또는 감소한 발현양상을 보인 단백질들(DEP)은 3개의 단백질을 제외하고 대부분 고위험 상태에서 감소하였다. 또한, 이 DEP들의 대부분은 주요정신질환 환자군의 선행 단백체 연구에서 보고된 바 있는 단백체 마커였다(12, 64, 65)

먼저 생물정보학 분석결과는 SP1, CTCF, HNFA1과 같은 upstream regulator들이 DEP들의 유의한 상위 조절자였다. SP1은 양극성장애 환자군을 대상으로 한 선행연구에서 취약한 전사인자로 보고된 바 있으며(66), CTCF는 학습, 인지기능과 관련된 유전자의 조절자로 보고된 바 있다(67). 따라서 인지기능 및 양극성장애 환자에게서 보고된 상위 조절자들이 고위험군에서 발현 변이를 보이는 단백질들의 상위 조절자들과 관련성을 보인 본 결과는, 고위험군의 병리학적인 변화가 질병의 초기 단계에서부터 시작되고 있음을 시사한다. 또한, 이러한 결과들은 완전히 증상이 발현된 환자들보다 고위험군 상태에서 더 활발한 병태생리학적 변화가 진행되고 있음을 암시한다.

CMGA, QSOX1 단백질은 CHR-BD군에서 HC, BD보다 증가한 발현양상을 보였으며, TPM4는 CHR-BD에서 BD보다 유의하게 증가한 발현양상을 보였다. CMGA의 발현량 증가는 선행연구에서 FEP, BD 군에서 보고된 바 있으며(8), CMGA의 기능은 염증을 유발(pro-inflammation)하고 신경전달물질의 이상(68)과 관련이 있었다. QSOX1은 정신질환과 관련하여 보고된 선행연구는 없었으나, 세포 분열 주기를 방해하여 유방암 환자의 생존율을 낮추는 단백질로 보고되었다(69). 그러므로 선행연구에서 보고된 바에 따르면

고위험군에서 증가한 단백질들은 신경염증과 같은 질병의 악화 및 진행과 관련이 있을 수 있고, 향후 고위험군의 질환 경과를 막기 위한 치료를 위한 표적이 될 수 있겠다.

고위험군 그룹에서 덜 발현된 단백질은 혈액 응고[10], 면역반응[11], 지질대사[12], 신경가소성의 발달[13]와 관련이 있었다. 정신적 스트레스는 카테콜아민과 세로토닌 경로를 자극하여, 혈액 응고 연쇄반응을 활성화하는 것으로 알려져 있다(70). 혈액 응고 경로와 관련된 A1BG, KNG1, KLKB1 또한, BD 환자에서 발현량이 감소한 것으로 선행연구에서 보고된 바 있다(71). 또한, 항염증 반응과 관련된 ITIH4와, 선천면역과 관련된 MASP2, MLB2 단백질은 환자군에서 발현량 감소가 보고된 바 있다. 따라서 항염증 반응 및 면역력과 관련된 단백질들의 발현량 감소는, 고위험군 상태가 면역반응에 취약한 상태임을 시사한다(65, 72).

콜레스테롤 대사의 초기 변이 또한, 많은 정신질환 연구에서 보고되고 있으며, 혈청 콜레스테롤의 감소는 신경수초와 시냅스 형성 및 발달과 관련된 기능 장애와 관련된다(65). NCHL1과 BDNF는 신경세포의 시냅스 가소성과 관련된 단백질로(73, 74), 본 연구를 통해 고위험군 상태에서 관련된 병리가 진행되고 있음을 확인할 수 있었다. Cortisol binding globulin(CBG)는 우울증, 양극성장애 환자군에서 보고된 바 있는 단백질로 만성 스트레스의 잠재적인 생물학적 지표로(75), 주요정신질환으로의 고위험 상태를 반영하는 유용한

[10] A1BG, VWF, FA5, FA10, KNG1, KLKB1
[11] ITIH4, MASP2, MBL2, LYSC, B2M
[12] APOA1, APOA2
[13] BDNF, NCHL1

마커가 될 가능성이 있겠다. 이러한 생물학적 경로 이상 외에도 칼슘 조절을 통하여 세포사멸을 방지하는 GSN(76), 신경독성 물질을 분해하는 CHLE, 산화 스트레스 감소기능이 있는 HPX, HGFA와 같은 다양한 신경 보호 단백질들이 고위험군에서 감소한 양상을 보였다 (65, 77, 78) 감소한 단백질들은 주요정신질환을 예방하는 경로와 관련된 급성 면역반응, 신경 보호 기능, 그리고 신경 가소성의 기능 장애와 관련이 있었다. 그러므로, 연구 1의 결과는 질병 경과를 예방하는 메커니즘의 기능 저하가 주요정신질환 고위험 상태에서부터 진행되고, 관찰될 수 있다는 이론을 뒷받침한다.

PRDX2는 ANOVA 분석에서 질병 진행(HC < CHR−BD < BD) 됨에 따라 발현량이 증가하는 양상을 보였다. 약물사용을 통계적으로 통제한 후 결과는 더 이상 유의하지 않았으나, 환자군에서 PRDX2의 발현량 증가는 이전 연구에서 보고된 바 있다(56, 79). 따라서 PRDX2가 질병의 진행을 반영하는 의미 있는 단백질 표지자 인지를 밝히기 위해서는 추가적인 종적 연구가 필요할 것으로 생각된다.

반면, 정상 대조군과 환자군을 비교한 선행연구와 결과와 다르게 (56), PRDX2를 제외하고 환자군−HC 간의 차이는 감지되지 않았다. 이는 Kim et al.(2021)에 의해 유의한 것으로 확인된 표적 펩타이드 중 3개의 펩타이드(PRDX2, F13A, GAPDH)만이 대상 물질의 안정성 및 측정 가능성 때문에 본 연구에 MRM 분석 후보 펩타이드로 포함되었기 때문일 수 있다. 그럼에도 선행연구에서 환자군과 HC간 구분되는 마커로 보고된 마커들이 본 연구에서 유의한 차이를 보이지 않았다 (12, 64, 65). 이러한 결과의 차이는 이전 선행연구들에서도 보고되었듯이(46, 49), 이러한 단백질 발현량 변화는 증상이 완전히 발현되고, 치료를 시작한 환자군에서 정상화 된 것을 반영한 것일 수

있다(80-83). 대부분의 본 연구에 참여한 환자들은 연구 참여시 약물치료에 제한을 두지 않았으나, 고위험군 대상자들은 3주 이상 항정신병약물을 복용했거나 6주 이상 기분조절제를 복용했다면 본 연구에서 배제되었으므로, 향후 연구에서는 이러한 약물 복용 및 약물치료의 기간을 통제할 필요가 있겠다.

본 연구의 한계점으로는 단면적으로 모집된 혈액 시료를 분석에 사용하였으므로, 이러한 단백질 변이가 실제 질병 진행 및 단계의 인과관계를 반영하는 마커들로 보기에는 어려움이 있다. 또한, 연구 1에서와 마찬가지로, 단백질 발현에 영향을 미칠 수 있는 교란요인(나이, 성별, 약물사용 여부)은 통계적으로 보정하였으나, 그 외에 중요한 요인(신체질량지수, 흡연 여부, 공복 여부, 약물복용의 기간, 복용량)들에 대해서는 영향력을 통제하지 못하였다. 그러나 연구 1은 고위험군의 혈중 단백체 마커를 환자군, 정상 대조군과 비교하였다는 점에서 상당히 의미 있는 단백체 마커들을 발견하였고, 병리학적 경과가 고위험 상태에서 진행되고 있음을 시사하는 결과를 보였다는 점에서 의미가 있다.

제 3 장 – 연구 2. 주요정신질환 고위험군 증상 심각도와 관련된 혈액 단백체 마커 탐색

다차원 증상 요인들을 이용한 잠재프로파일 분석

연구 2에서는 연구 1에서 적용된 고위험군의 분류 기준의 한계를 극복하기 위해 고위험군의 이질적인 증상 도메인을 종합하고, 이질적인 특성을 증상에 기반하여 유형화하고자 하였다. 이를 위해 고위험 대상자들이 가장 높은 빈도로 호소하며, 연구 1에 적용된 주된 임상 영역인 정신증 양성증상, 조증 증상, 우울, 불안을 평가한 임상척도 영역을 종합하고, 평가된 임상 증상을 기반으로 유형화하기 위하여 잠재프로파일 분석(Latent profile analysis, LPA)을 수행하였다.

잠재프로파일 분석은 이질적인 특성을 유형화하고 이를 기반으로 새로운 하위 잠재집단을 구분하는 방법이다(84). 본 연구에서는 1) 정신증 관련 2개의 지표(SIPS(Structured Interview for Prodromal Syndrome) 양성, PQ-B(Prodromal Questionnaire-Brief), 2) 3개의 조증 지표(BPSS-AP(Bipolar prodrome symptom scale-abbreviated prospective)의 하위 척도인 증상 심각도, 빈도, 역치), 3) 2개의 우울 지표(QIDS(Quick Inventory of Depressive Symptomatology), HAM-D), 4) 3개의 불안 지표(HAM-A, STAI-S/T(State-Trait Anxiety Inventory) 영역, 총 4개의 영역, 10개의 지표를 포함하여 LPA 모형을 구성하였다.

LPA 분석을 통해 연구 2에서는 증상의 차원으로 분류된 집단의 단백질 발현량을 비교하고자 하였다. 뿐만 아니라, 다양한 증상영역과 단백질 발현량의 상관관계를 분석하여, 증상의 중증도를 반영하는 단백체 마커를 탐색하고자 하였다.

제 1 절 연구 대상 및 방법

1. 연구 대상자: 주요정신질환 고위험군

연구 1에 기술된 주요정신질환 고위험군의 대상자 90명이 연구 1의 주요정신질환 고위험군에 포함되었다.

2. 임상증상평가도구

고위험 대상자들의 이질적인 특성을 유형화하기 위하여 LPA에 투입된 지표는 다음의 평가도구를 토대로 선정하였다. 또한, 면담자 평가 척도는 평가에 숙련된 간호사 또는 정신건강의학과 전문의와 대면 인터뷰를 통하여 평가하였다.

LPA 분석에 사용된 임상 평가도구

2-1. 전구기 증후군을 위한 구조화된 면담(Structured Interview for Prodromal Syndrome, SIPS)

SIPS는 주요정신질환 전구기 증후군을 진단하는데 사용되는 구조화된 진단 인터뷰 도구이다(28, 85). 이전 선행연구들에서 정신증의 전구기 환자를 진단에서 SIPS의 유효성을 확인하였다. Miller et al. 연구진들은 SOPS 기준을 사용하여 평가한 전구기 대상자 중 양성 예측률은 평가 후 12개월에 50%, 24개월에 67%를 보였다(29). 이러한 연구결과는 SIPS가 전구기 환자 증상의 심각성을 측정할 뿐만 아니라 전구기 증후군 경과 예측에 유용할 수 있음을 시사한다. 한글판 신뢰도 및 타당화 연구에 따르면, 40명의 대상자 중 12.5%가 1년 추적 기간 동안

주요정신질환으로 진단되었으며, 평가자 간 신뢰도 0.96, 내적 일관성 신뢰도 0.83으로 양호하였다(85). SIPS는 19개의 문항이 4가지 영역(전구적 양성증상-5문항, 전구적 음성증상-6문항, 전구적 파괴증상-4문항, 전구적 일반증상-4문항)으로 나누어 평가하며, 각 문항의 증상의 심각도에 따라 0-6으로 평가한다. 각 문항의 평가 기준 및 구체적인 질문에 대하여 문항마다 상세히 기술되어 있다. LPA 모형에는 전구기 환자의 정신증 증상 특성을 반영하기 위하여 양성증상 영역의 총점이 투입되었다.

2-2. 전구기 질문지-축약본(Prodromal Questionnaire-Brief, PQ-B)

PQ는 92개의 질문으로 구성되며, 정신증 전구기 상태 진단을 위한 선별도구로 개발되었다(86). PQ-B는 평가시간의 단축과 효율성을 높이기 위해 21개의 문항으로 축약하기 위해 개발되었다(87). PQ와 PQ-B 평가도구 모두 90%의 높은 민감도를 보이지만, PQ-B는 높은 특이성을 가진 효과적인 도구로 알려져 있다. 각 문항의 질문에서 설명하는 경험 여부를 표시하고, 경험이 있음에 "예"를 선택한 문항에 대해서 5점 리커트 척도를 사용하여 정도를 사용하여 정도를 평가한다. 7개 이상의 문항에 대한 경험이 있음에 '예'라고 응답한 경우 고위험군으로 의심할 수 있다 (88). 한글판 신뢰도 및 타당화 연구에 따르면, 총점 7점을 cut-off 점수로 하였을 때, 민감도 72.2, 특이도 71.4로 가장 적합하였으며, 내적 일관성 신뢰도는 0.83 이었다(86),

2-3. 양극성 전구기 증상 척도 -전향적 평가용 축약판(Bipolar Prodrome Symptom Scale-Abbreviated Prospective, BPSS-AP)

BPSS-full prospective는 양극성장애 전구기 환자를 평가하기

위하여 SIPS의 구조와 매우 유사하게 개발되었다(30). BPSS는 반구조화된 면담 도구로 DSM-IV의 양극성장애, 주요우울장애의 조증 삽화, 주요 우울 삽화 등의 정신병리학에 기초한 질문들을 포함하며, DSM 진단기준의 존재여부, 중등도, 약화된 증상의 유무를 파악할 수 있다. 또한, 증상의 심각도와 증상의 빈도를 모든 문항에 대해서 평가한다. 본 연구에서는 SIPS, PANSS, HAM-D 평가도구를 함께 사용하였으므로, 일반증상과 우울 증상의 문항이 제외된 BPSS-AP 축약판으로 평가하였다. 축약판은 10개의 조증 관련 문항과 1개의 일반 증상 문항으로 구성되어 있다 (89). 각 문항은 지난 1개월간의 중증도, 빈도에 따라 각 0-6점으로 평가하며, 각 문항이 환자의 기능에 영향이 있을 경우 4점 이상으로 평가한다. 각 문항별로 증상의 심각도와 빈도를 종합한 4단계의 임계점수 기준(Threshold 1-4 [14])을 적용하였고, 임상적으로 의미 있는 구체적인 증상상태로 분류하였다. 각 문항에 적용된 임곗값에 따라 전체 임곗값의 총합으로 문항의 총합을 구하였다. 현재까지 본 평가척도의 한글판 신뢰도 및 타당화 연구는 보고된 바 없으며, LPA 모형에는 전구기 환자의 조증 증상 특성을 반영하기 위하여 3가지(조증 심각도 총합, 조증 빈도의 총합, 조증 역치의 총합) 총점이 개별적으로 투입되었다.

2-4. 간이 우울 증상 평가 척도(Quick Inventory of Depressive Symptomatology, QIDS)

[14] Threshold 1 (경미한 증상-경미한 빈도) : 증상점수≤3, 빈도점수≤2,
Threshold 2 (경미한 증상-중등도 빈도) : 증상점수≤3, 빈도점수≥3,
Threshold 3 (중증도 증상-경미한 빈도) : 증상점수≥4, 빈도점수≤3,
Threshold 4 (중증도 증상-중증도 빈도) : 증상점수≥4, 빈도점수≥3

16개의 문항으로 구성된 QIDS척도는 자가보고와 면담자 평가 모두 사용할 수 있는 평가 척도로, 본 연구에서는 대상자의 주관적인 우울 증상을 평가하기 위하여 자가보고척도로 사용하였다(90). QIDS는 DSM-IV 주요우울장애 진단기준 9개의 영역을 포괄하며, 각 문항은 0-3점으로 평가한다. 점수가 높을수록 주관적인 우울감의 정도가 심한 것을 반영한다. 한글판 신뢰도 및 타당화 연구에 따르면, 내적 일관성 신뢰도는 0.73 이었다(90). LPA 모형에는 전구기 환자의 주관적 우울 증상 심각도를 반영하기 위하여 QIDS 총점이 투입되었다.

2-5. 상태-기질 불안 척도(State-Trait Anxiety Inventory STAI)

STAI는 대상자의 불안 상태를 평가하는 자가 보고식 척도로, 일반적으로 대상자가 경험하는 불안 정도를 평가하는 20개의 문항과(STAI-Trait, STAI-T), 현재 느끼는 불안 정도를 평가하는 20개의 문항(STAI-State, STAI-S)으로 구성된다(91, 92). 역채점 문항을 포함하고 있으며, 각 문항은 1(전혀 그렇지 않음)-4점(대단히 그렇다)으로 평가 된다. 점수가 높을수록 불안의 정도가 높은 것을 의미한다. 한글판 신뢰도 및 타당화 연구에 따르면, 내적 일관성 신뢰도는 0.91이었다(92). LPA 모형에는 전구기 환자의 주관적 현재 불안, 특성 불안을 반영하기 위하여 STAI-T, STA-S 총점이 투입되었다.

LPA 분석에 포함되지 않은 임상 평가도구

아동기 외상 질문지-축약형(Childhood Trauma Questionnaire-short form, CTQ-SF)

CTQ-SF는 아동 또는 청소년기에 경험한 5가지 유형의 학대나 방임을 평가하는 자가 보고식 설문지로 총 28개의 문항으로 구성되어 있다(93, 94). 또한, 남용을 최소화하거나 보고 거부 등 응답에 대한 타당성을 평가하기 위한 3개의 문항을 포함하고 있다. 대상자는 각 문항에 해당하는 경험의 정도에 따라 0-5점으로 평가한다. 역채점 문항을 포함하며, 총점이 높을수록 외상 경험의 정도가 심함을 의미한다. 한글판 신뢰도 및 타당화 연구에 따르면, 내적 일관성 신뢰도는 0.88이었으며, 5개의 하부요인(신체학대, 성 학대, 정서학대, 정서방임, 신체 방임)의 설명력은 64%였다(93).

바렛 충동성 척도(Barratt Impulsiveness Scale, BIS)

BIS는 특성 충동성을 측정하는 자기보고 척도로, 30개 문항으로 구성되어 있으며, 11개의 역채점 문항을 포함한다. 각 항목은 증상 빈도에 따라 1-4점으로 평가되고, 총점은 충동성의 정도를 나타낸다(95). 한글판 신뢰도 및 타당화 연구에 따르면, 내적 일관성 신뢰도는 0.78이었으며, 원저 및 한국판의 3개의 하부요인(주의 충동성, 운동 충동성, 무계획 충동성)으로 구분 하였다(95, 96).

생물학적 리듬 평가 설문지(Biological Rhythms Interview of Assessment in Neuropsychiatry, BRIAN)

BRIAN은 특히 정신질환이 있는 환자들의 일주기 리듬을 평가하기 위한 평가도구로, 지난 2주간 환자의 생물학적 리듬에 대해서 평가한다(97). 설문은 총 21가지 항목으로 구성되어 있다. 일주기 리듬의 유지 문제를 수면, 활동, 사회적 활동, 식사패턴, 시간형으로 분류하여 문제의 빈도를 평가한다. 1-4점(1-전혀 그렇지 않음, 2-드물게 있음,

3-때때로 있음, 4-흔히 있음)으로 평가하며, 점수가 높을수록 해당 일주기 리듬에 더 큰 이상이 있음을 의미한다. 한글판 신뢰도 및 타당화 연구에 따르면, 내적 일관성 신뢰도는 0.91이었다 (98).

아침형 종합 척도(Composite Scale of Morningness, CSM)

CSM은 대상자의 일주기 리듬을 평가하기 위해 개발된 설문지로 총 13개의 문항으로 구성되어 있다(99, 100). 각 문항은 대상자가 아침 또는 저녁 활동을 선호하는 지와 관련된 질문으로 구성되어 있으며, 점수가 높을수록 아침 활동형을 의미한다. 3개의 문항은 1-5점으로, 나머지 10개의 문항은 1-4점으로 평가한다. 한글판 신뢰도 및 타당화 연구에 따르면, 내적 일관성 신뢰도는 0.79이었다 (100).

계절성 패턴 평가 설문지(Seasonal Pattern Assessment Questionnaire, SPAQ)

SPAQ는 수면, 사회활동, 기분, 체중, 식욕 및 에너지 수준의 계절적 변화를 평가하는 자가 보고식 설문지로, 6개의 항목으로 구성되어 있다(101, 102). 각 항목은 0-4점으로 평가되며, 6개의 문항의 총합은 전반적 계절성 점수(Global seasonality score, GSS)를 반영한다. 또한, 계절성 변화가 일상생활에 미치는 영향을 0-5점로 평가한다. SPAQ를 기반으로 Kasper et al.(103)연구진들은 계절성 정동장애(Seasonal affective disorder, SAD; GSS≥11점, 중증도≥2)와 역치하 계절성 정동장애(Subsyndromal SAD, S-SAD; GSS 9-10점, 중증도≥2 또는 GSS≥11점, 중증도≥1)로 정의하였다. 본 연구에서는 GSS와 각 항목의 점수, SAD 여부를 변수로 사용하였다.

한국판 세계 보건기구 삶의 질 간편형 척도(WHO Quality of Life-

BREF, WHOQOL-BREF)

WHOQOL-BREF 도구는 전반적인 삶의 질, 신체적 건강상태, 심리적, 사회적 삶의 영역에 대한 만족도를 평가하는 자가 보고식 도구이다(104, 105). 총 26개의 문항으로 첫 두 문항은 전반적인 삶의 질에 대한 평가, 전반적인 건강상태에 대한 만족도를 평가하며, 나머지 24개의 문항은 4개의 하위영역(신체적 건강, 심리적 건강, 사회적 관계, 환경)으로 나뉘어 평가된다. 각 문항은 1-5점으로 평가되며, 점수가 높을수록 각 문항에 긍정을, 점수가 낮을수록 부정을 의미한다. 한글판 신뢰도 및 타당화 연구에 따르면, 내적 일관성 신뢰도는 0.90이었으며 각 영역에서도 높은 일관성을 보였다 (104).

코너-데이비슨 회복력 척도(Connor-Davidson Resilience Scale, CD-RISC)

CD-RISC는 대상자의 회복 탄력성을 평가하는 자가 보고식 설문지로 총 25개의 문항으로 구성되어 있다. Dr. Davison과 Dr. Kathryn이 개발한 척도로 저자에게 사용료를 지불 후 사용허가를 받았으며, 저자가 허가한 한국어판의 척도(내적 일관성 신뢰도 0.93)를 사용 하였다(106, 107). 각 문항은 0-4점으로 평가하며, 점수가 높을수록 회복 탄력성이 높은 것을 반영한다. LPA 모형에는 전구기 환자의 회복 탄력성을 반영하기 위하여 CD-RISC 총점이 투입되었다.

전반적 기능 -사회와 역할 평가(Global Functioning: Social and Role Scales, GF-social, GF-role)

대상자의 기능 저하를 평가하기 위하여 흔히 사용되는 GAF 척도는 전구기 상태가 호발하는 연령(청소년기-초기성인기)에서 특징적으로

관찰되는 기능 저하를 다루지 못하는 단점이 있다(108). GF 도구는 이러한 단점을 보완하고, 고위험 상태의 호발 연령에서 특징적인 기능 저하를 평가하기 위하여 평가영역을 역할, 사회생활로 분류하였다. 또한, 각 영역별로 현재, 지난 한 해 동안의 가장 높은 점수, 낮은 점수를 나누어 평가하여 기능의 변화양상을 파악할 수 있도록 설계되었다. 문항은 1-10점으로 평가되며, 점수가 낮을수록 기능 저하를 반영한다.

그 외 LPA 모형에 포함된 HAM-D, HAM-A 평가 척도는 및 LPA 모형에 포함되지 않은 GAF, CGI-S에 대한 설명은 연구 1에서 기술된 내용과 동일하다.

3. 혈장 단백질 분석

본 연구의 분석에 사용된 고위험군의 혈액 시료는 서울대학교병원 기분장애 클리닉 단일 기관에서 채혈하였다. 또한, 모든 연구 참여자들은 임상 증상 평가와 동일한 일자에 혈액을 채취하였다. 채취한 혈액은 서울대학교병원 의생명연구원 융합연구 협력부의 단백체학 실험실에 전달하였고, 단백체학 실험실에서 혈장 시료의 전처리 및 질량분석을 수행하였다. 혈장 시료의 전처리와 단백질 정량 방법은 연구 1 -3. 혈장 단백질 분석에 기술된 방법과 동일하다.

4. 통계분석

4-1. 잠재프로파일 분석(LPA)

10개의 지표를 기반으로 LPA분석을 통해 잠재집단을 도출하였다. 잠재집단의 계층 수를 결정하기 위하여, 계층 수를 하나씩 증가시키며

통계지수(BIC, saBIC, LMRT, BLRT)[15]를 검토하였다. BIC, saBIC는 그 값이 낮을수록 양호한 모형임을 나타내며, LMRT, BLRT는 유의 수준에 따라 분류된 계층의 유의성을 반영한다. 통계지수를 기반으로 잠재계층별 분류율, 분류된 집단의 해석을 종합적으로 고려하여 집단 수를 선정하였다(109). LPA분석을 위해 Mplus version 8 프로그램을 사용하였다.

4-2. 사회인구학적 변수 및 임상, 단백체 변인의 분석

LPA에서 분류된 세 그룹의 잠재계층 집단의 특성을 파악하기 위해 집단의 사회학적, 임상, 단백체 변인을 통계적으로 비교하였다.

단백체 정량 값의 분포가 정규분포를 만족하지 않고, 편중되어 있어 Log2 변환을 통해 전처리하였고, 이후 Shapiro-Wilk 검정을 통해 정규성을 평가하였다. 범주형 변수는 피어슨 카이제곱 검정(Pearson's chi-squared test), 연속형 변수는 ANOVA 분석하였다. 상관분석은 임상 증상과 단백질의 발현량 간의 상관관계를 평가하기 위해서 사용되었다. 다중 비교를 위해 보정하기 위해 Benjamini-Hochberg false discovery rate(FDR)을 보정한 유의확률 q 값을 사용하였다.

생물정보학 분석은 Database for Annotation, Visualization and Integrated Discovery(DAVID v6.8, http://david.ncifcrf.gov/)를 통해 gene ontology(GO)분석(62)을 수행하였다. 통계분석은 R Statistical Software 버전 4.0.2 (63) 및 IBM SPSS 25.0(IBM Inc., Armonk, NY)를 사용하여 수행되었다.

---

[15] BIC, Bayesian information criteria; saBIC, sample adjusted Bayesian information criteria; LMR, Lo-Mendell-Rubin likelihood adjusted test; BLRT, bootstrap likelihood ratio test

유의 수준은 $p < 0.05$, $q < 0.05$ 양측 검정을 사용하였다.

제 2 절 결과

1. 잠재집단의 분류 및 분류된 집단의 하위 지표 점수 비교

LPA 분석결과, 표2-1의 잠재집단 수에 따른 BIC 값을 보면 집단 수를 4개까지 늘렸을 때도 감소하였다(표2-1). 집단수가 3개일 때 적합도 지수 LMR, BLRT 모형이 유의하였으나(p <0.05), 집단수가 4개일 때의 모형은 유의하지 않았다(표2-1). 따라서 최종적으로 본 연구 2에서는 3개의 잠재집단을 도출하였으며, 세 집단의 세부지표의 점수 및 통계분석결과를 그림2-1에 제시하였다.

그림 2-1에 따라 분류된 잠재집단의 세부점수를 비교해 보면, 잠재집단1은 10개의 점수가 다른 두 잠재집단과 비교하여 모두 낮은 분포를 보였다. 잠재집단 2는 정신증 증상지표인 SIPS 양성, PQ-B 총점 점수 및 우울(HAM-D, QIDS), 불안(HAM-A, STAI-S, STAI-T)의 점수가 다른 두 잠재집단과 비교하여 유의하게 높았다. 반면 조증 지표인 BPSS 하위 척도는 잠재집단 3과 비교하여 낮았다. 잠재집단 3은 다른 두 잠재집단과 비교하여 BPSS 하위척도 점수가 유의하게 높았고, 우울(HAM-D, QIDS), 불안(STAI-S, STAI-T)의 점수가 잠재집단 1보다는 높고, 잠재집단 2보다는 유의하게 낮았다(그림 2-1).

따라서 LPA에서 분류된 세 집단의 임상 요인의 특성을 반영하여, 잠재집단1-낮은 심각도군(Low severity group, LS), 잠재집단2-우울 정신증 우세군(Depressive-psychosis dominant group, DPD), 잠재집단3-조증 우세 군(Manic dominant group, MD)으로 정의하였다.

2. 세 잠재집단의 사회인구학 및 임상특성 요인 비교

표2-2는 LS, DPD, MD 세 군의 사회인구학 특성을 비교한 결과이다. 연구 1(표1-1)의 고위험군 대상자 모집기준에 따라 CHR-PRS로 분류된 집단은 MD(2명, MD의 9.1%)보다 LS(11명, LS의 39.3%), DPD(16명, DPD의 40.0%)으로 MD로 분류되는 빈도가 나머지 두 군과

비교하여 유의하게 낮았다. 과거 약물치료력(현재 중단)의 빈도는 DPD(1명, 2.5%)로 LS(9명, 32.1%)보다 유의하게 낮았다. 벤조디아제핀 약물의 사용 빈도가 LS(1명, 3.6%)에서 DPD(19명, 47.5%), MD(9명, 40.9%)보다 낮았다(표2-2).

표 2-3은 세 잠재집단의 임상특성을 비교한 결과로, CGI-S가 LS 군에서 다른 두 군과 비교하여 유의하게 낮았다. 아동기 외상척도(CTQ)의 하위 척도 중 정서 및 신체 방임의 점수는 DPD군에서 LS군보다 유의하게 높았다. 충동성(BIS) 총점 및 무계획 충동성 점수 또한, DPD군에서 LS군보다 유의하게 높았으며, 주의충동성은 LS군보다 DPD, MD 군에서 유의하게 더 높았다. 생물학적 리듬평가(BRIAN) 총점 및 활동 점수는 DPD군에서 LS보다 더 높았으며, 이는 전반적인 일주기 리듬 및 활동 일주기리듬이 DPD군에서 LS군보다 더 심각한 이상이 있을 수 있음을 시사한다. 사회적 활동 일주기 리듬 평가 점수는 DPD, MD군에서 LS군보다 더 유의하게 높았다. 체중의 계절성(SPAQ)영향은 MD군에서 LS보다 더 유의하게 높았다. 삶에 대한 만족도(WHOQOL-BREF) 총점 및 신체 만족도는 LS, MD, DPD순서로 높았으며, 전반적, 심리적, 사회적 관계 만족도 하위척도는 DPD군이 다른 두 군과 비교하여 유의하게 낮았다. 회복력 점수(CD-RISC)는 DPD군에서 다른 두 군보다 유의하게 낮았으며, 현재 전반적 기능 점수(GAF-present) DPD군에서 LS 군보다 유의하게 낮았으나 증상 발병 전 및 최근 1년간의 GAF점수는 세 군간 유의한 차이가 없었다(표 2-3).

3. 세 잠재집단의 혈중 단백질의 발현량 비교 결과

158개의 표적 펩타이드(92개의 단백질) 중에서 2개의 단백질(FA5, PRDX2)가 $p < 0.05$에서 유의한 차이를 보였으나 $q < 0.05$에서 유의하지

않았다(표 2-4).

4. 임상지표와 단백질 발현량의 상관관계 분석결과

연구 2를 통해 평가된 19개의 임상지표와 158개의 펩타이드(92개의 단백질) 발현량의 상관관계를 분석하였고, 그 중 최소 2개의 임상지표와 유의한 상관관계를 보인 펩타이드를 표 2-5, 그림 2-2에 요약하였다.

C4BPA, PLTP는 7개의 임상지표와 유의한 상관관계를 보였다. 특히 주관적 불안(STAI-T/S)과 우울(HAM-D, QIDS) 지표와는 유의한 양의 상관관계를, 일주기리듬-아침형, 삶의 만족도 및 회복탄력성(WHOQOL-BREF, CD-RISC)와는 유의한 음의 상관관계를 보였다(표 2-5, 그림 2-2).

증상의 도메인로 살펴보면, 우울, 불안지표와 모두 양의 상관관계를 보인 단백질은 LBP, HGFL, A1AG2였으며, LBP는 CSM, WHOQOL-BREF, CD-RISC와 음의 상관관계를, HGLF는 SIPS-일반과 양의 상관관계를, A1AG2는 CSM과 음의 상관관계를 보였다. 그 외 주관적인 우울감(QIDS)와 양의 상관관계를 보인 단백질은 15개[16]로 이 중 4개의 단백질(A1BG, ITIH3, TETN, PZP)이 조증 지표(BPSS-심각도)와 음의 상관관계를 보였다(표 2-5, 그림 2-2). SIPS-양성점수와 음의 상관관계를 보인 단백질은 3개(PRDX2, BGH3, TLN1)으로 PRDX는 조증 지표(BPSS-심각도, BPSS-빈도)와도 음의 상관관계를 보였다. BGH3는 CSM과 양의 상관관계를, TLN1은 BPSS-심각도와 양의 상관관계를 보였다(표 2-5, 그림 2-2). PRGP2는 불안(HAM-A),

[16] C4BPA, PLTP, LBP,, A1BG, CRP, TPM4, VTNC, CERU, ITIH3, ECM1, B2MG, SAMP, TEN, PZP

SIPS-음성, BIS점수와 음의 상관관계를 보였으며, ITIH4도 SIPS-음성, BIS점수와 음의 상관관계를 보였다(표 2-5, 그림 2-2).

5. 생물정보학(Bioinformatics) 분석결과

대부분의 혈소판 탈과립 과정과 관련이 있었으며, 세포 단백질 대사과정, 급성기 반응, 엔도펩티다아제 활성의 억제 조절과 관련 있었다. 세포 성분 분석결과는 대부분의 단백체 마커가 세포외 공간의 구성 및 소포체와 관련 있었다. 분자 기능 분석에 따르면 엔도펩티다아제 억제제의 활성과, KEGG경로는 보체 및 혈액 응고 연쇄반응과 관련이 있었다(그림 2-3).

6. 결과 요약

정신증, 조증, 우울, 불안척도를 사용하여 잠재계층 분석결과 3개의 계층으로 분류 되었으며, 분류결과 전반적으로 낮은 중증도의 LS군, 정신증과 우울, 불안의 기분증상이 동반된 DPD군, 조증이 주된 증상이나 우울, 불안의 기분증상이 동반된 MD군으로 분류 되었다.

세 잠재집단에서 $p < 0.05$에서 유의한 차이를 발현량 차이를 보인 단백질은 2개(PRDX2, FA5)였으나, $q < 0.05$에서는 유의한 차이를 보이지 않았다(표2-4). LS군과 비교해 MD군에서 유의하게 감소한 PRDX2는 조증의 중증도(BPSS 심각도, BPSS 빈도) 유의한 음의 상관관계 SIPS 양성척도와 유의한 음의 상관관계를 보였다.

제 3 절 고찰

이질적이고 비교적 불명확한 증상이 관찰되는 고위험 대상자들의

다양한 증상 도메인을 종합하고, 이질적인 특성을 유형화하기 위하여 총 4개의 영역(10개의 지표)를 이용하여 프로파일 분석(Latent profile analysis, LPA)을 수행하였다. 그 결과 선행 분류 프로토콜의 범주(CHR-PRS, CHR-BD, CHR-MDD, CHR-PL, CHR-Extended)으로 구분되지 않고, 전반적으로 낮은 중증도의 LS군, 정신증과 우울, 불안의 기분증상이 동반된 DPD군, 조증이 주된 증상이나 우울, 불안의 기분증상이 동반된 MD 군으로 분류 되었다. 특히 분류된 LS군은 고위험 상태의 대상자들의 임상 상태를 단순히 진단명으로 구분하지 않고, 표현형과 중증도로 구분하는 단계모델(Staging Model, (11, 110, 111)을 반영하는 결과일 수 있다. 또한, 이 단계 모델(112, 113)은 무증상 및 전구기 단계에서 조기에 질병의 경과 상태를 반영할 수 있는 생물학적, 임상적, 환경적 마커의 통합을 강조한다. 따라서 연구 2에서 발견된 증상 심각도를 반영하는 단백체 마커들은 전구기 환자들의 질병단계를 예측하는 마커로 활용될 수 있을 것으로 보인다.

연구 2에서 발견된 단백질 의미 있는 단백질들의 대부분은 또한, 선행연구들을 통하여 정신질환 환자와 관련된 단백체 마커로 보고된 마커들이었다. 증상의 중증도와 관련된 단백질들을 선행연구 및 생물정보학 분석을 기반으로 관여하는 경로 및 기능에 따라 분류하여 살펴보면 다음과 같다.

첫째, 가장 많은 단백질들(12개, 38.7%)이 혈소판의 탈 과립화 및 혈액 응고 연쇄 반응과 관련되었다. 정신적 스트레스는 카테콜아민과 세로토닌 경로를 자극하여, 혈액 응고 연쇄반응을 활성화 하는 것으로 알려져 있다(70). 또한, 이 경로와 관련된 단백질들 중 KNG1(15, 71), A1BG(114), VWF(115), ITIH3(116), TETN(116) 단백질은 선행연구에서 정신질환과의 관련성이 보고된 바 있다. 대부분의 혈액

응고반응과 관련된 단백질들은 증상의 심각도와 양의 상관관계를 보였다. 따라서 혈액 응고반응과 관련된 단백체 마커의 증가는 질병의 악화 및 중증 상태임을 반영하는 것임을 시사한다.

또한, 면역반응의 이상은 주요정신질환의 관계는 선행연구들에서 보고 되어 왔다(65, 72). 면역반응과 관련된 대부분의 단백질들[17] 증상이 심각도와 단백질 발현량이 양의 상관관계를 보였으며, 선행연구에서 정신질환과의 관련성이 보고된 단백질이었다 (15, 65, 115, 117, 118). 반면, PRGP2(PGLYRP2)는 임상 증상이 더 심할수록 발현량이 더 감소 하였다. PRGP2는 세균 세포벽 성분인 펩티도글리칸(peptidoglycan)을 분해 에 관여하는 것으로 알려져 있으며(119). 선행연구에서 정상 대조군과 비교하여 주요정신질환 환자군 유의한 발현량의 차이가 보고된 바 있다(15, 65). 또한, 펩티도글리칸 인식 단백질은 장-뇌의 신호를 매개하여, 뇌의 발달에 영향을 미칠 수 있다는 동물연구가 보고된 바 있다(120). 따라서 선행연구 및 본 연구의 결과를 종합해 볼 때 PRGP의 외부의 침입에 반응하는 보호인자의 감소를 반영하며, 그로 인한 정신증상의 악화를 반영할 수 있음을 시사 한다.

콜레스테롤 대사의 초기 변이 또한, 많은 정신질환 연구에서 보고되고 있으며, 혈청 콜레스테롤의 감소는 신경수초와 시냅스 형성 및 발달과 관련된 기능 장애와 관련된다(65). 콜레스테롤 대사와 관련된 APOA1, APOB 단백질은 발현량이 감소할수록 증상 심각도와 유의한 음의 상관관계를 보였으며, 이는 지질대사의 저하가, 신경수초의 및 시냅스 형성의 저하에 영향을 미쳐, 정신증상의 악화로 이어질 수

[17] LBP, CRP, SAMP(APCS), C4BPA, B2MG(B2M), HGFL(MST1)

있음을 시사하는 결과이다.

주된 생물학적 경로 이상 외에도 외부 요인 및 신경에 보호 하는 물질로 알려진 물질들이 증상의 심각도와 음의 상관관계를 보였다. 세포사멸을 억제하는 GSN은 SIPS-음성증상과 음의 상관관계를 보였으며(76), 신경독성물질을 분해하는 CHLE는 및 항산화기능의 PRDX는 조증 및 일주기 증상의 심각도와 음의 상관관계를 보였다 (65, 77, 78). 따라서 이러한 보호인자들의 감소는 질병의 경과에 부정적인 영향을 주며, 증상악화로 이어질 수 있음을 시사한다.

그 외 구리운반 단백질인 CERU(CP), 비타민 D운반 단백질인 VTDB(GC), 인지질 운반단백질(PLTP)은 증상의 심각도와 양의 상관관계를 보였으며, 선행연구에서 정신질환과의 관련성이 보고된 바 있다(15, 65, 114). 그 외 본 연구에서 통계적으로 증상의 심각도와 유의한 상관관계를 보였으나, 선행연구에서 보고된 적 없는 VTNC, LYAM1(SELL), TPM4, SPP2, ECM1, TLN1, TPM4 물질에 대해서는 추가적인 연구가 필요할 것으로 보인다.

본 연구의 한계점으로는 단면적으로 모집된 혈액시료를 분석에 사용하였으므로, 이러한 단백질 변이가 실제 질병 진행 및 단계의 인과관계를 반영하는 마커로 확정하기에는 어려움이 있다. 또한, 단백질발현에 영향을 미칠 수 있는 교란요인 들에 대해서는 영향력을 통제하지 못하였다. 그러나 연구 2는 이질적 성격의 고위험군을 증상 특성을 유형화하고, 증상의 심각도를 반영하는 의미 있는 혈중 단백체 마커를 발견하였다는 점에서 매우 의미가 있겠다, 또한, 고위험군 환자의 질병 진행 관여하는 생물학적 경로 이상을 탐색하였다는 점에서 의미가 있다.

제 4 장 – 연구 3. 주요정신질환 고위험군의 임상 경과와 관련된 혈액 단백체 마커 탐색

연구3을 통해 첫 방문 시 채혈 및 정량된 단백체 마커 중

주요정신질환 고위험군의 경과를 예측할 수 있는 단백체 마커를 탐색하고자 하였다. 뿐만 아니라 초기 면담 시 평가 및 보고된 고위험군 환자들의 임상평가자료를 활용하여, 고위험군의 임상 경과를 예측할 수 있는 임상적 마커에 대해서도 탐색하고자 하였다. 이를 혼합 모델(임상+단백체 모델)을 구성하고, 이질적인 증상을 호소하는 고위험군 대상자가 내원하였을 때, 단백체 마커를 활용하는 경우 모델의 예측력 이 향상될 수 있는지 확인하고자 하였다.

고위험군의 임상 경과는 다음의 두 가지 기준으로 하였다.

기준 1) 경과 중 SCID를 통하여 주요정신질환으로 진단되는 경우,

기준 2) 정기적으로 평가한 CGI-S 점수를 기준으로, 마지막 평가 시 증상 호전의 여부

이 경과 결과(Outcome)를 기반으로 하여, 초기 단백체 정보가 평가 이후 주요정신질환으로의 발병 여부뿐만 아니라, 적극적인 약물치료 개입이 필요한 고위험군 대상자인지 구분해낼 수 있는지 확인하고자 하였다. 기준 1)을 결과로 하는 Model 1, 기준 2)를 결과로 하는 Model 2, 총 2개의 결과를 예측하는 분석을 나누어 진행하였다. 각 모델은 a. 임상 변인들이 투입된 모델, b. 단백체 변인들이 투입된 모델, c. 모든 변인(임상+단백체 변인)이 포함된 혼합 모델로 다시 나누어 분석하였다. 연구3의 분석결과를 통해 다양한 임상 증상을 호소하는 이 호소하는 고위험군의 임상진단, 중증도와 관련된 혈액 단백체 마커를 발굴하고, 경과 예측 모델을 발견하고자 하였다.

제 1 절 연구 대상 및 방법

1. 연구 대상자

연구 1에 기술된 주요정신질환 고위험군의 대상자 90명이 연구

3의 연구에 포함되었다.

2. 연구 방법

주요정신질환으로 등록된 고위험군 대상자는 초기 평가 후 2개월 간격으로(초기 평가 후 24개월까지) 대면 또는 유선으로 정신건강의학과 전문의와 면담을 통해 평가를 받았다. 면담을 통해 면담의는 대상자가 주요정신질환으로 발병하였는지 평가하였으며, 발병이 의심되는 환자는 대면평가를 통하여 발병 여부를 확인하였다. 연구 종결 시점에서 대상자가 연락되지 않아 평가가 조기 종결된(drop-out) 대상자들을 대상으로 임상 경과를 확인하기 위하여 유선평가를 추가 실시하였다.

3. 임상증상평가도구

3-1. 임상 경과평가도구

SCID

면담자는 대면 또는 유선평가에서 주요정신질환의 발병이 의심되는 고위험군 대상자를 구조화된 면담 도구인 SCID를 통해 발병 여부를 확인하였다.

CGI-S

면담자는 2개월 간격으로 고위험군 대상자의 임상 상태를 CGI-S 평가 척도를 이용하여 평가하였다.

3-2. 초기 평가도구 및 예측 모델에 사용한 임상 변인(표 3-1)

주요정신질환의 임상 경과 예측 모델에 사용된 사회 인구학적 변인 및 초기 평가 임상 변인은 표 3-1에 정리하였다. 각 평가도구(SIPS, PANSS, HAM-D, HAM-A, YMRS, BRIAN, CGI-S, GAF, GF, STAI, CD-RISC, WHOQOL-BREF, CSM, PQ-B, BIS, CTQ, SPAQ) 에 대한 기술은 연구 1, 2와 동일하다.

BPSS-AP

연구 3에서는 각 문항은 최근 1개월간의 증상과 살면서 경험한 증상 두 기간에 대하여 각각 질문한 응답을 사용하였다. 연구 2에 기술된 요인 외에도 2개의 요인(조증의 긍정적인 측면-예, 들뜬 기분, 에너지 증가, 조증의 부정적인 측면-예, 예민, 위험 행동 추구)으로 나누어 각 증상의 심각도를 구분하여 변인에 포함하였다(89).

4. 혈장 단백질 분석

혈장 시료의 전처리와 단백질 정량 방법은 연구 1, 연구 2에 기술된 방법과 동일 하며, 총 158개의 펩타이드를 MRM-MS 방법으로 정량화하였다. 정량 분석 및 통계 모델에 사용한 158개의 단백질에 대한 정보는 표 3-2와 같다.

5. 통계분석

5-1. 사회인구학적 변수 및 임상, 단백체 변인의 집단간 비교 분석

단백체 정량 값의 분포가 정규분포를 만족하지 않고, 편중되어 있어 Log2 변환을 통해 전처리하였고, 이후 Shapiro-Wilk 검정을 통해 정규성을 평가하였다. 범주형 변수는 피어슨 카이제곱 검정(Pearson's

chi-squared test), Fisher의 정확한 검정으로, 연속형 변수는 독립 표본 t 검정 및 Mann-Whiteny의 U검정으로 분석하였다. 다중 비교를 위해 보정하기 위해 Benjamini-Hochberg false discovery rate(FDR)을 보정한 유의확률 q 값을 사용하였다.

5-2. 변인 선택(Model selection)

고위험군 대상자의 경과를 반영하는 임상 및 단백체 마커를 선별하기 위하여, 총 268개의 변인(임상 변인 110개, 단백체 변인 158개)을 분석에 포함하였다(표 3-1, 표 3-2).

Penalized 회귀분석은 분석에 사용되는 변수의 수가 많고, 상대적으로 적은 샘플 수일 때 발생할 수 있는 과적합 오류를 줄이는 분석법이다. Penalized 회귀분석법 중 하나인 최소 절대 수축 및 선택 연산(Least absolute shirinkage and selection operator, LASSO) 분석법은 모델에 의미 있는 변수를 선택하고 과적합을 줄일 수 있는 회귀분석법이다. 구분력이 낮은 변인들의 회귀계수는 0으로 모형에서 제거되며, 계수가 0이 아닌 변수들만 선택된다. 본 연구에서는 예측 모델에 필요한 변수들을 선택하기 위하여 LASSO 분석을 사용하였으며, R 소프트웨어의 “glmnet”패키지를 사용하였다. 또한, 과적합을 더욱 줄이기 위하여 10배 교차검증(10-fold cross validation, 10-CV)을 사용하여 가장 적합하고 정규화된 모델을 산출하는 수축 조절 변수 람다(λ)값을 선택하였다. 데이터는 무작위로 10개의 단위로 분할되었고, 10개 중 9개의 단위가 학습 데이터(Training data)로, 나머지 1 단위의 데이터는 테스트 데이터(Test data)로 교차검증을 위해 사용되었다.

또한, 선행연구의 분석법을 기반으로(23, 121, 122) 학습

데이터에서 10-CV-LASSO 회귀분석을 100번 반복하여, 100개의 모델을 생성하였다. 이를 통해 10-CV로 무작위 분할로 인한 데이터의 작은 변화에 대한 모델선택의 민감도를 평가하여 모델선택의 불확실성에 대하여 평가를 조사할 수 있었다. 각 변수들에 대해서는 생성된 100개의 모델에서 변수가 선택된 비율을 계산하였고, 이를 선택 부분(Selection fraction, SF)으로 정의하였다. SF는 0-1 사이의 값으로, 모델에서 각 변인의 상대적 중요성을 반영하였다.

5-3. Akaike information criterion(AIC) and Akaike weights(w)

AIC는 모형이(상대적 척도로) 주어진 데이터에 얼마나 근접한지를 평가하며, AIC값이 가장 낮은 모형이 최적의 모형으로 간주 된다(123-125). 또한, 변수의 수(k)와 비교해 상대적으로 표본 크기(n)가 적으므로, bias-corrected 버전의 AICc를 선택하였다(126, 127). 각 모델에 대하여 Akaike 가중치(w) 값을 계산하였고, w를 각 모델이 이 데이터의 최적모델에 근접할 확률로 정의하였다(124). w는 0-1 사이의 값으로 모든 모델의 가중치 합은 1이었다. 각 모델에 대해 선택된 변수 조합이 최적모델일 확률을 추정하기 위해 해당 모델의 w를 합산하였다(122).

AICc 및 Akaike 가중치(w) 계산에 사용되는 공식은 다음과 같다(124, 125).

$$AIC = -2L + 2k$$

(L: log-likelihood, k: 모델에서 선택된 변수의 개수)

본 연구에서는 사용된 AICc(bias-corrected 버전의 AIC)의 공식은 다음과 같다(126, 127)

$$AICc = AIC + \frac{2k(k+1)}{(n-k-1)}$$

(n: 표본 수, k: 모델에서 선택된 변수의 개수)

개별 모델(Model i)의 Akaike 가중치(wi)는 AICc와 가장 낮은 AICc의 차로 계산되었다.

$$\triangle_i = AICc_i - AICc_{min}$$

$$wi = \frac{\exp(-\frac{1}{2}\Delta i)}{\sum_{r=1}^{R}\exp(-\frac{1}{2}\Delta r)}$$

통계분석은 R Statistical Software 버전 4.0.2 (63) 및 IBM SPSS 25.0(IBM Inc., Armonk, NY)를 사용하여 수행되었다.

유의 수준은 $p < 0.05$, $q < 0.05$ 양측 검정을 사용하였다.

5-4. 변수 추출과 모델 평균화

각 모델의 wi를 산출하여 w > 0.9인 높은 중요도를 보이는 고유 모델은 조절 모수(λ) 추정과 예측을 해당 단일모델만을 기반으로 할 수 있다 (128). 이 경우 이 고유 모델의 각 변수들의 회귀계수는 해당 모델들의 평균으로 계산되었다.

그러나, 높은 중요도를 보이는 고유의 단일 모델이 부재한 경우, 이는 모델의 선택의 불확실성을 반영하는 것이므로, 결과 예측의 재현성을 확보하기 위하여 모든 100개의 모델을 모두 사용하여 변수 추출 및 모델 평균을 산출하였다. 이 경우 SF이 0.8이상인 변수들만 모델에 포함하여 과적합을 제한하였고, 각 변수의 적합한 회귀계수 추정치를 얻기 위하여 100개 모델에 대한 평균값으로 산출하였다.

5-5. 예측성능(Predictive performance)

모델의 예측성능 평가는 훈련데이터와 테스트 데이터에 각각 적용하였으며, area under the receiver operating characteristics(AUROC, R소프트웨어의 RORC패키지(129))를 사용하여 평가하였다. 각 모델의 AUC 성능 평가의 기준은 다음과 같다(0.9-1: 우수, 0.8-0.9 : 양호, 0.7-0.8 : 보통, 0.6-0.7: 불량, 0.5-0.6: 실패)(130).

제 2 절 결과

주요정신질환 고위험군 증상의 임상 경과와 관련된 임상 증상 및 혈액

단백체 마커를 선별하고, 예측 모델을 도출하기 주요정신질환으로의 발병 여부, 증상 회복 유무(마지막 평가한 CGI-S 점수 기준)를 결과 변수로 설정하여 모델을 평가하였다. 연구 3에서 분석한 모델의 종류와 사용된 데이터에 대한 기술은 표 3-3에 정리하였다.

1. Model 1 -주요정신질환으로의 발병 여부 예측 모델

주요정신질환으로의 발병된 대상자는 총 14명(15.6%)로, 이 중 11명이 제 2형 양극성장애로, 2명이 조현병, 1명이 주요우울장애로 발병하였다. 조현병 환자의 초기 평가 후 발병까지의 평균 기간과 표준편차는 4.0±2.8개월이었으며, 제 2형 양극성장애는 14.2±9.1개월, 주요우울장애는 12개월 이었다. 따라서 예측 모델의 결과 변수로 발병군(Transition group, CHR-T 14명), 비발병군(Non-transition group, CHR-NT 76명)이 분석되었다. Model 1a는 110개 임상 변인을 예측변수로 사용하였고, Model 1b는 158개의 펩타이드를 예측변수로, Model 1c는 268개의 모든 변인(임상+단백체)을 예측변수로 사용하였다(표3-3).

1-1. 주요정신질환 발병군(CHR-T)과 비발병군(CHR-NT) 의 사회인구학 및 임상특성 요인 비교

표 3-4는 두 군의 사회인구학 및 임상특성을 비교한 결과이다. CHR-T(12명, 85.7%)에서 CHR-NT(28명, 36.8%)군보다 초기 평가 시 CHR-BD로 진단된 빈도가 높았으며, 초기 평가 이후 연구에 참여한 기간이 CHR-T군(23.4±9.5)에서 CHR-NT군(17.1±10.0) 보다 더 길었다. 그 외의 나이, 성별, 치료력 및 약물사용력의 두 군간 유의한 차이는 관찰되지 않았다(표3-4).

두 군의 초기 임상평가 척도의 점수를 비교하였을 때 BPSS-현재 심각도, BPSS-현재 T1 threshold, BPSS-과거 T1 threshold, BPSS-현재-부정적 조증과 관련된 증상의 점수가 CHR-T군에서 CHR-NT군보다 높았다(표3-5). 그 외에 임상평가 척도에서 유의한 차이를 보인 임상특성은 없었다(표3-5).

1-2. 주요정신질환 발병군(CHR-T)과 비발병군(CHR-NT) 의 혈중 단백질의 발현량 비교 결과

158개의 표적 펩타이드(92개의 단백질) 중에서 19개의 펩타이드(17개의 단백질)가 $p < 0.05$에서 유의한 차이를 보였으나 $q < 0.05$에서 차이를 두 군간 유의한 단백체 마커의 차이는 관찰되지 않았다(표3-6, 그림3-1).

1-3. 주요정신질환 발병 예측 모델(Model 1) 분석결과

1-3A. Model 1a(모델의 선택-110개의 임상 변인)

110개의 임상 변인을 100 repeated 10-CV LASSO 회귀모델로 분석한 결과 모델의 중요도(Model probability, w) >0.9 이상인 Unique 모델은 관찰되지 않았다(표3-7A). 총 110개의 임상 변인 중 0-32개의 변인이 모델에서 선택되었으며, 단 한 번도 선택되지 않은 변수는 76개(69.1%)였다(그림3-2 Model 1a). SF ≥0.8 인 변수의 개수는 9였다. 선택된 변수들의 조합을 기반으로 Unique 모델을 관찰하였을 때 27개의 Unique 모델이 관찰되었다(표3-7A). 28개의 변수로 구성된 모델이 가장 높은 빈도수(100번 중 24번)를 보였으나 모델의 중요도(w)는 9.85×10^{-9}로 높지 않았다(표3-7A). 반면, 가장

높은 중요도를 보인 모델(2개의 변인)은 선택된 빈도가 100번 중 5번이었으나 w 0.367로 가장 높은 중요도를 보였다(표3-7A). 따라서 선택된 모델의 빈도수가 모델의 중요도를 반영하지 않음을 확인할 수 있었다. 따라서 100번의 모델에서 80번이상 선택된 9개의 변인(SF ≥0.8)을 최종 Model 1a의 변인으로 선택하였다(표3-8).

1-3B. Model 1b(모델의 선택-158개의 단백체 변인)

158개의 임상 변인을 100 repeated 10-CV LASSO 회귀모델로 분석한 결과 모델의 중요도(Model probability, w) >0.9 이상인 Unique 모델은 관찰되지 않았다(표3-7B). 총 158개의 단백체 변인 중 0-32개의 변인이 모델에서 선택되었으며, 단 한 번도 선택되지 않은 변수는 125개(79.1%)였다(그림3-2 Model 1b). SF ≥0.8 인 변수의 개수는 15였다. 선택된 변수들의 조합을 기반으로 Unique 모델을 관찰하였을 때 27개의 Unique 모델이 관찰되었다(표3-7B). 30개의 변수로 구성된 모델이 가장 높은 빈도수(100번 중 23번)를 보였으나 모델의 중요도(w)는 $1.50X10^{-9}$로 높지 않았다(표3-7B). 따라서 선택된 모델의 빈도수가 모델의 중요도를 반영하지 않음을 확인할 수 있었다. 따라서 Model 1b에서도 100번의 모델에서 80번이상 선택된 15개의 변인(SF ≥0.8)을 최종 Model 1b의 변인으로 선택하였다(표3-8).

1-3C. Model 1c(모델의 선택-268개의 임상 및 단백체 변인)

268개의 모든 변인(110개의 임상 변인+158개의 단백체 변인)을 100 repeated 10-CV LASSO 회귀모델로 분석한 결과 모델의 중요도(Model probability, w) >0.9 이상인 Unique 모델은 관찰되지

않았다(표3-7C). 총 268개의 임상 변인 중 0-36개의 변인이 모델에서 선택되었으며, 단 한 번도 선택되지 않은 변수는 227개(79.4%)였다(그림3-2 Model 1c). SF ≥0.8 인 변수의 개수는 7개(5개의 임상 변인 + 2개의 단백체 변인)였다. 선택된 변수들의 조합을 기반으로 Unique 모델을 관찰하였을 때 29개의 Unique 모델이 관찰되었다(표3-7C). 35개의 변수로 구성된 모델이 가장 높은 빈도수(100번 중 13번)를 보였으나 모델의 중요도(w)는 $2.91X10^{-17}$로 높지 않았다(표3-7C). 따라서 선택된 모델의 빈도수가 모델의 중요도를 반영하지 않음을 확인할 수 있었다. 따라서 Model 1c에서도 100번의 모델에서 80번이상 선택된 7개의 변인(SF ≥0.8)을 최종 Model 1c의 변인으로 선택하였다(표3-8).

Model 1a-1c의 성능을 학습데이터(72명)와 테스트 데이터(18명) AUC로 확인하였다(그림3-3). SF ≥0.8인 변수들을 선택하여 각 모델의 성능을 비교하였을 때(그림 3-3a), Model 1a의 AUC는 학습데이터 0.97, 테스트데이터에서 0.90으로 모두 '우수'한 성능을 보였다. 반면 Model 1b의 AUC는 학습데이터에서 1.00(우수)였으나, 테스트데이터에서는 0.67로 불량한 성능을 보였다. 이는 학습데이터에서 과적합한 모델이 선택 되었음을 반영한다. Model 1c의 AUC는 학습데이터 0.95, 테스트데이터에서 0.93으로 모두 '우수'한 성능을 보였다(그림 3-3a). 모델의 과적합 오류를 줄이기 위하여 SF ≥0.85인 변수들을 선택하여 모델의 성능을 비교하였으나, Model 1b의 학습데이터는 정확도 0.95, 학습데이터의 정확도는 0.53으로 여전히 과적합 오류를 보였다(그림 3-3b). 마지막으로 SF ≥0.8인 변수들 중

발병군과 비발병군에서 유의한 통계적 차이를 보인 변수들(표 3-5, 3-6)을 선별하여 모델의 성능을 비교 하였을 때, Model 1b의 학습데이터의 정확도는 0.79로 이전 모델의 정확도 보다 낮았으나 테스트데이터의 정확도는 0.87로 양호한 정확도를 보였다(그림 3-3c).

2. Model 2 –주요정신질환의 증상 경과 예측 모델

초기 평가 후 1회 이상의 경과 평가를 한 대상자 83명(92.2%, 전체 90명 중)로 마지막 평가 시 CGI-S를 기준으로 CGI-S가 2점 이하인 군을 증상 호전 군(Recovered), CGI-S가 3점 이상인 군을 증상 비호전군(Not-recovered)으로 분류 하였다. 분류 결과 호전군은 39명(47.0%), 비호전군은 44명(53.0%)으로 분류 되었다. 본 분석에서는 예측 모델의 결과 변수로 증상 호전 유무로 하여 분석하였다. Model 2a는 110개 임상 변인을 예측변수로 사용하였고, Model 2b는 158개의 펩타이드를 예측변수로, Model 2c는 268개의 모든 변인(임상+단백체)을 예측변수로 사용하였다(표3-3). 마지막으로 Model 2d는 Model 2a, Model 2b에서 선택된(SF ≥0.8) 임상, 단백체 마커들을 포함하여 Model의 성능을 평가하였다.

2-1. 증상 호전군(Recovered)과 비호전군(Not recovered) 의 사회인구학 및 임상특성 요인 비교

표 3-9는 두 군의 사회인구학 및 임상특성을 비교한 결과이다. 남성의 빈도가 증상 비호전군(25명, 56.8%), 호전군(10명, 25.6%)으로 비호전군에서 유의하게 더 많았다(p=0.007). 초기 평가 이후 연구에 참여한 기간이 호전군(22.5±8.1)에서 비호전군(16.6±9.4)보다 더 길었다(p=0.003). 그 외의 나이, 성별, 치료력 및 약물사용력의 두

군간 유의한 차이는 관찰되지 않았다(표3-9).

두 군의 초기 임상평가 척도의 점수를 비교하였을 때, 증상 비호전군에서 호전군보다 양성증상의 심각도(SIPS-양성), 불안의 심각도(HAM-A), 불규칙한 생활리듬(BRIAN-총점, BRIAN-식이)의 점수가 높았으며, PQ-B threshold 점수 이상인 빈도가 높았다. 반면 초기 평가시의 GAF점수, 삶의 만족도(WHOQOL-BREF 총점, 신체, 정신심리학), 회복탄력성(CD-RISC), 아침 활동형(CSM)의 점수가 증상 호전군에서 비호전군보다 더 높았다(표3-10). 초기 평가시의 CGI-S점수는 증상 호전군(3.4±0.7)과 비호전군(3.7±0.8)에서 유의한 차이가 없었으나(p=0.103), 마지막 평가시의 CGI-S점수는 증상 호전군(1.6±0.5)과 비호전군(3.6±0.8)에서 유의한 차이를 보였다(표3-10).

2-2. 증상 호전군(Recovered)과 비호전군(Not-recovered)의 혈중 단백질의 발현량 비교 결과

158개의 표적 펩타이드(92개의 단백질) 중에서 8개의 펩타이드(6개의 단백질)가 p <0.05에서 유의한 차이를 보였으나 q < 0.05에서 차이를 두 군간 유의한 단백체 마커의 차이는 관찰되지 않았다(표3-6). 두 군간 단백질 발현량의 차이를 비교하였을 때 대부분의 단백질들이 호전군에서 비호전군보다 발현량이 감소한 패턴을 보였다(그림 3-4)

2-3. 주요정신질환의 증상 경과 예측 모델(Model 2) 분석결과

2-3A. Model 2a(모델의 선택-110개의 임상 변인)

110개의 임상 변인을 100 repeated 10-CV LASSO 회귀모델로

분석한 결과 모델의 중요도(Model probability, w) >0.9 이상인 Unique 모델은 관찰되지 않았다(표3-12A). 총 110개의 임상 변인 중 0-46개의 변인이 모델에서 선택되었으며, 단 한 번도 선택되지 않은 변수는 58개(52.7%)였다(그림3-5 Model 2a). SF ≥0.8 인 변수의 개수는 8였다. 선택된 변수들의 조합을 기반으로 Unique 모델을 관찰하였을 때 28개의 Unique 모델이 관찰되었다(표3-12A). 43개의 변수로 구성된 모델이 가장 높은 빈도수(100번 중 11번)를 보였으나 모델의 중요도(w)는 $3.62X10^{-31}$로 매우 낮았다(표3-12A). 따라서 모델의 선택된 빈도수가 모델의 중요도를 반영하지 않음을 확인할 수 있었다. 따라서 100번의 모델에서 80번이상 선택된 8개의 변인(SF ≥0.8)을 최종 Model 2a의 변인으로 선택하였다(표3-13).

2-3B. Model 2b(모델의 선택-158개의 단백체 변인)

158개의 임상 변인을 100 repeated 10-CV LASSO 회귀모델로 분석한 결과 모델의 중요도(Model probability, w) >0.9 이상인 Unique 모델은 관찰되지 않았다(표3-12B). 총 158개의 단백체 변인 중 0-50개의 변인이 모델에서 선택되었으며, 단 한 번도 선택되지 않은 변수는 102개(64.6%)였다(그림3-5 Model 2b). SF ≥0.8 인 변수의 개수는 8였다. 선택된 변수들의 조합을 기반으로 Unique 모델을 관찰하였을 때 48개의 Unique 모델이 관찰되었다(표3-12B). 48개의 변수로 구성된 특정 모델이 가장 높은 빈도수(100번 중 10번)를 보였으나 모델의 중요도(w)는 $7.80X10^{-56}$로 매우 낮았다(표3-12B). 따라서 선택된 모델의 빈도수가 모델의 중요도를 반영하지 않음을 확인할 수 있었다. 따라서 Model 2b에서도 100번의 모델에서 80번이상

선택된 8개의 변인(SF ≥0.8)을 최종 Model 2b의 변인으로 선택하였다(표3-13).

2-3C. Model 2c(모델의 선택-268개의 임상 및 단백체 변인)

268개의 모든 변인(110개의 임상 변인+158개의 단백체 변인)을 100 repeated 10-CV LASSO 회귀모델로 분석한 결과 모델의 중요도(Model probability, w) >0.9 이상인 Unique 모델은 관찰되지 않았다(표3-7C). 총 268개의 총 변인 중 0-48개의 변인이 모델에서 선택되었으며, 단 한 번도 선택되지 않은 변수는 207개(77.2%)였다(그림3-5, Model 2c). SF≥0.8 인 변수의 개수는 14개(9개의 임상 변인 + 5개의 단백체 변인)였다. 선택된 변수들의 조합을 기반으로 Unique 모델을 관찰하였을 때 51개의 Unique 모델이 관찰되었다(표3-12C). 모델 2a, 2b와 달리 Unique 모델간의 선택된 빈도수가 모델간 큰 차이는 없었으며 각 모델은 1-6번의 빈도로 선택되었다(표3-12C). 유의미한 중요도를 보이는 Unique 모델이 관찰되지 않았으므로, Model 1c에서도 100번의 모델에서 80번이상 선택된 14개의 변인(SF ≥0.8)을 최종 Model 2c의 변인으로 선택하였다(표3-13).

Model 2a-2d의 성능을 학습데이터(66명)와 테스트 데이터(17명) AUC로 확인하였다. 먼저, SF ≥0.8인 변수들을 선택하여 각 모델의 성능을 비교하였을 때(그림 3-6a), Model 2a의 AUC는 학습데이터 0.92으로 '우수'이었으나, 학습데이터의 AUC는 0.63 '불량'의 성능을 보였다. 또한, Model 2b의 AUC는 학습데이터 0.88 '양호', 테스트 데이터 0.59으로 '불량'한 성능을 보였다. Model 2c의 AUC는

학습데이터에서 0.96(우수)였으나, 테스트데이터에서는 0.70로 '양호'한 성능을 보였다. Model 2d의 AUC는 학습데이터에서 0.99(우수)였으나, 테스트데이터에서는 0.71로 '양호'한 성능을 보였다. 이는 학습데이터에서 과적합한 모델이 선택 되었음을 반영한다(그림 3-6a). 모델의 과적합 오류를 줄이기 위하여 SF ≥0.85인 변수들을 선택하여 모델의 성능을 비교하였을 때, Model 2a-2d의 훈련데이터의 정확도는 0.84-0.99로 '양호-우수'의 정확도를 보였고 테스트 데이터의 정확도는 Model 2a를 제외하고 모두 0.7 이상으로 보통의 정확도를 보였다(그림 3-6b).

제 3 절 고찰

연구 3에서는 주요정신질환 고위험군 대상자들의 임상 경과를 예측할

수 있는 임상 및 단백질 변인을 발굴하고, 예측 모델을 구축하기 위하여 LASSO 회귀분석을 통하여 변수선택, 모델성능 평가를 수행하였다.

주요정신질환의 발병여부를 예측하는 임상 및 단백체 예측 마커를 분석한 모델 1의 결과는 다음과 같다. 모델 1a 분석을 통해 선택된 주요정신질환의 발병과 관련된 초기 평가 임상 변인은, 초기 평가 시 CHR-BD로 분류 될 수록, 약물치료의 과거력이 있으나 평가 당시 복용을 중단한 경우, 평가 당시 기분조절제를 복용력, 정신질환의 가족력, 연구 참여 기간, BPSS-T1 threshold, HAM-D 무의욕/지연, WHOQOL-BREF 정신적, CD-RISC는 Model 1a 분석에서 선택된 임상 변인이었다. Model 1b와 Model 1c를 통해 선택된 단백체 마커는 총 15개로 면역반응 및 염증반응[18], 응고연쇄반응[19]과 관련된 단백질이었다. 정신질환과 질환에서 병태생리학적으로 혈액 응고 연쇄반응을 활성화는 선행연구를 통해 보고된 바 있으며(70), 면역반응의 이상 또한, 선행연구를 통하여 꾸준히 보고 되었다(65, 72). Model 1의 성능을 비교해보면 훈련데이터에서는 Model 1b(단백체모델)의 예측력이 1.0로 가장 높은 정확도를 보였으나, 테스트 데이터에서는 0.67으로 불량한 예측 정확도를 보였다. 다만 임상 변인과 단백체 변인을 모두 포함한 혼합 모형의 정확도는 훈련데이터(AUC 0.97) 및 테스트 데이터(AUC 0.90) 모두 양호한 예측력을 보였다. 따라서 단백체 데이터 단독으로 보다는 임상 변인을 함께 사용하였을 때 주요정신질환의 발병을 더 안정된 정확도로 예측할 수 있음을 확인하였다.

주요정신질환의 증상 호전 여부를 예측하는 임상 및 단백체 예측 마커를 분석한 모델2의 결과는 다음과 같다. 모델 2a 분석을 통해

[18] CO9, CMGA, ITIH4, MIF, SAMP
[19] A1AT, VWF, HGFA, F13A, PPIA, A1BG, MASP

선택된 주요정신질환의 발병과 관련된 초기 평가 임상 변인은, 초기 평가 시 CHR-PL로 분류 될 수록, 연구 참여 기간, BPSS-T3 threshold, HAM-A, GAF-현재, WHOQOL-BREF 정신심리학적요인, CSM, PQ-B Threshold 는 Model 2a 분석에서 선택된 임상 변인이었다. 이중 연구 참여 기간, GAF-현재, WHOQOL-BREF 정신심리학적요인, CSM, PQ-B Threshold는 혼합모형 Model 1c에서도 선택되었다. Model 2b와 Model 2c를 통해 선택된 단백체 마커는 총 9개로 또한, 면역반응 및 염증반응[20], 응고연쇄반응[21]과 관련된 단백질들을 포함하였다(65, 72). 특히 Cortisol binding globulin(CBG)는 우울증, 양극성장애 환자군에서 보고된 바 있는 단백질로 만성 스트레스의 잠재적인 생물학적 지표로(75), 임상 경과를 예측할 수 있는 유용한 마커가 될 가능성이 있겠다. Model 2의 성능을 비교해보면 훈련데이터에서는 Model 2d(혼합모델)의 예측력이 0.99로 가장 높은 정확도를 보였으나, 테스트 데이터에서는 0.71으로 보통 예측 정확도를 보였다. 모델 2의 경우 테스트 데이터에서 0.59-0.71의 불량한 예측력을 보였는데, 이는 적은 샘플수로 인한 과적합의 문제로 보인다. 이러한 문제를 해결하기 위하여 SF ≥ 0.85인 변수들로 선별하여 모델의 성능을 비교하였을 때 테스트 데이터의 예측도가 0.66-0.79로 SF ≥ 0.80 변수를 포함한 모델의 정확도 보다 과적합 오류가 감소하였다.

본 연구의 한계점으로는 1) 주요정신질환으로 발병한 환자 14명 중 12명(85.7%) 제 2형 양극성장애로 발병하였다. 모델 1a를 통해 선택된 임상 변인들도 대부분 조증 및 치료력과 관련된 변인들로, 조현병 및 주요우울증의 발병과 관련된 요인들은 반영하지 못하였을 수 있다.

---

[20] SAMP
[21] FA12, KLKB1, VWF, SPP2, VWF

2)주요정신질환자들의 연구 참여기간은 2-24개월로 다양하였고, 각 환자들의 증상 호전 여부는 개인의 마지막 평가시점을 기준으로 하였다. 따라서 증상의 호전여부를 평가한 시점이 대상자마다 상이하므로, 본 연구에서 판단한 결과(증상 호전, 비호전)이 실제 대상자의 임상 경과를 제대로 반영하지 못하였을 가능성이 있겠다. 3) 훈련데이터와 테스트 데이터의 성능의 차이는 연구 3에서 구축한 모델의 불안정성을 반영한다. 따라서 종단적인 단백체 변화량, 임상양상의 변화량을 고려하여, 모델에 투입되는 변수들의 안정성을 확보하는 과정이 추후 필요할 것으로 보인다.

제 5 장 - 종합요약

연구 1-3을 통해 주요정신질환 고위험군의 진단, 중증도와 관련된

혈액 단백체 마커를, 경과를 반영하는 단백체 마커를 탐색하였고, 연구 1-3에서 확인된 단백체 마커를 그림 4에 요약하였다.

먼저, 데이터 마이닝 및 선행연구를 토대로 선정된 92개의 단백질 중 연구 1-3에서 유의미한 결과를 보인 단백질은 총 54개의 단백질(58.7%)였다. 따라서 고위험군의 대상자의 병태 생리를 반영하는 단백체 마커들을 본 연구를 통하여 압축할 수 있었다. 그 중 2개 이상의 연구에서 유의미한 결과를 보인 단백질은 19개(20.7%)였다. 특히, 고위험군의 상태를 진단하고(연구 1), 증상의 심각도를 반영하며(연구 2), 증상의 경과와 관련된(연구3) 모두에서 유의미한 결과를 단백체 마커는 5가지 마커(ITIH4, SAMP, A1BG, VWF, CHLE)로 모두 선행연구에서 주요정신질환과 관련성이 보고된 바 있는 단백체 마커였다(15, 65, 114, 115). 그림 4에 따르면 연구 1-3을 통해 유의미한 결과를 보인 단백질 들은 대부분 환자군의 선행연구에서 보고된 지질대사 이상, 혈액 응고연쇄반응, 면역반응, 항산화 반응과 관련된 생물학적 경로와 관련 되었다(70-72, 114, 131). 따라서 이러한 결과들은 생물학적 경로 이상이 주요정신질환으로 진단되기 이전 고위험 상태에서부터 시작되었음을 시사하는 결과일 수 있겠다.

LPA분석을 통해 다양한 증상을 유형화 하였을 때, 연구 1의 프로토콜의 분류기준과 달리 LS, DPD, MD로 유형화 되었다. 또한, 연구 2를 통해 증상의 심각도를 반영하는 29개의 단백질들은 질병의 각 단계를 반영하는 단백체 마커로 활용될 수 있을 것이다.

마지막으로 초기에 측정된 단백질 중 주요정신질환의 경과를 예측하는 단백체 마커 22종류를 선별할 수 있었다. 이 마커들 중 특히 CBG는 선행연구에서 만성 스트레스 상태를 반영하고, 약물치료반응과 관련된 것으로 보고된 단백질로, 이는 매우 의미 있는 발견이다(75).

향후 종단적 연구를 통하여 이 단백질의 종단적 변화량과 임상 경과 사이의 상호작용에 대하여 결과를 확인한다면, 임상 경과를 예측하는 의미있는 단백체 마커가 될 가능성이 높아 보인다.

하지만 본 연구에서는 초기 평가 시 정량화된 단백질 분석자료를 사용하였고, 증상의 경과를 평가한 시점이 대상자마다 상이하므로, 본 연구에서 판단한 임상 경과가 임상 경과를 제대로 반영하지 못하였을 가능성이 있다. 또한, 주요정신질환의 경과에는 초기 평가된 요인들 외에도 다양한 중간요인(약물치료 등)이 경과에 영향을 줄 수 있으나, 본 연구에서는 초기 평가된 요인만을 포함하였다. 모델3의 낮은 예측성능은 이러한 한계를 반영하는 결과일 수 있겠다.

향후 선별된 단백체 마커들의 임상적 유용성을 확보하고, 모델 성능의 안정성을 확보하기 위해서는, 종단적으로 확보된 혈액시료의 단백체 분석을 통해 발현량 변화와 임상 경과 사이의 상호작용에 대한 분석이 수반되어야 할 것으로 보인다.

참고 문헌

1. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012;380(9859):2197-223.

2. Go D-S, Kim Y-E, Paik J-W, Roh S, Yoon S-J. A comparison of disease burden and the government budget for mental health in Korea. Journal of Mental Health. 2020:1-8.

3. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2163-96.

4. Woods SW, Carlson JP, McGlashan TH. DSM-5 and the 'Psychosis Risk Syndrome': The DSM-5 proposal is better than DSM-IV. Psychosis. 2010;2(3):187-90.

5. Fusar-Poli P, Schultze-Lutter F, Cappucciati M, Rutigliano G, Bonoldi I, Stahl D, et al. The Dark Side of the Moon: Meta-analytical Impact of Recruitment Strategies on Risk Enrichment in the Clinical High Risk State for Psychosis. Schizophr Bull. 2016;42(3):732-43.

6. Fusar-Poli P, Nelson B, Valmaggia L, Yung AR, McGuire PK. Comorbid depressive and anxiety disorders in 509 individuals with an at-risk mental state: impact on psychopathology and transition to psychosis. Schizophr Bull. 2014;40(1):120-31.

7. Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, et al. Identification of a Common Neurobiological Substrate for Mental Illness. JAMA Psychiatry. 2015;72(4):305-15.

8. Guest PC. Proteomic Methods in Neuropsychiatric Research. 1st ed. 2017.. ed: Cham : Springer International Publishing : Imprint: Springer; 2017.

9. Janiri D, Moser DA, Doucet GE, Luber MJ, Rasgon A, Lee WH, et al. Shared Neural Phenotypes for Mood and Anxiety Disorders: A Meta-analysis of 226 Task-Related Functional Imaging Studies. JAMA Psychiatry. 2020;77(2):172-9.

10. Lisa M. McTeague, Ph.D. ,, Julia Huemer, M.D. ,, David M. Carreon, M.D. ,, Ying Jiang, M.D. ,, Simon B. Eickhoff, Dr.Med. ,, Amit Etkin, M.D., Ph.D. Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. American Journal of Psychiatry. 2017;174(7):676-85.

11. McGorry PD, Purcell R, Hickie IB, Yung AR, Pantelis C, Jackson HJ. Clinical staging: a heuristic model for psychiatry and youth mental health. Medical Journal of Australia. 2007;187(S7):S40-S2.

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

13. Ren J, Zhao G, Sun X, Liu H, Jiang P, Chen J, et al. Identification of plasma biomarkers for distinguishing bipolar depression from major depressive disorder by iTRAQ-coupled LC-MS/MS and bioinformatics analysis. Psychoneuroendocrinology. 2017;86:17-24.

14. Rodrigues-Amorim D, Rivera-Baltanás T, Vallejo-Curto MdC, Rodriguez-Jamardo C, de las Heras E, Barreiro-Villar C, et al. Proteomics in Schizophrenia: A Gateway to Discover Potential Biomarkers of Psychoneuroimmune Pathways. Frontiers in Psychiatry. 2019;10(885).

15. Li Y, Zhou K, Zhang Z, Sun L, Yang J, Zhang M, et al. Label-free quantitative proteomic analysis reveals dysfunction of complement pathway

in peripheral blood of schizophrenia patients: evidence for the immune hypothesis of schizophrenia. Mol Biosyst. 2012;8(10):2664-71.

16. Schmidt A, Cappucciati M, Radua J, Rutigliano G, Rocchetti M, Dell'Osso L, et al. Improving Prognostic Accuracy in Subjects at Clinical High Risk for Psychosis: Systematic Review of Predictive Models and Meta-analytical Sequential Testing Simulation. Schizophr Bull. 2017;43(2):375-88.

17. Lin A, Reniers RLEP, Wood SJ. Clinical staging in severe mental disorder: evidence from neurocognition and neuroimaging. British Journal of Psychiatry. 2013;202(s54):s11-s7.

18. Sartori JM, Reckziegel R, Passos IC, Czepielewski LS, Fijtman A, Sodré LA, et al. Volumetric brain magnetic resonance imaging predicts functioning in bipolar disorder: A machine learning approach. Journal of Psychiatric Research. 2018;103:237-43.

19. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma Proteome Profiling to Assess Human Health and Disease. Cell systems. 2016;2(3):185-95.

20. Geyer PE, Holdt LM, Teupser D, Mann M. Revisiting biomarker discovery by plasma proteomics. Mol Syst Biol. 2017;13(9):942-.

21. Moos T, Høyer PE. Detection of plasma proteins in CNS neurons: conspicuous influence of tissue-processing parameters and the utilization of serum for blocking nonspecific reactions. J Histochem Cytochem. 1996;44(6):591-603.

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

23. Shin D, Rhee SJ, Lee J, Yeo I, Do M, Joo EJ, et al. Quantitative

Proteomic Approach for Discriminating Major Depressive Disorder and Bipolar Disorder by Multiple Reaction Monitoring-Mass Spectrometry. J Proteome Res. 2021;20(6):3188-203.

24. Lee TY, Lee J, Lee HJ, Lee Y, Rhee SJ, Park DY, et al. Study Protocol for a Prospective Longitudinal Cohort Study to Identify Proteomic Predictors of Pluripotent Risk for Mental Illness: The Seoul Pluripotent Risk for Mental Illness Study. Front Psychiatry. 2020;11:340.

25. Hartmann JA, Yuen HP, McGorry PD, Yung AR, Lin A, Wood SJ, et al. Declining transition rates to psychotic disorder in "ultra-high risk" clients: Investigation of a dilution effect. Schizophrenia Research. 2016;170(1):130-6.

26. Yung AR, Yuen HP, Berger G, Francey S, Hung T-C, Nelson B, et al. Declining Transition Rate in Ultra High Risk (Prodromal) Services: Dilution or Reduction of Risk? Schizophr Bull. 2007;33(3):673-81.

27. Hartmann JA, Nelson B, Spooner R, Paul Amminger G, Chanen A, Davey CG, et al. Broad clinical high-risk mental state (CHARMS): Methodology of a cohort study validating criteria for pluripotent risk. Early Interv Psychiatry. 2019;13(3):379-86.

28. Tandy J. Miller, Ph.D. ,, Thomas H. McGlashan, M.D. ,, Joanna Lifshey Rosen, Psy.D. ,, Lubna Somjee, Ph.D. ,, Philip J. Markovich, B.A. ,, Kelly Stein, Ph.D. , and, et al. Prospective Diagnosis of the Initial Prodrome for Schizophrenia Based on the Structured Interview for Prodromal Syndromes: Preliminary Evidence of Interrater Reliability and Predictive Validity. American Journal of Psychiatry. 2002;159(5):863-5.

29. Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Cannon T, Ventura J, et al. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive

validity, interrater reliability, and training to reliability. Schizophr Bull. 2003;29(4):703-15.

30. Correll CU, Olvet DM, Auther AM, Hauser M, Kishimoto T, Carrión RE, et al. The Bipolar Prodrome Symptom Interview and Scale-Prospective (BPSS-P): description and validation in a psychiatric sample and healthy controls. Bipolar Disorders. 2014;16(5):505-22.

31. Fusar-Poli P, De Micheli A, Rocchetti M, Cappucciati M, Ramella-Cravaro V, Rutigliano G, et al. Semistructured Interview for Bipolar At Risk States (SIBARS). Psychiatry Research. 2018;264:302-9.

32. Hartmann JA, Nelson B, Spooner R, Paul Amminger G, Chanen A, Davey CG, et al. Broad clinical high-risk mental state (CHARMS): Methodology of a cohort study validating criteria for pluripotent risk. Early Intervention in Psychiatry. 2019;13(3):379-86.

33. Saraf G, Moazen-Zadeh E, Pinto JV, Ziafat K, Torres IJ, Kesavan M, et al. Early intervention for people at high risk of developing bipolar disorder: a systematic review of clinical trials. Lancet Psychiatry. 2021;8(1):64-75.

34. Howie LD, Pastor PN, Lukacs SL. Use of medication prescribed for emotional or behavioral difficulties among children aged 6-17 years in the United States, 2011-2012. NCHS Data Brief. 2014(148):1-8.

35. Duffy A, Horrocks J, Doucette S, Keown-Stoneman C, Grof P, Andreazza A, et al. Immunological and neurotrophic markers of risk status and illness development in high-risk youth: understanding the neurobiological underpinnings of bipolar disorder. International Journal of Bipolar Disorders. 2014;2(1):4.

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

37. 이중서, 배승오, 안용민, 박두병, 노경선, 신현균, et al. 한국판 Hamilton 우울증 평가 척도의 신뢰도, 타당도 연구. 신경정신의학. 2005;44(4):456-65.

38. Park HW, Lee, E.H., Kim, D.K., Yu, B.H., Lee, D.S., Kim, J.H. The factor structure of the Korean hamilton depression rating scale (K-HDRS): A confirmatory factor analysis. J Korean Neuropsychiatr Assoc. 2009;48(1):21-8.

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

40. 정희연, 조현상, 주연호, 신현균, 이중서, 황석현, et al. Young 조증 평가 척도의 한국판 표준화 연구. 신경정신의학. 2003;42(2):263-9.

41. Mohammadi Z, Pourshahbaz A, Poshtmashhadi M, Dolatshahi B, Barati F, Zarei M. Psychometric Properties of the Young Mania Rating Scale as a Mania Severity Measure in Patients With Bipolar I Disorder. Practice in Clinical Psychology. 2018;6(3):175-82.

42. Thompson E. Hamilton Rating Scale for Anxiety (HAM-A). Occupational Medicine. 2015;65(7):601-.

43. 김창윤. 정신장애 평가도구: 하나의학사; 2001. 278 p.

44. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull. 1987;13(2):261-76.

45. 이중서, 안용민, 신현균, 안석균, 주연호, 김승현, et al. 한국판 양성 및 음성증후군 척도(Positive and Negative Syndrome Scale)의 신뢰도와 타당도. 신경정신의학. 2001;40(6):1090-105.

46. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont). 2007;4(7):28-37.

47. Yi JS CM, Lee BW, Lee SI. Reliability and Validity of the

Korean Version of Global Assessment of Functioning Scale. Korean J Schizophr Res. 2003;6(2):118-24.

48. 정명훈. 한국어판 총괄기능평가척도(Global Assessment of Functioning Scale)의 신뢰도와 타당도. 춘천: 翰林大學校 大學院; 2003.

49. First MB. Structured Clinical Interview for DSM-IV Axis I Disorders. Biometrics Research Department. 1997.

50. Hahn OS, Ahn JH, Song SH, Cho M, Kim J-K, Bae JN. Development of Korean version of Structured Clinical Interview Schedule for DSM-IV Axis I Disorder: Interrater reliability. J Korean Neuropsychiatr Assoc. 2000;39:362-72.

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

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

53. Mongan D, Föcking M, Healy C, Susai SR, Heurich M, Wynne K, et al. Development of Proteomic Prediction Models for Transition to Psychotic Disorder in the Clinical High-Risk State and Psychotic Experiences in Adolescence. JAMA Psychiatry. 2021;78(1):77-90.

54. Khoury R, Nasrallah HA. Inflammatory biomarkers in individuals at clinical high risk for psychosis (CHR-P): State or trait? Schizophr Res. 2018;199:31-8.

55. Mantere O, Saarela M, Kieseppä T, Raij T, Mäntylä T, Lindgren M, et al. Anti-neuronal anti-bodies in patients with early psychosis. Schizophr Res. 2018;192:404-7.

56. Kim H, Rhee SJ, Lee H, Han D, Lee TY, Kim M, et al. Identification of altered protein expression in major depressive disorder and bipolar disorder patients using liquid chromatography-tandem mass spectrometry. Psychiatry Res. 2021;299:113850.

57. Lee H, Rhee SJ, Kim J, Lee Y, Kim H, Lee J, et al. Predictive protein markers for depression severity in mood disorders: A preliminary trans-diagnostic approach study. J Psychiatr Res. 2021;142:63-72.

58. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26(7):966-8.

59. McTeague LM, Goodkind MS, Etkin A. Transdiagnostic impairment of cognitive control in mental illness. J Psychiatr Res. 2016;83:37-46.

60. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, et al. Plasma proteomic signature of age in healthy humans. Aging Cell. 2018;17(5):e12799.

61. Miike K, Aoki M, Yamashita R, Takegawa Y, Saya H, Miike T, et al. Proteome profiling reveals gender differences in the composition of human serum. PROTEOMICS. 2010;10(14):2678-91.

62. Huang C-C, Luo Q, Palaniyappan L, Yang AC, Hung C-C, Chou K-H, et al. Transdiagnostic and Illness-Specific Functional Dysconnectivity Across Schizophrenia, Bipolar Disorder, and Major Depressive Disorder. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2020;5(5):542-53.

63. R Development core team. R: A language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.

64. Thomas EA, Dean B, Pavey G, Sutcliffe JG. Increased CNS levels of apolipoprotein D in schizophrenic and bipolar subjects: implications for the pathophysiology of psychiatric disorders. Proc Natl Acad Sci U S A. 2001;98(7):4066-71.

65. Song YR, Wu B, Yang YT, Chen J, Zhang LJ, Zhang ZW, et al. Specific alterations in plasma proteins during depressed, manic, and euthymic states of bipolar disorder. Braz J Med Biol Res. 2015;48(11):973-82.

66. Zhou X, Barrett TB, Kelsoe JR. Promoter variant in the GRK3 gene associated with bipolar disorder alters gene expression. Biol Psychiatry. 2008;64(2):104-10.

67. Sams DS, Nardone S, Getselter D, Raz D, Tal M, Rayi PR, et al. Neuronal CTCF Is Necessary for Basal and Experience-Dependent Gene Regulation, Memory Formation, and Genomic Structure of BDNF and Arc. Cell Rep. 2016;17(9):2418-30.

68. Willis M, Leitner I, Jellinger KA, Marksteiner J. Chromogranin peptides in brain diseases. J Neural Transm (Vienna). 2011;118(5):727-35.

69. Knutsvik G, Collett K, Arnes J, Akslen LA, Stefansson IM. QSOX1 expression is associated with aggressive tumor features and reduced survival in breast carcinomas. Mod Pathol. 2016;29(12):1485-91.

70. Hoirisch-Clapauch S, Nardi AE, Gris JC, Brenner B. Coagulation and mental disorders. Rambam Maimonides Med J. 2014;5(4):e0036.

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

72. Foldager L, Köhler O, Steffensen R, Thiel S, Kristensen AS, Jensenius JC, et al. Bipolar and panic disorders may be associated with hereditary defects in the innate immune system. J Affect Disord. 2014;164:148-54.

73. Fernandes BS, Gama CS, Ceresér KM, Yatham LN, Fries GR, Colpo G, et al. Brain-derived neurotrophic factor as a state-marker of mood episodes in bipolar disorders: a systematic review and meta-regression analysis. J Psychiatr Res. 2011;45(8):995-1004.

74. Calabrò M, Mandelli L, Crisafulli C, Lee SJ, Jun TY, Wang SM, et al. Neuroplasticity, Neurotransmission and Brain-Related Genes in Major Depression and Bipolar Disorder: Focus on Treatment Outcomes in an Asiatic Sample. Adv Ther. 2018;35(10):1656-70.

75. Vieta E, Gastó C, Osaba MM, Otero A, Nieto E, Pintor L, et al. Cortisol-binding globulin levels in bipolar disorder. Eur Psychiatry. 1997;12(1):11-5.

76. Spinardi L, Witke W. Gelsolin and diseases. Subcell Biochem. 2007;45:55-69.

77. Frye MA, Nassan M, Jenkins GD, Kung S, Veldic M, Palmer BA, et al. Feasibility of investigating differential proteomic expression in depression: implications for biomarker development in mood disorders. Transl Psychiatry. 2015;5:e689.

78. Russo AJ. Decreased Serum Hepatocyte Growth Factor (HGF) in Individuals with Bipolar Disorder Normalizes after Zinc and Anti-oxidant Therapy. Nutr Metab Insights. 2010;3:49-55.

79. Lee S-Y, Wang T-Y, Lu R-B, Wang L-J, Li S-C, Tu C-Y, et al. Identification of potential plasma protein biomarkers for bipolar II disorder: a preliminary/exploratory study. Scientific Reports. 2021;11(1):9452.

80. Tramontina JF, Andreazza AC, Kauer-Sant'anna M, Stertz L, Goi J, Chiarani F, et al. Brain-derived neurotrophic factor serum levels before and after treatment for acute mania. Neurosci Lett. 2009;452(2):111-3.

81. Fernandes BS, Gama CS, Maria Ceresér K, Yatham LN, Fries GR, Colpo G, et al. Brain-derived neurotrophic factor as a state-marker of mood episodes in bipolar disorders: A systematic review and meta-regression analysis. Journal of Psychiatric Research. 2011;45(8):995-1004.

82. Passos IC, Mwangi B, Vieta E, Berk M, Kapczinski F. Areas of controversy in neuroprogression in bipolar disorder. Acta Psychiatrica Scandinavica. 2016;134(2):91-103.

83. Berk M, Kapczinski F, Andreazza AC, Dean OM, Giorlando F, Maes M, et al. Pathways underlying neuroprogression in bipolar disorder: Focus on inflammation, oxidative stress and neurotrophic factors. Neuroscience & Biobehavioral Reviews. 2011;35(3):804-17.

84. Tein JY, Coxe S, Cham H. Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. Struct Equ Modeling. 2013;20(4):640-57.

85. Jung MH, Jang JH, Kang DH, Choi JS, Shin NY, Kim HS, et al. The reliability and validity of the korean version of the structured interview for prodromal syndrome. Psychiatry Investig. 2010;7(4):257-63.

86. Loewy RL, Bearden CE, Johnson JK, Raine A, Cannon TD. The prodromal questionnaire (PQ): preliminary validation of a self-report screening measure for prodromal and psychotic syndromes. Schizophr Res. 2005;79(1):117-25.

87. Preti A, Raballo A, Kotzalidis GD, Scanu R, Muratore T, Gabbrielli M, et al. Quick Identification of the Risk of Psychosis: The Italian Version of the Prodromal Questionnaire-Brief. Clin Pract Epidemiol Ment Health.

2018;14:120-31.

88. Jang YE, Lee TY, Hur J-W, Kwon JS. Validation of the Korean Version of the Prodromal Questionnaire-Brief Version in Non-Help-Seeking Individuals. Psychiatry Investig. 2019;16(2):109-14.

89. Van Meter A, Guinart D, Bashir A, Sareen A, Cornblatt BA, Auther A, et al. Bipolar Prodrome Symptom Scale - Abbreviated Screen for Patients: Description and validation. Journal of Affective Disorders. 2019;249:357-65.

90. 홍진표, 전대종, 임아영, 박수빈, 박승진. 한국어판 자기보고형 간이우울증상평가척도(QIDS-SR)의 신뢰도 및 타당도. 우울조울병. 2013;11(1):44-50.

91. Spielberger CD. State-Trait anxiety inventory. The Corsini encyclopedia of psychology. 2010:1-.

92. Lee Y-J BS-Y, Choi S, Lee H-K, Kim B-W, Kim W, et al. . Korean State-Trait Anxiety Inventory Application Study in Middle and High School Students. J Korean Neuropsychiatr Assoc. 2008;47:471-80.

93. Kim D, Park SC, Yang H, Oh DH. Reliability and validity of the korean version of the childhood trauma questionnaire-short form for psychiatric outpatients. Psychiatry Investig. 2011;8(4):305-11.

94. Bernstein DP, Fink L, Handelsman L, Foote J, Lovejoy M, Wenzel K, et al. Initial reliability and validity of a new retrospective measure of child abuse and neglect. The American journal of psychiatry. 1994;151(8):1132-6.

95. Lee S-R, Lee W-H, Park J-S, Kim S-M, Kim J-W, Shim J-H. The Study on Reliability and Validity of Korean Version of the Barratt Impulsiveness Scale-11-Revised in Nonclinical Adult Subjects. J Korean Neuropsychiatr Assoc. 2012;51(6):378-86.

96. Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. J Clin Psychol. 1995;51(6):768-74.

97. Cho C-H, Jung S-Y, Kapczinski F, Rosa AR, Lee H-J. Validation of the Korean Version of the Biological Rhythms Interview of Assessment in Neuropsychiatry. Psychiatry Investig. 2018;15(12):1115-20.

98. Chul-Hyun C, Seo-Yeon J, Flávio K, Adriane RR, Heon-Jeong L. Validation of the Korean Version of the Biological Rhythms Interview of Assessment in Neuropsychiatry. PSYCHIATRY INVESTIGATION. 2018;15(12):1115-20.

99. 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.

100. 이형영, 윤진상, 국승희. 아침형-저녁형의 측정을 위한 한국어판 조합척도의 교차 타당화. 신경정신의학. 1999;38(2):297-305.

101. Baek JH, Kim JS, Huh I, Lee K, Park JH, Park T, et al. Prevalence, behavioral manifestations and associated individual and climatic factors of seasonality in the Korean general population. Comprehensive Psychiatry. 2015;57:148-54.

102. 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. Archives of General Psychiatry. 1984;41(1):72-80.

103. Kasper S, Wehr TA, Bartko JJ, Gaist PA, Rosenthal NE. Epidemiological Findings of Seasonal Changes in Mood and Behavior: A Telephone Survey of Montgomery County, Maryland. Archives of General Psychiatry. 1989;46(9):823-33.

104. Min SK, Kim KI, Lee CI, Jung YC, Suh SY, Kim DK. Development of

the Korean versions of WHO Quality of Life scale and WHOQOL-BREF. Quality of Life Research. 2002;11(6):593-600.

105. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med. 1998;28(3):551-8.

106. Baek H-S, Lee K-U, Joo E-J, Lee M-Y, Choi K-S. Reliability and Validity of the Korean Version of the Connor-Davidson Resilience Scale. Psychiatry Investig. 2010;7(2):109-15.

107. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety. 2003;18(2):76-82.

108. Cornblatt BA, Auther AM, Niendam T, Smith CW, Zinberg J, Bearden CE, et al. Preliminary Findings for Two New Measures of Social and Role Functioning in the Prodromal Phase of Schizophrenia. Schizophr Bull. 2007;33(3):688-702.

109. Nylund-Gibson K, Choi AY. Ten frequently asked questions about latent class analysis. Translational Issues in Psychological Science. 2018;4(4):440-61.

110. Scott J, Leboyer M, Hickie I, Berk M, Kapczinski F, Frank E, et al. Clinical staging in psychiatry: a cross-cutting model of diagnosis with heuristic and practical value. Br J Psychiatry. 2013;202(4):243-5.

111. McGorry PD, Nelson B, Goldstone S, Yung AR. Clinical staging: a heuristic and practical strategy for new research and better health and social outcomes for psychotic and related mood disorders. Can J Psychiatry. 2010;55(8):486-97.

112. Nieman DH, McGorry PD. Detection and treatment of at-risk mental state for developing a first psychosis: making up the balance. The Lancet

Psychiatry. 2015;2(9):825-34.

113. McGorry P, Keshavan M, Goldstone S, Amminger P, Allott K, Berk M, et al. Biomarkers and clinical staging in psychiatry. World psychiatry : official journal of the World Psychiatric Association (WPA). 2014;13(3):211-23.

114. Xu HB, Zhang RF, Luo D, Zhou Y, Wang Y, Fang L, et al. Comparative proteomic analysis of plasma from major depressive patients: identification of proteins associated with lipid metabolism and immunoregulation. Int J Neuropsychopharmacol. 2012;15(10):1413-25.

115. Domenici E, Wille DR, Tozzi F, Prokopenko I, Miller S, McKeown A, et al. Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections. PLoS One. 2010;5(2):e9166.

116. Jaros JAJ, Martins-de-Souza D, Rahmoune H, Rothermundt M, Leweke FM, Guest PC, et al. Protein phosphorylation patterns in serum from schizophrenia patients and healthy controls. Journal of Proteomics. 2012;76:43-55.

117. Lee MY, Kim EY, Kim SH, Cho K-C, Ha K, Kim KP, et al. Discovery of serum protein biomarkers in drug-free patients with major depressive disorder. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2016;69:60-8.

118. Schwarz E, Guest PC, Rahmoune H, Martins-de-Souza D, Niebuhr DW, Weber NS, et al. Identification of a blood-based biological signature in subjects with psychiatric disorders prior to clinical manifestation. World J Biol Psychiatry. 2012;13(8):627-32.

119. Raivisto T, Heikkinen AM, Silbereisen A, Kovanen L, Ruokonen H, Tervahartiala T, et al. Regulation of Salivary Peptidoglycan Recognition Protein 1 in Adolescents. JDR Clin Trans Res. 2020;5(4):332-41.

120. Arentsen T, Qian Y, Gkotzis S, Femenia T, Wang T, Udekwu K, et al. The bacterial peptidoglycan-sensing molecule Pglyrp2 modulates brain development and behavior. Mol Psychiatry. 2017;22(2):257-66.

121. Han SYS, Cooper JD, Ozcan S, Rustogi N, Penninx BWJH, Bahn S. Integrating proteomic, sociodemographic and clinical data to predict future depression diagnosis in subthreshold symptomatic individuals. Translational Psychiatry. 2019;9(1):277.

122. Cooper JD, Han SYS, Tomasik J, Ozcan S, Rustogi N, van Beveren NJM, et al. Multimodel inference for biomarker development: an application to schizophrenia. Translational Psychiatry. 2019;9(1):83.

123. Burnham KP, Anderson DR, Huyvaert KP. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behavioral Ecology and Sociobiology. 2011;65(1):23-35.

124. Burnham KP, Anderson DR. Multimodel Inference:Understanding AIC and BIC in Model Selection. Sociological Methods & Research. 2004;33(2):261-304.

125. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Parzen E, Tanabe K, Kitagawa G, editors. Selected Papers of Hirotugu Akaike. New York, NY: Springer New York; 1998. p. 199-213.

126. HURVICH CM, TSAI C-L. Regression and time series model selection in small samples. Biometrika. 1989;76(2):297-307.

127. Sugiura N. Further analysts of the data by akaike' s information criterion and the finite corrections. Communications in Statistics - Theory and Methods. 1978;7(1):13-26.

128. Burnham KPA, D. R. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. second edn ed: Springer; 2002.

129. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005;21(20):3940-1.

130. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.

131. Perez M, Farrant J. Immune reactions and mental disorders. Psychological Medicine. 1988;18(1):11-3.

Abstract

A Study on Clinical variables and Blood Proteins related to Diagnosis, reflect Severity, and Progress of High-Risk State of Psychiatric Disorders

Hyunju Lee
Department of Medicine (Clinical Medical Science)
The Graduate School
Seoul National University

Introduction: Early diagnosis and treatment of major mental disorders are very important in improving the prognosis and quality of life of young age. The existing diagnostic system has limitations in diagnosing high-risk patients with non-specific symptoms, and it is necessary to find biological markers that can reflect the disease status of high-risk groups. Therefore, in this study, blood proteins were quantitatively analyzed for high-risk groups of major mental illness, and protein markers reflecting the state of high-risk groups and disease progress were explored in various ways.

Method: 90 clinical high-risk subjects recruited from 7 medical

institutions in Seoul Clinical evaluation and blood collection were performed on the same day, and blood proteins were quantified using 158 target peptides with multiple reaction monitoring-mass spectrometry. The analyzed protein expression in the high-risk group was compared with that in the patient group and the normal control group. Protein markers reflecting the high-risk group status were explored(Study1), In addition, through latent profile analysis, heterogeneous characteristics of high-risk were categorized, and correlation analysis was performed to search for proteins reflecting symptom severity(Study2). 10-crossfold LASSO regression was repeated 100 times for each model, features predict transition to major psychiatric disorders, and recovery of symptoms were selected. Predictive performance of the models were measured by area under the receiver operating characteristics (Study3).

Results: Through study 1, Twenty-seven protein markers reflecting high-risk conditions were identified. Most of proteins were reported in previous studies related to psychiatric disorders. In study 2, clinical high-risk groups were categorized into three subgroups, and it was different from categorized group in study 1. Twenty-nine protein markers with significant correlation with two or more of the 19 clinical indicators could be selected. In study 3, clinical variables (9 markers in clinical models, 5 markers in mixed models) and protein (15 markers in proteomic models, 2 markers in mixed models) were selected to predict the transition to psychiatric disorders. The performance of the mixed model (Model 1c) was showed excellent performance in the test data (AUC 0.90-0.97). On the other hand, clinical variables (8 markers in clinical models, 9 markers mixed models) and protein variables (8 markers in clinical models, 5

markers in mixed models) were selected to predict symptom recovery of symptom. However, the performance of the prediction model was fair at AUC 0.70−0.75 in the test data. Protein markers discovered through study 1−3 were integrated and most of the significant proteins were related to the blood coagulation cascade, immune response, lipid metabolism, and antioxidant response pathways reported in previous study.

Conclusion: Through this study, it was confirmed that meaningful biological alterations reported in the patient group were also observed in the patients with high−risk state. In addition, it was possible to identify common protein markers that could reflect the severity of symptoms, and predict the progression of disease. However, in order to validate our results and confirm the stability of the discovered protein markers, it will be necessary to evaluate the longitudinal variation of the protein markers in the future. Nevertheless, the protein and clinical indicators discovered in this study are expected to be used as useful markers for early diagnosis of the conditions of high−risk states in the future and predicting the course of the disease.

Keywords : Major psychiatric disorder, High−risk state, Transition, Progress, Protein, Predictor

Student Number: 2020−32637

Table 1-1. Inclusion and exclusion criteria of clinical high risk group[a]

Subgroup	Inclusion criteria	Exclusion criteria
CHR-PRS	• **Attenuated psychosis group.** Meeting all of the following criteria: (1) At least one SOPS score rated 3–5 points from PSP1 to PSP5 (2) Symptoms occurred during the past years, and the current score is higher than 12 months ago (3) Symptoms appear at least once a week in the past month. • **Brief limited intermittent psychotic symptoms group.** Meeting all of the following criteria: (1) At least one of SOPS score of 6 points was obtained from PSP1 to PSP5 (2) Symptoms have started within the last 3 months (3) Symptoms now appear more than once a month, at least several minutes per day.	• antipsychotics are prescribed for more than 30 days
CHR-BD	Bipolar at risk state criteria (BARS criteria) Common inclusion criteria: Meeting criteria for at least one group from the below in the last 12 months. • **Subtype 1.** Subthreshold mania group : manic episodes lasting more than two days, less than four days. • **Subtype 2.** Depression + cyclothymic features group: Meeting all of the following criteria: (1) at least a week, there is depressed mood or loss of interest or pleasure (2) two or more items from diagnostic criteria A of major depressive disorder in DSM (3) multiple episodes with subthreshold manic symptoms not meeting group I criteria and numerous episodes with depressive symptoms. • **Subtype 3.** Depression + genetic risk group: The criteria for depression listed above AND the presence of first-degree relatives with bipolar disorder. • **Subtype 4.** Cyclothymic features and genetic risk group: Meeting all of the following criteria: (1) Numerous episodes with subthreshold manic symptoms, not meeting group I criteria and numerous episodes with depressive symptoms (2) the presence of first-degree relatives with bipolar disorder. • **Subtype 5.** Subthreshold mixed episode group: Meeting all of the following criteria: (1) Subthreshold mania (2) depressed mood nearly every day but less than five consecutive days. • **Subtype 6.** Mood swings group: There is a recent onset of mood instability.	• history of psychosis that lasted more than seven days in SIPS evaluation regardless of treatment • history of mood stabilizer treatment for six weeks or more • history of antipsychotic treatment for more than three weeks • history of bipolar I and II disorder.
CHR-MDD	• Meeting all of the following criteria in the last 12 months: (1) at least 1 week of a depressed mood or loss of interest or pleasure (2) two or more items on diagnostic criteria A for major depressive disorder from the DSM.	• major depressive episodes lasting more than two weeks.

[a] The recruitment and exclusion criteria for high-risk groups include the contents of the previously published study protocol (Lee TY, Lee J, Lee HJ, Lee Y, Rhee SJ, Park DY, et al. Study Protocol for a Prospective Longitudinal Cohort Study to Identify Proteomic Predictors of Pluripotent Risk for Mental Illness: The Seoul Pluripotent Risk for Mental Illness Study. Front Psychiatry. 2020;11:340)

CHR-PL	Subjects who complained of psychotic or mood symptoms not meeting the criteria above but meeting all of the following criteria: (1) symptoms experienced by the subject are causing distress and help-seeking (2) cares where the clinician judges that the follow-up observation is necessary due to the possibility of future mental illness. Extended undifferentiated risk group: (1) Those who have been diagnosed with anxiety disorders during the last 5 years (2) those with a family history of psychosis, bipolar disorders or anxiety disorders.	• history of full-threshold disorder (psychosis, bipolar, severe major depressive disorder, borderline disorder) that lasted more than seven days in SIPS evaluation regardless of treatment
CHR-Extended	Participants who have been diagnosed with the following mental illness before the age of 25 and has not yet passed 5 years; bipolar II disorder, cyclothymia, bipolar disorder NOS, recurrent MDD (recurrent MDD, regardless of the duration after diagnosis in case of recurrent MDD).	

Table 1-2. Demographic and clinical characteristics of the study subjects related to HC, CHR-BD, and BD

Characteristics	HC (n=75)	CHR-BD (n=40)	BD[a] (n=55)	*p*-value[b]
Age, years (mean ± SD)	23.9 ± 4.1	22.1 ± 4.2	23.0 ± 4.5	0.940
Sex (Female), n (%)	38 (50.7)	27 (67.5)	34 (61.8)	0.180
Education, years (mean ± SD)	**15.0 ± 2.1**	**13.7 ± 2.3**	**13.9 ± 2.2**	**0.001[c]**
Age of symptom onset, years (mean ± SD)		16.4 ± 4.1	**18.3 ± 4.9**	**0.050**
Duration of untreated period, month (mean ± SD)		55.4 ± 54.6	43.6 ± 39.0	0.466
Psychopharmacology use, n (%)		25 (62.5)	46 (83.6)	**0.030**
Antipsychotics, n (%)		**18 (45.0)**	**41 (74.5)**	**0.003**
Mood stabilizers, n (%)		**15 (37.5)**	**43 (78.2)**	**<0.001**
Antidepressants, n (%)		5 (12.5)	7 (12.7)	>0.999
Benzodiazepines, n (%)		11 (27.5)	17 (30.9)	0.822
HAM-D Total, mean ± SD		10.5 ± 6.4	11.0 ± 6.0	0.693
YMRS Total, mean ± SD		**2.0 ± 2.4**	**3.6 ± 4.2**	**0.025**
CGI-S, mean ± SD		3.3 ± 0.7	3.3 ± 0.8	0.763

[a] BD-I (n=7), BD-II (n=48)

[b] Significant differences between the HC, CHR-BD, BD samples were examined using the chi-square test for categorical variables, and analysis of variance (ANOVA) and independent samples t-test for continuous variables. Significant findings at *p*-value<0.050 are in **bold**

[c] Healthy controls showed longer duration of education when compared to CHR-BP and BD group (Tukey HSD)

Table 1-3. Classification of subgroup of CHR-BD according to BARS criteria

		Classified subtype[a]			
	Total (n=39)	Subtype 1 (n=5)	Subtype 2 (n=28)	Subtype 3 (n=6)	*p*-value[b]
Use of psychotropics, n (%)	25 (64.1)	3 (60.0)	18 (64.3)	4 (66.7)	>0.999
Antipsychotics, n (%)	18 (46.2)	2 (40.0)	12 (42.9)	4 (66.7)	0.545
Mood stabilizers, n (%)	15 (38.5)	1 (20.0)	12 (42.9)	2 (33.2)	0.602
Antidepressants, n (%)	5 (12.8)	0 (0.0)	4 (14.3)	1 (16.7)	0.648
Benzodiazepines, n (%)	11 (27.5)	1 (20.0)	8 (28.6)	2 (33.3)	0.887
HAM-D Total, mean ± SD	10.5 ± 6.5	8.8 ± 4.5	11.1 ± 6.8	8.8 ± 6.4	0.632
YMRS Total, mean ± SD	2.0 ± 2.5	4.6 ± 4.7	1.8 ± 1.8	0.5 ± 0.8	0.078
CGI-S, mean ± SD	3.3 ± 0.7	3.4 ± 0.6	3.4 ± 0.7	2.8 ± 0.8	0.241

[a] BARS criteria : (1) subthreshold mania (n=5, 12.5%) , (2) depression + cyclothymic features (n=28, 70.0%), (3) depression + genetic risk (n=6, 15.0%), (4) cyclothymic features and genetic risk (n=0, 0.0%), (5) subthreshold mixed episode (n=0, 0.0%), and (6) mood swings (n=1, 2.5%).

[b] Significant differences between the subtype 1, subtype 2, and subtype 3 were examined using the Fisher's exact test for categorical variables, and Kruskal-Wallis H test for continuous variables.

Table 1-4[a]. Differentially expressed peptides (ANOVA) identified by LC-MRM analysis between HC, CHR-BD, and BD.

Protein ID	Gene name	Peptide	HC vs. CHR-BD vs. BD			CHR-BD vs. BD		CHR-BD vs. HC		BD vs. HC	
			F-statistics	p-value (ANOVA)	q-value	Mean difference	p-value (Tukey's HSD)	Mean difference	p-value (Tukey's HSD)	Mean difference	p-value (Tukey's HSD)
Over-expressed in CHR-BD											
O00391	QSOX1	DTGAALLAESR	9.898	<0.001	**<0.001**	0.568	**<0.001**	0.303	**0.034**	-0.265	**0.043**
P10645	CMGA	ELQDLALQGAK	8.834	<0.001	**<0.001**	0.783	**<0.001**	0.387	0.076	-0.396	**0.038**
P67936	TPM4	TIDDLEEK	6.459	0.002	**0.011**	0.919	**0.001**	0.484	0.117	-0.435	0.108
Under-expressed in CHR-BD											
P00742	FA10	TGIVSGFGR	13.774	<0.001	**<0.001**	-0.439	**<0.001**	-0.236	**0.009**	0.202	0.014
P61626	LYSC	WESGYNTR	13.549	<0.001	**<0.001**	-0.381	**<0.001**	-0.167	**0.048**	0.214	0.003
P06396	GELS	EGGQTAPASTR	16.341	<0.001	**<0.001**	-0.411	**<0.001**	-0.385	**<0.001**	0.026	0.921
P02647	APOA1	LLDNWDSVTSTFSK	6.186	0.003	**0.016**	-0.234	**0.003**	-0.192	**0.011**	0.042	0.761
P02652	APOA2	EQLTPLIK	12.961	<0.001	**<0.001**	-0.397	**<0.001**	-0.252	**0.002**	0.145	0.080
P23560	BDNF	QYFYETK	8.208	<0.001	**<0.001**	-0.005	**<0.001**	-0.003	**0.016**	0.002	0.230
Q14624	ITIH4	FAHTVVTSR	11.662	<0.001	**<0.001**	-0.337	**<0.001**	-0.207	**0.006**	0.131	0.077
P51884	LUM	ILGPLSYSK	14.454	<0.001	**<0.001**	-0.365	**<0.001**	-0.292	**<0.001**	0.074	0.446
P04275	VWF	VTVFPIGIGDR	7.594	0.001	**0.007**	-0.652	**<0.001**	-0.436	**0.018**	0.216	0.293
P01042	KNG1	QVVAGLNFR	12.402	<0.001	**<0.001**	-0.349	**<0.001**	-0.238	**0.001**	0.111	0.162
Q04756	HGFA	VANYVDWINDR	7.823	0.001	**0.007**	-0.365	**<0.001**	-0.229	**0.025**	0.136	0.201
P03952	KLKB1	DSVTGTLPK	6.517	0.002	**0.011**	-0.332	**0.001**	-0.224	**0.030**	0.108	0.363
P03952	KLKB1	TGAVSGHSLK	11.084	<0.001	**<0.001**	-0.345	**<0.001**	-0.228	**0.004**	0.118	0.152
P43251	BTD	LSSGLVTAALYGR	13.258	<0.001	**<0.001**	-0.371	**<0.001**	-0.274	**<0.001**	0.097	0.274
P04217	A1BG	SGLSTGWTQLSK	14.974	<0.001	**<0.001**	-0.375	**<0.001**	-0.275	**<0.001**	0.100	0.219
P04217	A1BG	ATWSGAVLAGR	14.920	<0.001	**<0.001**	-0.398	**<0.001**	-0.279	**<0.001**	0.120	0.143
P61769	B2MG	VNHVTLSQPK	6.605	0.002	**0.011**	-0.333	**0.002**	-0.251	**0.014**	0.082	0.569
O00533	NCHL1	VIAVNEVGR	14.163	<0.001	**<0.001**	-0.362	**<0.001**	-0.287	**<0.001**	0.076	0.425
O00187	MASP2	WPEPVFGR	8.857	<0.001	**<0.001**	-0.345	**<0.001**	-0.201	**0.027**	0.144	0.103
P02790	HEMO	QGHNSVFLIK	16.828	<0.001	**<0.001**	-0.357	**<0.001**	-0.131	0.080	0.226	**<0.001**
P06276	CHLE	FWTSFFPK	7.111	0.001	**0.007**	-0.281	**0.001**	-0.105	0.323	0.176	**0.023**
P32119	PRDX2	QITVNDLPVGR	10.091	<0.001	**<0.001**	-0.421	**0.046**	0.252	0.283	0.672	**<0.001**
P07225	PROS	QLAVLDK	7.324	0.001	**0.007**	-0.269	**0.001**	-0.133	0.115	0.136	**0.066**

q-value means adjusted *p*-value for multiple comparison by using Benjamini-Hochberg FDR. Proteins significant findings at *q*-value<0.050 are in **bold**

[a] Table 1-4 Summarize statistically significant peptides among 143 peptides when analyzed by ANOVA. Multiple comparison was performed by Tukey's HSD.

P02743	SAMP	QGYFVEAQPK	5.129	0.007	**0.035**	-0.441	**0.005**	-0.225	0.198	0.216	0.164
P11226	MBL2	EEAFLGITDEK	5.992	0.003	**0.016**	-0.515	**0.002**	-0.305	0.079	0.210	0.226
P08185	CBG	AQLLQGLGFNLTER	4.754	0.010	**0.048**	-0.509	**0.008**	-0.221	0.345	0.288	0.115
P08185	CBG	ITQDAQLK	5.322	0.006	**0.031**	-0.271	**0.004**	-0.164	0.094	0.107	0.290

Table 1-5[a]. Differentially expressed peptides (ANCOVA[b]) identified by LC-MRM analysis between HC, CHR-BD, and BD.

Protein ID	Gene name	Peptide	HC vs CHR-BD vs BD		CHR-BD vs BD		CHR-BD vs HC		BD vs HC	
			F-statistics	p-value (ANCOVA)	Mean difference[c]	p-value (Tukey's HSD)	Mean difference	p-value (Tukey's HSD)	Mean difference	p-value (Tukey's HSD)
Over-expressed in CHR-BD										
O00391	QSOX1	DTGAALLAESR	6.541	**0.002**	0.485	**0.003**	0.346	**0.042**	-0.139	0.673
P10645	CMGA	ELQDLALQGAK	7.353	**<0.001**	0.740	**0.002**	0.550	**0.023**	-0.190	0.704
P67936	TPM4	TIDDLEEK	4.581	**0.012**	0.880	**0.009**	0.197	0.765	-0.683	0.093
Under-expressed in CHR-BD										
P00742	FA10	TGIVSGFGR	10.153	**<0.001**	-0.362	**<0.001**	-0.315	**0.002**	0.047	0.894
P06396	GELS	EGGQTAPASTR	12.980	**<0.001**	-0.413	**<0.001**	-0.309	**0.001**	0.104	0.555
P02647	APOA1	LLDNWDSVTSTFSK	5.007	**0.008**	-0.176	**0.063**	-0.219	**0.013**	-0.042	0.880
P02652	APOA2	EQLTPLIK	9.735	**<0.001**	-0.311	**0.002**	-0.323	**<0.001**	-0.012	0.992
P23560	BDNF	QYFYETK	5.898	**0.003**	-0.005	**0.006**	-0.004	**0.038**	<0.001	0.828
Q14624	ITIH4	FAHTVVTSR	10.047	**<0.001**	-0.312	**<0.001**	-0.251	**0.004**	0.061	0.766
P51884	LUM	ILGPLSYSK	12.749	**<0.001**	-0.338	**<0.001**	-0.313	**<0.001**	0.025	0.958
P04275	VWF	VTVFPIGIGDR	6.362	**0.002**	-0.619	**0.005**	-0.490	**0.029**	0.129	0.825
P01042	KNG1	QVVAGLNFR	9.129	**<0.001**	-0.300	**<0.001**	-0.228	**0.009**	0.072	0.685
Q04756	HGFA	VANYVDWINDR	7.218	**<0.001**	-0.350	**0.003**	-0.291	**0.014**	0.059	0.871
P03952	KLKB1	DSVTGTLPK	6.814	**0.001**	-0.265	**0.030**	-0.346	**0.003**	-0.080	0.773
P03952	KLKB1	TGAVSGHSLK	10.152	**<0.001**	0.302	**<0.001**	-0.301	**<0.001**	0.000	>0.999
P43251	BTD	LSSGLVTAALYGR	9.784	**<0.001**	-0.313	**<0.001**	-0.286	**0.002**	0.028	0.954
P04217	A1BG	SGLSTGWTQLSK	15.383	**<0.001**	-0.362	**<0.001**	-0.336	**<0.001**	0.026	0.951
P04217	A1BG	ATWSGAVLAGR	13.086	**<0.001**	-0.367	**<0.001**	-0.315	**<0.001**	0.052	0.840
P61769	B2MG	VNHVTLSQPK	6.964	**0.001**	-0.253	**0.048**	-0.369	**0.002**	-0.116	0.600
O00533	NCHL1	VIAVNEVGR	12.776	**<0.001**	-0.354	**<0.001**	-0.287	**<0.001**	0.067	0.732
O00187	MASP2	WPEPVFGR	7.371	**<0.001**	-0.302	**0.002**	-0.247	**0.014**	0.055	0.846
P02790	HEMO	QGHNSVFLIK	11.968	**<0.001**	-0.308	**<0.001**	-0.251	**0.001**	0.057	0.754
P06276	CHLE	FWTSFFPK	4.793	0.010	-0.214	**0.033**	-0.213	**0.030**	0.000	>0.999
P07225	PROS	QLAVLDK	6.402	0.002	-0.187	**0.040**	-0.248	**0.003**	-0.061	0.762
P61626	LYSC	WESGYNTR	9.411	**<0.001**	-0.359	**<0.001**	-0.190	0.061	0.169	0.182
P11226	MBL2	EEAFLGITDEK	5.172	0.007	-0.534	**0.006**	-0.289	0.198	0.245	0.409

Significant findings at *p*-value<0.050 are in **bold**

[a] Table 1-5 Summarize statistically significant peptides among 143 peptides when analyzed by ANCOVA. Multiple comparison was performed by Tukey's HSD.

[b] Covariates : Age, sex, use of antipsychotics, mood stabilizers, antidepressants, and benzodiazepines

[c] Adjusted mean difference

P08185	CBG	AQLLQGLGFNLTER	3.160	0.045	-0.456	**0.048**	-0.297	0.257	0.159	0.744
P08185	CBG	ITQDAQLK	4.616	0.011	-0.212	0.064	-0.249	**0.021**	-0.036	0.938
P02743	SAMP	QGYFVEAQPK	6.151	0.003	-0.346	0.062	-0.494	**0.003**	-0.147	0.669
P32119	PRDX2	QITVNDLPVGR	1.675	0.191	-0.159	0.690	0.236	0.430	0.394	0.170

Table 1-6. Demographic and clinical characteristics of the HC, CHR-PRS and FEP

Characteristics	HC (n=45)	CHR-PRS (n=29)	FEP (n=41)	*p*-value[d]
Age, years (mean ± SD)	22.7 ± 3.6	20.7 ± 3.6	22.5 ± 5.0	0.111
Sex (Female), n (%)	24 (53.3)	11 (37.9)	19 (46.3)	0.430
Education, years (mean ± SD)	**14.9 ± 2.2**	**12.8 ± 1.2**	**13.3 ± 2.2**	**<0.001[e]**
Employment				0.064
Unemployed	3 (6.8)	8 (27.6)	10 (24.4)	
Employed	8 (18.2)	1 (3.4)	6 (14.6)	
Student or military	33 (75.0)	20 (69.0)	25 (61.0)	
BMI, kg/m2 (mean ± SD)		22.0 ±3.1	23.1 ±3.8	0.185
Family history of psychiatric disorders, n (%)		15 (51.7)	13 (31.7)	0.092
Psychopharmacology use, n (%)		**13 (44.8)**	**31 (75.6)**	**0.009**
Antipsychotics, n (%)		**8 (27.6)**	**30 (73.2)**	**<0.001**
Mood stabilizers, n (%)		**4 (13.8)**	**0 (0.0)**	**0.014**
Antidepressants, n (%)		5 (17.2)	2 (4.9)	0.089
Benzodiazepines, n (%)		9 (31.0)	13 (31.7)	0.952
PANSS Total, mean ± SD		**57.1 ±16.3**	**69.7 ±14.1**	**0.001**
Positive total, mean ± SD		**11.5 ±2.7**	**18.0 ±4.9**	**<0.001**
Negative total, mean ± SD		14.4 ±7.1	16.9 ±5.1	0.103
General total, mean ± SD		31.2 ±8.3	35.0 ±7.7	0.054
HAM-D Total, mean ± SD		11.6 ±5.3	11.0 ±4.8	0.651
HAM-A Total, mean ± SD		10.1 ±5.2	10.8 ±5.5	0.621
YMRS Total, mean ± SD		2.6 ±3.1	1.6 ±2.9	0.197
GAF-present, mean ± SD		**53.9 ±8.4**	**45.7 ±10.1**	**0.001**

Significant findings at *p*-value<0.050 are in **bold**

[d] Significant differences between the HC, CHR-PRS and FEP were examined using the chi-square test for categorical variables, and ANOVA and independent samples t-test for continuous variables.

[e] HC showed longer duration of education when compared to CHR-PRS and FEP group (Tukey's HSD)

Table 1-7[a]. Differentially expressed peptides (ANOVA) identified by LC-MRM analysis between HC, CHR-PRS, and FEP.

			HC vs CHR-PRS vs FEP			CHR-PRS vs FEP		CHR-PRS vs HC		FEP vs HC	
Protein ID	**Gene name**	**Peptide**	F-statistics	p-value (ANOVA)	q-value	Mean difference	p-value (Tukey's HSD)	Mean difference	p-value (Tukey's HSD)	Mean difference	p-value (Tukey's HSD)
Q9UGM5	FETUB	AIFYMNNPSR	9.457	<0.001	**<0.001**	**-0.478**	**<0.001**	-0.262	0.044	0.216	0.075
Q14624	ITIH4	FAHTVVTSR	8.610	<0.001	**<0.001**	**-0.331**	**<0.001**	**-0.290**	**0.002**	0.041	0.849
P01042	KNG1	QVVAGLNFR	9.544	<0.001	**<0.001**	**-0.340**	**<0.001**	**-0.290**	**0.001**	0.050	0.773
P04217	A1BG	SGLSTGWTQLSK	8.196	<0.001	**<0.001**	**-0.255**	**0.007**	**-0.321**	**<0.001**	-0.066	0.646
P04217	A1BG	ATWSGAVLAGR	13.983	<0.001	**<0.001**	**-0.427**	**<0.001**	**-0.363**	**<0.001**	0.064	0.676
P02790	HEMO	QGHNSVFLIK	7.042	0.001	**0.013**	**-0.288**	**0.001**	-0.177	0.054	0.111	0.238
P12259	FA5	SQHLDNFSNQIGK	7.267	0.001	**0.013**	**-0.332**	**0.003**	**-0.324**	**0.003**	0.008	0.995
P12259	FA5	NFFNPPIISR	7.702	0.001	**0.013**	**-0.456**	**0.001**	**-0.335**	**0.013**	0.121	0.485
P03952	KLKB1	TGAVSGHSLK	7.978	0.001	**0.013**	**-0.247**	**0.009**	**-0.316**	**<0.001**	-0.068	0.622
O00533	NCHL1	VIAVNEVGR	7.645	0.001	**0.013**	**-0.341**	**<0.001**	**-0.206**	**0.047**	0.136	0.193
P32119	PRDX2	QITVNDLPVGR	7.303	0.001	**0.013**	**-0.640**	**0.005**	-0.050	0.964	0.590	0.003
P06396	GELS	EGGQTAPASTR	6.880	0.002	**0.022**	-0.178	0.213	**-0.378**	**0.001**	-0.200	0.088
P43251	BTD	LSSGLVTAALYGR	6.568	0.002	**0.022**	**-0.255**	**0.007**	**-0.274**	**0.003**	-0.019	0.965
P00742	FA10	TGIVSGFGR	6.053	0.003	**0.031**	**-0.333**	**0.005**	**-0.308**	**0.009**	0.024	0.963
P11226	MBL2	EEAFLGITDEK	5.864	0.004	**0.038**	**-0.593**	**0.004**	**-0.474**	**0.022**	0.119	0.739
P51884	LUM	ILGPLSYSK	5.648	0.005	**0.045**	**-0.280**	**0.003**	-0.190	0.059	0.090	0.454

q-value means adjusted p-value for multiple comparison by using Benjamini-Hochberg FDR. Proteins significant findings at *q*-value<0.050 are in bold

[a] Table 1-7 Summarize statistically significant peptides among 143 peptides when analyzed by ANOVA. Multiple comparison was performed by Tukey's HSD.

Table 1-8. Differentially expressed peptides (ANCOVA) identified by LC-MRM analysis between HC, CHR-PRS, and FEP

			HC vs CHR-PRS vs FEP		CHR-PRS vs FEP		CHR-PRS vs HC		FEP vs HC	
Protein ID	Gene name	Peptide	F-statistics	p-value (ANCOVA[a])	Mean difference[b]	p-value (Tukey's HSD)	Mean difference	p-value (Tukey's HSD)	Mean difference	p-value (Tukey's HSD)
CHR-P < HC, FEP										
P12259	FA5	SQHLDNFSNQIGK	5.561	0.005	**-0.312**	**0.031**	**-0.321**	**0.008**	-0.009	0.997
P12259	FA5	NFFNPPIISR	4.724	0.011	**-0.394**	**0.024**	**-0.325**	**0.034**	0.069	0.888
P01042	KNG1	QVVAGLNFR	8.375	<0.001	**-0.338**	**0.004**	**-0.318**	**0.001**	0.019	0.980
P02790	HEMO	QGHNSVFLIK	9.212	<0.001	-0.375	**<0.001**	**-0.227**	**0.013**	0.147	0.241
Q14624	ITIH4	FAHTVVTSR	9.429	<0.001	-0.399	**<0.001**	**-0.029**	**0.003**	0.108	0.533
P00742	FA10	TGIVSGFGR	8.629	<0.001	**-0.429**	**0.003**	**-0.396**	**0.001**	0.033	0.962
P43251	BTD	LSSGLVTAALYGR	6.863	0.002	**-0.267**	**0.029**	**-0.314**	**0.002**	-0.047	0.890
P04217	A1BG	SGLSTGWTQLSK	8.804	<0.001	**-0.318**	**0.005**	**-0.329**	**<0.001**	-0.010	0.994
P04217	A1BG	ATWSGAVLAGR	12.925	<0.001	**-0.462**	**<0.001**	**-0.354**	**<0.001**	0.107	0.542
CHR-P < FEP										
P32119	PRDX2	QITVNDLPVGR	5.645	0.005	**-0.769**	**0.005**	-0.088	0.909	0.682	0.016
Q9UGM5	FETUB	AIFYMNNPSR	5.686	0.004	**-0.460**	**0.003**	-0.215	0.175	0.245	0.179
O00533	NCHL1	VIAVNEVGR	5.471	0.005	**0.345**	**0.006**	-0.213	0.068	0.132	0.448
CHR-P < HC										
P06396	GELS	EGGQTAPASTR	6.189	0.003	-0.288	0.074	**-0.389**	**0.002**	-0.101	0.721
P51884	LUM	ILGPLSYSK	5.287	0.006	-0.207	0.102	**-0.276**	**0.006**	-0.068	0.775
P03952	KLKB1	TGAVSGHSLK	8.180	<0.001	-0.220	0.085	**-0.358**	**<0.001**	-0.138	0.371
P11226	MBL2	EEAFLGITDEK	4.810	0.010	-0.477	0.084	**-0.568**	**0.011**	-0.091	0.911

Statistically significant differences were analyzed by ANCOVA. Multiple comparison was performed by Tukey's HSD. Significant findings at *p*-value<0.050 are in **bold**

[a] Covariates : Age, sex, use of antipsychotics, mood stabilizers, antidepressants, and benzodiazepines

[b] Adjusted mean difference

Table 1-9. Comparison of baseline demographics and clinical characteristics between CHR-PRS vs. CHR-BD

	CHR-PRS (n = 29)	CHR-BD (n = 40)	Statistics[a]	*p*-value
Age, years, mean ± SD	20.7 ± 3.6	22.1 ± 4.2	465.0	0.162
Gender, male, n (%)	**18 (62.1)**	**13 (32.5)**	**5.9**	**0.015**
Body mass Index, kg/m^2, mean ± SD	22.0 ± 3.1	22.6 ± 3.9	539.5	0.627
Education, years, mean ± SD	12.8 ± 1.2	13.7 ± 2.3	438.5	0.079
Social economic status			0.4	0.601
Unstable, n (%)	10 (34.5)	11 (27.5)		
Stable, n (%)	19 (65.5)	29 (72.5)		
Psychiatric family history, n (%)	14 (48.3)	25 (62.5)	1.4	0.239
History ot treatment			2.2	0.535
No, n (%)	6 (20.7)	5 (12.5)		
Medication-present, n (%)	13(44.8)	25 (62.5)		
Medication-past, n (%)	6 (20.7)	6 (15.0)		
Psychosocial intervention, n (%)	4 (13.8)	4 (10.0)		
Use of psychotropic medication, n (%)	13 (44.8)	25 (62.5)	2.1	0.145
Use of antipsychotics, n (%)	8 (27.6)	18 (45.0)	2.2	0.208
Use of Mood stabilizers, n (%)	4 (13.8)	15 (37.5)	4.7	0.054
Use of Anti-depressant, n (%)	5 (17.2)	5 (12.5)	FE	0.732
Use of Benzodiazepines, n (%)	9 (31.0)	11 (27.5)	0.1	0.793
PANSS : total, mean ± SD	**57.1 ± 16.3**	**47.5 ± 8.2**	**365.0**	**0.009**
PANSS : positive, mean ± SD	**11.5 ± 2.7**	**9.8 ± 2.3**	**334.0**	**0.002**
PANSS : negative, mean ± SD	**14.4 ± 7.1**	**10.7 ± 3.5**	**387.5**	**0.019**
PANSS : general, mean ± SD	**31.2 ± 8.3**	**27.1 ± 5.4**	**408.5**	**0.037**
HAM-D : total, mean ± SD	11.6 ± 5.3	10.5 ± 6.4	501.5	0.342
HAM-D : anhedonia/retardation, mean ± SD	4.2 ± 2.4	3.8 ± 2.1	493.5	0.290
HAM-D : guilt/agitation, mean ± SD	2.9 ± 2.0	2.9 ± 2.3	560.5	0.815
HAM-D : somatic, mean ± SD	1.6 ± 1.6	1.3 ± 1.6	481.0	0.213
HAM-D : insomnia, mean ± SD	1.6 ± 1.5	1.6 ± 1.8	556.5	0.772
HAM-D : appetite, mean ± SD	0.3 ± 0.7	0.6 ± 0.9	507.5	0.292
HAM-A total, mean ± SD	10.1 ± 5.2	9.9 ± 7.3	512.5	0.414
YMRS : total, mean ± SD	2.6 ± 3.1	2.0 ± 2.5	531.0	0.539
CGI-S, mean ± SD	3.6 ± 0.7	3.3 ± 0.7	443.0	0.063
GAF : present, mean ± SD	**53.9 ± 8.4**	**60.3 ± 9.7**	**346.5**	**0.005**
GAF : past one year, mean ± SD	**66.6 ± 13.4**	**75.9 ± 11.4**	**361.0**	**0.008**
GAF : before symptom onset, mean ± SD	81.4 ± 7.6	77.9 ± 13.4	492.0	0.287

Significant findings at p-value < 0.050 are in **bold**

[a] Significant differences between the CHR-PRS and CHR-BD were examined using the chi-square test for categorical variables, and independent samples t-test for continuous variables.

Table 1-10[a]. Differentially expressed peptides identified by LC-MRM analysis between CHR-PRS and CHR-BD

Protein ID	Gene names	Peptide	Mean difference (CHR-PRS vs. CHR-BD)	Student's t	*p*-value	*q*-value
Over-expressed in CHR-PRS (CHR-PRS > CHR-BD)						
P21333	FLNA	AGVAPLQVK	1.00	-3.1	**0.003**	0.395
P10645	CMGA	ELQDLALQGAK	0.55	-2.1	**0.043**	0.528
P14174	MIF	LLCGLLAER	0.38	-2.3	**0.027**	0.528
P12259	FA5	NFFNPPIISR	0.24	-2.5	**0.015**	0.528
P04275	VWF	VIVIPVGIGPHANLK	0.25	-2.0	**0.049**	0.528
P04275	VWF	EAPDLVLQR	0.33	-2.0	**0.049**	0.528
P32119	PRDX2	QITVNDLPVGR	0.41	-2.4	**0.020**	0.528
P37802	TAGL2	QMEQISQFLQAAER	0.62	-2.3	**0.026**	0.528
P37802	TAGL2	TLMNLGGLAVAR	0.63	-2.2	**0.033**	0.528
Q9Y490	TLN1	GLAGAVSELLR	0.60	-2.2	**0.031**	0.395
Q9Y490	TLN1	TLAESALQLLYTAK	0.70	-2.9	**0.005**	0.528
P18206	VINC	ELLPVLISAMK	0.48	-2.2	**0.034**	0.528
Over-expressed in CHR-BD (CHR-PRS < CHR-BD)						
P06396	GELS	TGAQELLR	-0.42	2.0	**0.046**	0.528
P05090	APOD	NILTSNNIDVK	-0.61	2.0	**0.047**	0.528

q-value means adjusted *p*-value for multiple comparison by using Benjamini-Hochberg FDR. Proteins significant findings at *p*-value<0.050 and *q*-value<0.050 are in **bold**

[a] Table 1-10 Summarize statistically significant peptides among 143 peptides when analyzed by Student's independent t-test

Table 2-1. Fit indices of LPA

Model	BIC	saBIC	LMR	BLRT	Proportion of class, n (%)			
			p-value	*p*-value	1	2	3	4
1 class	2634.06	2570.94	N/A	N/A	90 (100%)			
2 class	2476.48	2378.64	0.027	<0.001	66 (73.3%)	24 (26.7%)		
3 class	**2368.18**	**2235.62**	**0.049**	**<0.001**	**28 (31.1%)**	**40 (44.4%)**	**22 (24.4%)**	
4 class	2342.78	2175.51	0.203	<0.001	15 (16.7%)	22 (24.4%)	36 (40.0%)	17 (18.9%)

Table 2-2. Comparison of baseline demographics between LS, DPD, and MD group

	Total (n=90)	Low severity (n = 28)	Depressive-psychosis dominant (n=40)	Manic dominant (n=22)	Statistics[a]	p-value	Post_hoc [b]
Age, years, mean ± SD	21.7 ±3.9	22.5 ±5.0	21.0 ±3.2	21.8 ±3.5	1.4	0.258	
Gender, male, n (%)	40 (44.4)	14 (50.0)	20 (50.0)	6 (27.3)	3.5	0.184	
Body mass Index, kg/m^2, mean ± SD	22.4 ±3.5	23.0 ±3.7	21.8 ±3.5	22.6 ±3.4	1.1	0.347	
Education, years, mean ± SD	13.4 ±1.9	13.8 ±2.5	13.0 ±1.4	13.6 ±1.9	1.3	0.269	
Social economic status					1.0	0.628	
Unstable, n (%)	25 (27.8)	6 (21.4)	13 (32.5)	6 (27.3)			
Stable, n (%)	65 (72.2)	22 (78.6)	27 (67.5)	16 (72.7)			
Initial classification							
CHR-PRS, n (%)	**29 (32.2)**	**11 (39.3)**	**16 (40.0)**	**2 (9.1)**	**7.1**	**0.029**	MD<LS, AP
CHR-BD, n (%)	40 (44.4)	12 (42.9)	14 (35.0)	14 (63.6)	4.8	0.088	
CHR-MDD, n (%)	2 (2.2)	2 (7.1)	0 (0.0)	0 (0.0)	FE	0.152	
CHR-PL, n (%)	4 (4.4)	0 (0.0)	2 (5.0)	2 (9.1)	FE	0.272	
CHR-Extended, n (%)	15 (16.7)	3 (10.7)	8 (20.0)	4 (18.2)	FE	0.631	
Duration of follow up, month, mean ± SD	18.1 ±10.1	17.3 ±12.8	17.3 ±9.5	20.6 ±7.1	0.9	0.417	
Psychiatric family history, n (%)	40 (44.4)	12 (42.9)	19 (47.5)	9 (40.9)	0.3	0.892	
History of treatment					**FE**	**0.032**	
No, n (%)	14 (15.6)	4 (14.3)	6 (15.0)	4 (18.2)			
Medication-present, n (%)	53 (58.9)	12 (42.9)	27 (67.5)	14 (63.6)			
Medication-past, n (%)	**13 (14.4)**	**9 (32.1)**	**1 (2.5)**	**3 (13.6)**			AP < LS
Psychosocial intervention, n (%)	10 (11.1)	3 (10.7)	6 (15.0)	1 (4.5)			
Use of psychotropic medication, n (%)	53 (58.9)	12 (42.9)	27 (67.5)	14 (63.6)	4.4	0.107	
Use of antipsychotics, n (%)	36 (40.0)	8 (28.6)	18 (45.0)	10 (45.5)	2.2	0.332	
Use of Mood stabilizers, n (%)	31 (34.4)	6 (21.4)	15 (37.5)	10 (45.5)	3.4	0.181	
Use of Anti-depressant, n (%)	15 (16.7)	4 (14.3)	7 (17.5)	4 (18.2)	FE	0.938	
Use of Benzodiazepines, n (%)	**29 (32.2)**	**1 (3.6)**	**19 (47.5)**	**9 (40.9)**	**15.6**	**<0.001**	LS<AP, MD

[a] Significant differences between groups were examined using the chi-square test for categorical variables, and ANOVA for continuous variables.

[b] Multiple comparison were performed by chi-square test for categorical variables, and Tukey's HSD for continuous variables.

Table 2-3. Comparison of clinical characteristics between LS, DPD, and MD group

	Missing, n (%)	Total (n=90)	Low severity (n = 28)	Depressive-psychosis dominant (n=40)	Manic dominant (n=22)	Statistics[a]	p-value	Post_hoc[b]
CGI-S : baseline, mean ± SD	0 (0.0)	**3.5 ±0.8**	**3.1 ±0.6**	**3.7 ±0.8**	**3.8 ±0.8**	**7.2**	**0.001**	LS<DPD,ABD
CTQ : total, mean ± SD	5 (5.6)	64.8 ±12.7	60.8 ±10.3	66.7 ±13.5	65.9 ±13.1	1.8	0.168	
CTQ : physical abuse, mean ± SD	5 (5.6)	9.5 ±4.6	9.5 ±4.6	9.3 ±4.4	10.0 ±5.0	0.2	0.842	
CTQ : sexual abuse, mean ± SD	5 (5.6)	5.9 ±2.0	5.6 ±1.3	6.3 ±2.6	5.6 ±1.3	1.5	0.239	
CTQ : emotional abuse, mean ± SD	5 (5.6)	11.9 ±5.5	10.8 ±4.9	12.2 ±5.6	12.7 ±6.0	0.8	0.455	
CTQ : emotional neglect, mean ± SD	5 (5.6)	**20.0 ±5.5**	**17.4 ±5.5**	**21.6 ±5.3**	**20.2 ±5.1**	**4.8**	**0.011**	LS<DPD
CTQ : physical neglect, mean ± SD	5 (5.6)	**10.3 ±3.1**	**9.0 ±2.0**	**11.1 ±3.3**	**10.4 ±3.4**	**3.4**	**0.039**	LS<DPD
BIS : total, mean ± SD	3 (3.3)	**70.7 ±10.7**	**65.1 ±8.9**	**74.2 ±10.9**	**71.4 ±9.9**	**6.5**	**0.002**	LS<DPD
BIS : attentional, mean ± SD	3 (3.3)	**19.0 ±5.1**	**16.4 ±6.1**	**20.1 ±4.3**	**20.3 ±3.9**	**5.7**	**0.005**	LS<DPD,ABD
BIS : motor, mean ± SD	3 (3.3)	31.2 ±6.8	31.1 ±9.3	31.9 ±5.5	29.9 ±5.2	0.6	0.564	
BIS : non-planning, mean ± SD	3 (3.3)	**20.6 ±5.5**	**17.7 ±6.1**	**22.2 ±5.0**	**21.2 ±4.0**	**6.3**	**0.003**	LS<DPD
BRIAN : total, mean ± SD	1 (1.1)	**52.4 ±11.7**	**47.0 ±11.9**	**56.0 ±10.7**	**52.6 ±10.9**	**5.3**	**0.007**	LS<DPD
BRIAN : sleep, mean ± SD	1 (1.1)	17.7 ±11.4	15.5 ±10.6	17.7 ±10.9	20.8 ±13.1	1.3	0.271	
BRIAN : activity, mean ± SD	1 (1.1)	**12.0 ±3.9**	**10.6 ±3.7**	**13.0 ±3.7**	**11.7 ±4.3**	**3.3**	**0.041**	LS<DPD
BRIAN : social, mean ± SD	1 (1.1)	**9.8 ±3.1**	**7.9 ±3.2**	**11.0 ±2.7**	**10.0 ±2.4**	**10.5**	**<0.001**	LS<DPD,ABD
BRIAN : eating, mean ± SD	1 (1.1)	9.7 ±3.0	8.7 ±3.0	10.4 ±3.0	9.7 ±2.9	2.7	0.075	
BRIAN : chronotype, mean ± SD	1 (1.1)	6.7 ±1.6	6.9 ±1.8	6.6 ±1.7	6.9 ±1.2	0.5	0.611	
CSM total, mean ± SD	5 (5.6)	27.8 ±7.1	29.4 ±7.2	26.7 ±6.9	27.7 ±7.3	1.0	0.356	
SPAQ : GSS total, mean ± SD	6 (6.7)	11.1 ±6.0	9.7 ±4.6	11.6 ±7.2	12.0 ±4.9	1.0	0.375	
SPAQ : sleep, mean ± SD	6 (6.7)	1.9 ±1.2	1.6 ±1.0	2.0 ±1.3	2.2 ±1.0	1.7	0.184	
SPAQ : social, mean ± SD	6 (6.7)	1.8 ±1.2	1.6 ±1.0	1.8 ±1.3	2.0 ±1.2	0.7	0.522	
SPAQ : general mood, mean ± SD	6 (6.7)	2.0 ±1.3	2.0 ±1.4	2.1 ±1.5	2.0 ±0.8	0.1	0.935	
SPAQ : weight, mean ± SD	6 (6.7)	**1.7 ±1.1**	**1.2 ±0.6**	**1.8 ±1.3**	**2.0 ±1.0**	**3.6**	**0.031**	LS<ABD
SPAQ : energy, mean ± SD	6 (6.7)	2.1 ±1.3	1.9 ±1.3	2.1 ±1.4	2.2 ±1.0	0.3	0.770	
SPAQ : appetite, mean ± SD	6 (6.7)	1.6 ±1.0	1.4 ±0.6	1.8 ±1.2	1.7 ±1.0	1.0	0.385	
SPAQ	6 (6.7)							
No, n (%)								
SAD+S-SAD, n (%)								
WHOQOL-BREF : total, mean ± SD	5 (5.6)	**67.5 ±14.5**	**79.8 ±12.7**	**58.3 ±10.2**	**70.5 ±11.2**	**29.1**	**<0.001**	DPD<MD<LS
WHOQOL-BREF : overall, mean ± SD	5 (5.6)	**5.0 ±1.3**	**5.6 ±1.3**	**4.3 ±1.2**	**5.6 ±0.9**	**13.4**	**<0.001**	DPD< LS,MD

[a] Significant differences between groups were examined using the chi-square test for categorical variables, and ANOVA for continuous variables

[b] Multiple comparison were performed by chi-square test for categorical variables, and Tukey's HSD for continuous variables.

WHOQOL-BREF : physical, mean ± SD	5 (5.6)	**18.0 ±4.9**	**22.1 ±4.3**	**15.3 ±3.1**	**18.4 ±4.9**	**23.1**	**<0.001**	DPD<MD<LS
WHOQOL-BREF : psychological, mean ± SD	5 (5.6)	**13.0 ±4.3**	**16.2 ±4.2**	**10.6 ±3.1**	**14.0 ±3.2**	**21.5**	**<0.001**	DPD<LS, MD
WHOQOL-BREF : social relationship, mean ± SD	5 (5.6)	**8.2 ±2.6**	**9.4 ±2.2**	**7.2 ±2.3**	**8.8 ±2.9**	**6.9**	**0.002**	DPD< LS, MD
WHOQOL-BREF : environment, mean ± SD	5 (5.6)	**23.2 ±5.3**	**26.4 ±4.7**	**20.9 ±5.2**	**23.7 ±4.1**	**10.2**	**<0.001**	DPD<LS
CD-RISC : total, mean ± SD	5 (5.6)	**40.7 ±17.5**	**51.4 ±17.6**	**31.8 ±14.9**	**45.0 ±13.3**	**13.4**	**<0.001**	DPD < LS,MD
GAF : present, mean ± SD	0 (0.0)	**57.5 ±9.4**	**61.4 ±8.8**	**55.5 ±8.1**	**56.4 ±11.2**	**3.7**	**0.030**	DPD<LS
GAF : past one year, mean ± SD	0 (0.0)	72.6 ±12.7	72.8 ±13.0	72.4 ±12.2	72.9 ±13.7	0.0	0.983	
GAF : before symptom onset, mean ± SD	0 (0.0)	80.5 ±10.6	80.9 ±8.6	80.5 ±8.3	80.0 ±15.9	0.0	0.957	
GF social : present, mean ± SD	0 (0.0)	6.0 ±1.3	6.2 ±1.6	5.8 ±1.0	6.3 ±1.5	1.4	0.255	
GF social : lowest, mean ± SD	0 (0.0)	5.7 ±1.4	5.6 ±1.5	5.6 ±1.2	6.0 ±1.5	0.8	0.451	
GF social : highest, mean ± SD	0 (0.0)	7.2 ±1.4	7.3 ±1.6	7.1 ±1.3	7.4 ±1.1	0.4	0.697	
GF role : present, mean ± SD	0 (0.0)	6.2 ±1.3	6.4 ±1.4	5.9 ±1.2	6.6 ±1.2	3.0	0.057	
GF role : lowest, mean ± SD	0 (0.0)	5.8 ±1.3	5.8 ±1.4	5.8 ±1.2	6.1 ±1.3	0.5	0.589	
GF role : highest, mean ± SD	0 (0.0)	7.4 ±1.3	7.4 ±1.4	7.3 ±1.3	7.7 ±1.1	0.6	0.547	

Table 2-4. Differentially expressed peptides (ANOVA) identified by LC-MRM analysis between LS, DPD, and MD group.

Protein ID	Gene name	Peptide	LS vs DPD vs MD			LS vs DPD		LS vs MD		DPD vs MD	
			F-statistics	p-value (ANOVA)	q-value[a]	Mean difference	p-value (Tukey's HSD)	Mean difference	p-value (Tukey's HSD)	Mean difference	p-value (Tukey's HSD)
Q9UGM5	FETUB	AIFYMNNPSR	1.467	0.236	0.997	-0.153	0.625	0.147	0.723	0.300	0.216
Q9UGM5	FETUB	LVVLPFPK	0.278	0.758	0.997	0.078	0.966	0.263	0.745	0.186	0.844
P05452	TETN	LDTLAQEVALLK	0.397	0.674	0.997	-0.070	0.957	0.167	0.829	0.237	0.649
P21333	FLNA	AGVAPLQVK	1.150	0.321	0.997	-0.472	0.362	-0.503	0.421	-0.031	0.996
P02748	CO9	LSPIYNLVPVK	0.703	0.498	0.997	0.005	0.999	-0.136	0.581	-0.141	0.509
P10643	CO7	AASGTQNNVLR	0.992	0.375	0.997	-0.250	0.546	0.074	0.961	0.324	0.418
P06396	GELS	HVVPNEVVVQR	0.804	0.451	0.997	-0.296	0.420	-0.201	0.740	0.095	0.924
P06396	GELS	EGGQTAPASTR	0.547	0.581	0.997	-0.106	0.582	-0.028	0.972	0.078	0.776
P06396	GELS	TGAQELLR	0.235	0.791	0.997	-0.141	0.809	-0.017	0.998	0.124	0.869
P06727	APOA4	LEPYADQLR	0.777	0.463	0.997	0.178	0.429	0.107	0.794	-0.071	0.889
P06727	APOA4	ISASAEELR	0.521	0.596	0.997	0.154	0.566	0.086	0.875	-0.068	0.908
P06727	APOA4	LAPLAEDVR	0.283	0.754	0.997	0.102	0.769	0.017	0.995	-0.086	0.853
P10645	CMGA	ELQDLALQGAK	0.865	0.425	0.997	0.028	0.995	-0.352	0.529	-0.379	0.427
P10645	CMGA	SGELEQEEER	0.380	0.685	0.997	-0.246	0.662	-0.122	0.926	0.124	0.914
P07225	PROS	VYFAGFPR	0.008	0.992	0.997	0.013	0.992	0.012	0.995	-0.001	1.000
P07225	PROS	QLAVLDK	0.132	0.877	0.997	0.040	0.890	0.043	0.904	0.003	0.999
P01009	A1AT	SVLGQLGITK	0.087	0.917	0.997	0.034	0.909	0.018	0.980	-0.016	0.982
P01009	A1AT	AVLTIDEK	0.340	0.713	0.997	-0.120	0.801	0.028	0.991	0.148	0.747
P02647	APOA1	DYVSQFEGSALGK	0.038	0.962	0.997	0.097	0.964	0.096	0.974	-0.001	1.000
P02647	APOA1	LLDNWDSVTSTFSK	1.019	0.365	0.997	0.108	0.355	0.091	0.571	-0.016	0.979
P02647	APOA1	QGLLPVLESFK	0.330	0.720	0.997	0.002	1.000	0.155	0.762	0.153	0.736
P02652	APOA2	SPELQAEAK	0.311	0.733	0.997	0.076	0.750	0.077	0.799	0.002	1.000
P02652	APOA2	EQLTPLIK	0.342	0.712	0.997	0.055	0.834	0.088	0.702	0.034	0.942
P04114	APOB	SVSLPSLDPASAK	0.288	0.751	0.997	0.125	0.838	0.184	0.751	0.059	0.967
P04114	APOB	TEVIPPLIENR	0.227	0.797	0.997	0.054	0.959	0.151	0.782	0.097	0.889
P04114	APOB	GVISIPR	0.660	0.519	0.997	0.275	0.504	0.223	0.712	-0.052	0.979
P05090	APOD	IPTTFENGR	1.031	0.361	0.997	0.002	1.000	0.237	0.432	0.235	0.387
P05090	APOD	NILTSNNIDVK	0.828	0.440	0.997	-0.040	0.991	0.377	0.555	0.417	0.437
P08519	APOA	GTLSTTITGR	0.563	0.572	0.997	0.404	0.609	0.040	0.996	-0.364	0.706
P23560	BDNF	QYFYETK	0.734	0.483	0.997	0.213	0.845	0.537	0.451	0.324	0.714
P00450	CERU	ALYLQYTDETFR	0.413	0.663	0.997	-0.029	0.972	-0.128	0.659	-0.100	0.748
P00450	CERU	DIFTGLIGPMK	0.386	0.681	0.997	-0.100	0.789	-0.147	0.681	-0.047	0.955
P00450	CERU	DDEEFIESNK	0.394	0.676	0.997	-0.371	0.776	0.091	0.989	0.462	0.712
P09341	GROA	NIQSVNVK	0.605	0.549	0.997	-0.083	0.899	-0.239	0.522	-0.156	0.727
P02790	HEMO	QGHNSVFLIK	0.377	0.687	0.997	0.067	0.662	0.041	0.889	-0.026	0.949
P02790	HEMO	GGYTLVSGYPK	0.226	0.799	0.997	-0.040	0.987	0.145	0.879	0.184	0.786
P02790	HEMO	LHIMAGR	0.127	0.881	0.997	-0.005	0.999	0.058	0.910	0.063	0.882
Q14624	ITIH4	FAHTVVTSR	1.762	0.178	0.997	0.029	0.939	-0.140	0.336	-0.168	0.164
Q14624	ITIH4	ILDDLSPR	0.004	0.996	0.997	-0.002	1.000	0.019	0.997	0.021	0.996
Q14624	ITIH4	GSEMVVAGK	1.213	0.302	0.997	0.084	0.635	-0.066	0.807	-0.150	0.289
Q14624	ITIH4	ETLFSVMPGLK	0.440	0.646	0.997	0.061	0.854	-0.052	0.919	-0.113	0.632
P03971	MIS	ELSVDLR	1.954	0.148	0.997	-0.585	0.551	0.592	0.633	1.177	0.131
P03971	MIS	YGNHVVLLLK	1.452	0.240	0.997	0.183	0.300	0.007	0.999	-0.176	0.382
P14174	MIF	LLCGLLAER	0.112	0.894	0.997	0.082	0.886	0.036	0.983	-0.046	0.968
P80511	S10AC	ELANTIK	0.698	0.500	0.997	-0.203	0.617	0.035	0.989	0.237	0.566
P02743	SAMP	VFVFPR	0.757	0.472	0.997	0.418	0.582	-0.058	0.992	-0.476	0.547
P02743	SAMP	VGEYSLYIGR	0.182	0.834	0.997	-0.077	0.919	-0.135	0.824	-0.058	0.959
P02743	SAMP	QGYFVEAQPK	0.168	0.846	0.997	-0.087	0.860	-0.092	0.881	-0.005	1.000
P08254	MMP3	DLVFIFK	1.998	0.142	0.997	0.181	0.561	0.407	0.118	0.226	0.462

[a] q-value means adjusted p-value for multiple comparison by using Benjamini-Hochberg FDR.

P01023	A2MG	TEHPFTVEEFVLPK	0.243	0.785	0.997	-0.179	0.879	0.080	0.981	0.259	0.793
P01023	A2MG	QGIPFFGQVR	0.130	0.879	0.997	-0.075	0.873	-0.061	0.936	0.014	0.996
P51884	LUM	ILGPLSYSK	0.310	0.734	0.997	0.063	0.758	0.067	0.791	0.004	0.999
Q86UX7	URP2	LTQLYEQAR	1.542	0.220	0.997	-0.542	0.436	-0.866	0.210	-0.323	0.774
P02774	VTDB	HLSLLTTLSNR	0.944	0.393	0.997	-0.127	0.436	-0.013	0.994	0.114	0.560
Q08380	LG3BP	ASHEEVEGLVEK	0.160	0.852	0.997	-0.024	0.997	0.181	0.895	0.205	0.849
P12259	**FA5**	**SQHLDNFSNQIGK**	**3.126**	**0.049**	**0.997**	0.237	0.057	0.048	0.911	-0.189	0.202
P12259	FA5	NFFNPPIISR	0.163	0.850	0.997	-0.032	0.953	-0.072	0.836	-0.040	0.939
P04275	VWF	VIVIPVGIGPHANLK	0.032	0.969	0.997	0.008	0.998	-0.027	0.982	-0.035	0.966
P04275	VWF	EAPDLVLQR	0.105	0.901	0.997	-0.077	0.897	-0.063	0.945	0.013	0.997
P04275	VWF	VTVFPIGIGDR	0.043	0.958	0.997	0.011	0.998	0.061	0.959	0.050	0.968
P04275	VWF	ILAGPAGDSNVVK	0.166	0.847	0.997	-0.092	0.847	-0.084	0.903	0.009	0.999
P01042	KNG1	TVGSDTFYSFK	0.469	0.627	0.997	-0.035	0.985	0.179	0.743	0.214	0.613
P01042	KNG1	QVVAGLNFR	0.059	0.943	0.997	0.017	0.975	-0.011	0.992	-0.028	0.942
P55058	PLTP	AGALQLLLVGDK	0.654	0.522	0.997	-0.156	0.591	0.002	1.000	0.157	0.630
P05154	IPSP	TLYLADTFPTNFR	0.144	0.866	0.997	-0.031	0.995	0.156	0.912	0.186	0.860
P05154	IPSP	EDQYHYLLDR	0.733	0.484	0.997	-0.183	0.701	0.097	0.928	0.280	0.490
P05154	IPSP	FSIEGSYQLEK	0.226	0.798	0.997	0.055	0.905	0.100	0.784	0.044	0.946
P00748	FA12	TEQAAVAR	0.413	0.663	0.997	0.131	0.685	0.135	0.741	0.004	1.000
P00748	FA12	VVGGLVALR	0.267	0.766	0.997	0.050	0.947	0.135	0.747	0.085	0.876
Q04756	HGFA	VANYVDWINDR	2.976	0.056	0.997	0.197	0.082	0.016	0.988	-0.182	0.158
P03952	KLKB1	DSVTGTLPK	0.079	0.924	0.997	0.036	0.951	-0.009	0.998	-0.045	0.934
P03952	KLKB1	TGAVSGHSLK	0.437	0.648	0.997	0.054	0.795	-0.025	0.965	-0.079	0.660
P03952	KLKB1	YSPGGTPTAIK	0.167	0.846	0.997	0.075	0.833	0.051	0.939	-0.024	0.984
P00742	FA10	TGIVSGFGR	0.151	0.860	0.997	-0.054	0.852	-0.042	0.930	0.012	0.993
Q13103	SPP24	VNSQSLSPYLFR	0.811	0.448	0.997	-0.046	0.876	0.083	0.727	0.129	0.414
Q13103	SPP24	VSAQQVQGVHAR	0.770	0.466	0.997	-0.272	0.471	-0.068	0.965	0.204	0.693
P20742	PZP	ATVLNYLPK	0.082	0.922	0.997	0.024	0.992	0.087	0.918	0.064	0.949
P20742	PZP	SSGSLLNNAIK	0.234	0.792	0.997	-0.123	0.931	0.125	0.946	0.248	0.779
P20742	PZP	YGAATFTR	0.021	0.979	0.997	0.017	0.995	0.041	0.977	0.024	0.991
P04406	G3P	VIPELNGK	0.995	0.374	0.997	0.113	0.783	0.276	0.340	0.163	0.646
Q16610	ECM1	NVALVSGDTENAK	0.264	0.769	0.997	-0.021	0.996	0.154	0.834	0.175	0.764
P02654	APOC1	EFGNTLEDK	0.469	0.627	0.997	-0.040	0.984	0.195	0.744	0.235	0.612
P18428	LBP	LAEGFPLPLLK	0.666	0.516	0.997	-0.203	0.679	0.071	0.965	0.273	0.546
P43251	BTD	LSSGLVTAALYGR	0.179	0.836	0.997	-0.051	0.821	-0.031	0.947	0.020	0.974
Q9UHG3	PCYOX	YQSHDYAFSSVEK	1.424	0.246	0.997	0.331	0.458	0.523	0.237	0.192	0.797
P06276	CHLE	NIAAFGGNPK	0.497	0.610	0.997	-0.037	0.989	0.236	0.717	0.273	0.600
P06276	CHLE	IFFPGVSEFGK	0.407	0.667	0.997	-0.162	0.926	0.260	0.863	0.423	0.641
P06276	CHLE	FWTSFFPK	0.052	0.949	0.997	0.026	0.955	0.002	1.000	-0.024	0.967
P26927	HGFL	SPLNDFQVLR	0.344	0.710	0.997	-0.104	0.707	-0.030	0.978	0.074	0.860
P04217	A1BG	HQFLLTGDTQGR	0.426	0.654	0.997	-0.066	0.947	0.142	0.828	0.208	0.628
P04217	A1BG	SGLSTGWTQLSK	0.638	0.531	0.997	0.085	0.527	0.075	0.688	-0.010	0.992
P04217	A1BG	ATWSGAVLAGR	0.073	0.930	0.997	-0.032	0.930	-0.008	0.996	0.023	0.966
P01019	ANGT	DPTFIPAPIQAK	0.017	0.983	0.997	-0.026	0.982	-0.016	0.995	0.010	0.997
P01019	ANGT	ALQDQLVLVAAK	0.231	0.794	0.997	-0.134	0.935	0.145	0.943	0.279	0.780
P61769	B2MG	VNHVTLSQPK	0.034	0.967	0.997	-0.023	0.973	0.000	1.000	0.024	0.976
P01024	CO3	LVAYYTLIGASGQR	0.125	0.882	0.997	-0.042	0.976	0.066	0.957	0.108	0.871
P01024	CO3	IHWESASLLR	0.014	0.986	0.997	-0.018	0.997	0.023	0.996	0.041	0.985
O00533	NCHL1	VIAVNEVGR	0.330	0.720	0.997	0.072	0.701	0.033	0.945	-0.039	0.914
P49747	COMP	DTDLDGFPDEK	0.545	0.582	0.997	-0.311	0.553	-0.205	0.824	0.106	0.942
P00488	F13A	VEYVIGR	0.312	0.733	0.997	-0.180	0.773	-0.003	1.000	0.177	0.807
P00488	F13A	GTYIPVPIVSELQSGK	0.067	0.936	0.997	-0.027	0.990	0.051	0.973	0.078	0.929
Q9NPH3	IL1AP	VAFPLEVVQK	0.136	0.873	0.997	0.081	0.921	0.121	0.872	0.040	0.983
Q06033	ITIH3	EVSFDVELPK	0.250	0.779	0.997	-0.133	0.916	0.116	0.952	0.248	0.768
Q06033	ITIH3	GISMLNK	0.352	0.705	0.997	0.050	0.910	-0.058	0.908	-0.108	0.684
Q06033	ITIH3	DYIFGNYIER	0.240	0.787	0.997	-0.160	0.817	-0.006	1.000	0.154	0.851
P04180	LCAT	SSGLVSNAPGVQIR	0.570	0.568	0.997	0.026	0.984	0.178	0.582	0.152	0.636
P61626	LYSC	WESGYNTR	0.058	0.944	0.997	0.001	1.000	-0.034	0.955	-0.035	0.946
O00187	MASP2	WPEPVFGR	0.493	0.613	0.997	0.090	0.616	0.023	0.977	-0.067	0.792
P11226	MBL2	FQASVATPR	0.848	0.432	0.997	0.256	0.415	0.208	0.647	-0.048	0.973

P11226	MBL2	EEAFLGITDEK	1.638	0.200	0.997	0.293	0.177	0.210	0.509	-0.083	0.885
P11226	MBL2	TEGQFVDLTGNR	1.001	0.372	0.997	0.160	0.598	0.263	0.355	0.103	0.830
P40925	MDHC	GEFVTTVQQR	0.111	0.895	0.997	-0.100	0.974	0.133	0.966	0.233	0.885
O60462	NRP2	DGDSESADLLGK	0.015	0.985	0.997	0.055	0.991	0.084	0.985	0.029	0.998
P19652	A1AG2	TEDTIFLR	0.369	0.692	0.997	-0.286	0.769	0.041	0.996	0.327	0.744
P02763	A1AG1	SDVVYTDWK	0.258	0.773	0.997	-0.127	0.789	-0.131	0.828	-0.004	1.000
Q96PD5	PGRP2	TFTLLDPK	1.668	0.195	0.997	0.123	0.191	0.037	0.892	-0.086	0.492
Q96PD5	PGRP2	WGAAPYR	0.290	0.749	0.997	-0.037	0.997	-0.382	0.769	-0.345	0.782
P14618	KPYM	LDIDSPPITAR	0.439	0.646	0.997	-0.059	0.957	-0.221	0.634	-0.162	0.753
P62937	PPIA	VSFELFADK	1.401	0.252	0.997	-0.057	0.943	0.254	0.429	0.311	0.234
P62937	PPIA	FEDENFILK	1.397	0.253	0.997	-0.060	0.986	0.593	0.370	0.653	0.252
P32119	**PRDX2**	**QITVNDLPVGR**	**5.101**	**0.008**	**0.997**	0.284	0.238	**0.641**	**0.005**	0.358	0.141
P22891	PROZ	APDLQDLPWQVK	0.060	0.942	0.997	-0.051	0.950	-0.003	1.000	0.049	0.961
P22891	PROZ	DFAEHLLIPR	0.056	0.945	0.997	-0.060	0.962	-0.080	0.948	-0.021	0.996
O00391	QSOX1	DTGAALLAESR	0.204	0.816	0.997	-0.092	0.834	-0.099	0.856	-0.007	0.999
P61224	RAP1B	LVVLGSGGVGK	0.664	0.518	0.997	-0.218	0.697	-0.348	0.506	-0.129	0.896
P06702	S10A9	LTWASHEK	0.242	0.786	0.997	-0.114	0.868	0.040	0.987	0.153	0.802
P14151	LYAM1	DNYTDLVAIQNK	0.797	0.454	0.997	-0.423	0.491	-0.050	0.992	0.372	0.620
P02741	CRP	AFVFPK	1.273	0.285	0.997	-0.384	0.628	-0.766	0.255	-0.382	0.673
P02741	CRP	ESDTSYVSLK	0.845	0.433	0.997	-0.254	0.628	-0.401	0.422	-0.147	0.874
P04003	C4BPA	QSSSYSFFK	2.798	0.066	0.997	-0.418	0.089	-0.046	0.977	0.371	0.189
P20851	C4BPB	ALLAFQESK	0.037	0.963	0.997	0.033	0.987	0.069	0.960	0.036	0.987
P01011	AACT	ADLSGITGAR	0.241	0.787	0.997	-0.014	0.998	0.144	0.841	0.159	0.786
P01011	AACT	NLAVSQVVHK	0.175	0.840	0.997	0.001	1.000	0.083	0.866	0.082	0.850
P08185	CBG	AQLLQGLGFNLTER	1.919	0.153	0.997	-0.018	0.995	0.362	0.236	0.380	0.162
P08185	CBG	ITQDAQLK	0.470	0.627	0.997	0.079	0.704	-0.007	0.998	-0.086	0.697
P05546	HEP2	SVNDLYIQK	0.179	0.837	0.997	0.023	0.995	0.157	0.844	0.134	0.867
P05546	HEP2	QFPILLDFK	0.215	0.807	0.997	-0.060	0.972	0.126	0.910	0.186	0.790
P05546	HEP2	TLEAQLTPR	0.003	0.997	0.997	0.009	0.997	0.003	1.000	-0.006	0.999
P04278	SHBG	IALGGLLFPASNLR	1.295	0.279	0.997	-0.398	0.250	-0.201	0.763	0.196	0.743
P04278	SHBG	QAEISASAPTSLR	0.945	0.393	0.997	-0.220	0.588	0.084	0.944	0.304	0.420
P04278	SHBG	LDVDQALNR	0.417	0.661	0.997	-0.227	0.634	-0.123	0.905	0.104	0.920
P37802	TAGL2	QMEQISQFLQAAER	0.277	0.759	0.997	-0.139	0.863	-0.224	0.750	-0.085	0.953
P37802	TAGL2	TLMNLGGLAVAR	0.686	0.506	0.997	-0.225	0.729	-0.396	0.484	-0.171	0.854
P02787	TRFE	DGAGDVAFVK	0.260	0.772	0.997	0.103	0.822	0.130	0.792	0.027	0.989
P02787	TRFE	EGYYGYTGAFR	0.121	0.886	0.997	-0.077	0.920	0.014	0.998	0.091	0.904
Q15582	BGH3	GDELADSALEIFK	0.187	0.830	0.997	0.083	0.837	0.016	0.995	-0.068	0.904
Q9Y490	TLN1	GLAGAVSELLR	1.523	0.224	0.997	-0.398	0.383	-0.571	0.230	-0.173	0.854
Q9Y490	TLN1	TLAESALQLLYTAK	2.861	0.063	0.997	-0.314	0.432	-0.698	0.049	-0.385	0.338
P22105	TENX	LQVVPVAANQR	1.006	0.370	0.997	-0.304	0.366	-0.089	0.936	0.215	0.647
P22105	TENX	ILISGLEPSTPYR	0.283	0.754	0.997	0.022	0.992	0.149	0.764	0.127	0.798
P22105	TENX	TLSPVLESPR	0.343	0.711	0.997	0.016	0.998	0.222	0.739	0.206	0.741
P67936	TPM4	TIDDLEEK	2.352	0.101	0.997	-0.425	0.290	0.185	0.837	0.610	0.115
P18206	VINC	ELLPVLISAMK	0.914	0.405	0.997	-0.154	0.785	-0.362	0.371	-0.208	0.682
P04004	VTNC	FEDGVLDPDYPR	0.362	0.697	0.997	-0.167	0.725	-0.018	0.997	0.149	0.802
P04004	VTNC	QPQFISR	0.498	0.610	0.997	-0.108	0.673	0.002	1.000	0.111	0.700
O75083	WDR1	DYSGQGVVK	0.620	0.540	0.997	-0.091	0.924	-0.305	0.520	-0.213	0.690
P63104	1433Z	SVTEQGAELSNEER	0.354	0.703	0.997	0.095	0.925	0.246	0.680	0.151	0.846
P63104	1433Z	FLIPNASQAESK	0.121	0.886	0.997	-0.052	0.988	-0.192	0.881	-0.140	0.926

Table 2-5. Pearson's correlation coefficients (r) between 19 clinical variables and 29 peptides[a]

Information			Index[b]	Pearson's correlation coefficient																		
Protein ID	Gene name	Peptide		SIPS : positive	SIPS : negative	SIPS : disorganized	SIPS : general	BPSS current : severity	BPSS current : frequency	YMRS	HAM-D	QIDS	HAM-A	STAI-T	STAI-S	BIS	CTQ	CSM	SPAQ : GSS	WHOQOL-BREF	CD-RISC	GAF : present
P55058	PLTP	AGALQLLLVGDK	7	0.160	0.069	-0.053	0.185	-0.126	-0.055	-0.146	.216*	.237*	0.002	.291**	.265*	-0.106	-0.096	-.209*	-0.048	-.234*	-.236*	-0.057
P04003	C4BPA	QSSSYSFFK	7	0.198	0.156	-0.003	0.200	-0.163	-0.001	-0.091	.230*	.390**	0.129	.331**	.325**	0.116	0.063	-.258*	0.086	-.315**	-.285**	-0.180
P18428	LBP	LAEGFPLPLLK	5	0.147	0.125	0.003	0.095	-.211*	-0.092	-0.091	0.098	.259*	0.000	.224*	.239*	-0.034	-0.052	-0.203	-0.060	-0.194	-.214*	-0.140
P26927	HGFL	SPLNDFQVLR	4	0.102	0.166	0.085	.221*	-0.144	-0.071	-0.070	.214*	0.202	0.107	.234*	.260*	0.036	0.001	-0.031	-0.003	-0.168	-0.117	-0.196
P19652	A1AG2	TEDTIFLR	3	0.161	0.180	0.026	0.142	-0.189	-0.006	0.027	0.129	.250*	0.030	0.185	.208*	0.099	-0.005	-.236*	-0.014	-0.126	-0.199	-0.167
P02741	CRP	AFVFPK	3	-0.006	0.007	-0.027	0.009	0.191	0.092	-0.059	0.104	.242*	0.036	0.204	0.157	-0.018	-0.097	-.270*	0.049	-.287**	-0.167	-0.020
P05452	TETN	LDTLAQEVALLK	3	0.182	0.130	-0.020	0.055	-.217*	-0.044	0.023	0.067	.213*	-0.021	0.140	0.174	0.074	0.080	-.220*	-0.044	-0.082	-0.171	-0.145
P02774	VTDB	HLSLLTTLSNR	3	0.016	0.030	-0.012	0.048	-0.178	-0.062	-0.014	0.170	0.206	0.068	0.196	.208*	-0.008	-0.025	-.250*	0.057	-0.148	-.228*	-0.084
P04275	VWF	VTVFPIGIGDR	3	0.168	0.024	0.086	.220*	-0.073	0.068	0.036	.223*	0.125	0.117	0.205	0.192	-.229*	0.077	0.055	0.089	-0.116	-0.070	-0.128
Q96PD5	PGRP2	TFTLLDPK	3	-0.191	-.317**	-0.107	-0.095	-0.096	-0.125	0.002	-0.165	-0.119	-.271**	-0.173	-0.129	-.250*	-0.073	0.090	-0.019	0.199	0.097	0.136
P32119	PRDX2	QITVNDLPVGR	3	-.215*	-0.092	-0.094	-0.028	-.237*	-.282**	-0.196	-0.028	0.054	0.009	0.032	0.042	0.066	-0.123	-0.057	-0.135	-0.024	-0.103	0.137
P00450	CERU	DDEEFIESNK	2	0.020	0.101	-0.058	0.203	-0.166	-0.056	-0.076	0.154	.260*	0.065	0.193	0.193	0.010	-.223*	-0.182	-0.107	-0.117	-0.207	-0.123
Q06033	ITIH3	EVSFDVELPK	2	0.152	0.115	-0.029	0.137	-.216*	-0.066	0.025	0.131	.241*	-0.030	0.127	0.156	-0.009	0.039	-0.136	0.052	-0.057	-0.113	-0.145
P04004	VTNC	FEDGVLDPDYPR	2	0.133	0.149	0.007	0.145	-0.169	-0.010	0.016	0.109	.235*	0.047	0.188	0.203	0.062	-0.032	-.253*	-0.054	-0.154	-0.192	-0.192
Q16610	ECM1	NVALVSGDTENAK	2	0.154	0.180	0.023	0.087	-0.183	-0.034	0.038	0.119	.231*	0.025	0.075	0.119	0.065	0.011	-0.183	-0.042	-0.104	-.220*	-0.188
P67936	TPM4	TIDDLEEK	2	-0.062	-0.037	-0.074	0.092	-0.162	-0.162	-0.090	.231*	.231*	0.123	0.193	0.206	0.030	-0.131	-0.042	-0.038	-0.179	-0.089	0.109
P61769	B2MG	VNHVTLSQPK	2	0.086	0.108	0.079	.253*	-0.059	0.024	0.040	0.182	.223*	-0.010	0.104	0.130	-0.040	0.002	-0.053	-0.050	-0.083	-0.105	-0.194
P02743	SAMP	VGEYSLYIGR	2	0.000	0.011	-0.036	0.067	0.024	0.104	0.056	0.105	.218*	0.085	0.146	0.191	0.059	-0.147	-0.171	-0.082	-.213*	-0.084	-0.003
P04217	A1BG	HQFLLTGDTQGR	2	0.139	0.162	-0.010	0.102	-.239*	-0.070	0.020	0.094	.213*	-0.024	0.163	0.173	0.063	0.001	-0.183	-0.015	-0.084	-0.148	-0.159
P20742	PZP	SSGSLLNNAIK	2	0.162	0.140	-0.025	0.116	-.209*	-0.047	0.007	0.083	.211*	-0.022	0.119	0.142	0.053	0.038	-0.149	-0.053	-0.038	-0.126	-0.131
P06276	CHLE	IFFPGVSEFGK	2	0.140	0.170	-0.002	0.101	-.209*	-0.028	0.043	0.077	0.183	-0.016	0.138	0.172	0.088	0.042	-.218*	-0.029	-0.092	-0.185	-0.147
P14151	LYAM1	DNYTDLVAIQNK	2	0.004	-0.034	-0.136	0.053	-0.053	-0.102	-0.151	0.143	0.181	0.066	.227*	.212*	0.074	-0.065	0.012	0.152	-0.200	-0.155	0.129
P01042	KNG1	TVGSDTFYSFK	2	0.151	0.195	0.042	0.144	-.222*	-0.074	0.049	0.092	0.173	-0.016	0.150	0.188	0.068	0.007	-0.197	-0.054	-0.108	-0.180	-.211*
Q15582	BGH3	GDELADSALEIFK	2	-.208*	-0.071	0.111	-0.054	0.102	-0.006	-0.022	-0.020	-0.163	0.026	-0.129	-0.166	-0.106	0.066	.229*	-0.096	0.085	0.148	0.000
P02647	APOA1	LLDNWDSVTSTFSK	2	0.040	-0.152	-0.070	-0.031	-0.110	-0.032	-0.122	-0.095	-0.135	-0.112	-0.102	-0.130	-.229*	-0.108	0.202	-0.020	.214*	0.061	0.104
Q13103	SPP24	VSAQQVQGVHAR	2	-0.014	-0.125	-0.189	0.045	-0.053	0.027	-0.064	0.061	0.112	0.025	0.021	0.020	-0.060	0.124	-.232*	.239*	-0.078	-0.096	-0.029
Q9Y490	TLN1	TLAESALQLLYTAK	2	-.268*	-0.045	0.031	0.131	.227*	0.159	0.171	0.199	0.056	0.087	0.054	0.020	0.081	-0.056	0.013	-0.015	-0.131	0.077	-0.049
Q14624	ITIH4	FAHTVVTSR	2	-0.140	-.266*	-0.098	-0.005	0.095	0.091	0.021	0.002	0.020	-0.134	0.013	0.030	-.219*	-0.085	-0.003	0.107	-0.018	0.045	-0.010
P04114	APOB	GVISIPR	2	-0.107	-0.197	-.235*	-.221*	-0.081	-0.064	-0.060	-0.086	0.012	-0.167	0.014	-0.043	-0.037	-0.034	0.024	-0.080	0.025	0.068	0.184

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[a] Proteins that showed significant correlation with at least two of the 10 clinical variables were selected

[b] Number of psychiatric symptom indexes with p values ≤ 0.05

Table 3-1. List of baseline 110 clinical features included in models

Age, years	HAM-D : somatic
Gender (Male)	HAM-D : insomnia
Gender (Female)	HAM-D : appetite
Body mass Index	HAM-A
Education, years	YMRS : total
Social economic status (unstable)	YMRS : psychotic mania
Social economic status (stable)	YMRS : irritable mania
Initial classification (CHR-PRS)	YMRS : core feature of mania
Initial classification (CHR-BD)	BRIAN : total
Initial classification (CHR-MDD)	BRIAN : sleep
Initial classification (CHR-PL)	BRIAN : activity
Initial classification (CHR-Extended)	BRIAN : social
Duration of follow up	BRIAN : eating
Psychiatric family history	BRIAN : chronotype
History of treatment (No)	CGI-S
History of treatment (Medication-present)	GAF : present
History of treatment (Medication-past)	GAF : past one year
History of treatment (Psychosocial Intervention)	GAF : before symptom onset
Use of psychotropic medication	GF social : present
Use of antipsychotics	GF social : lowest
Use of Mood stabilizers	GF social : highest
Use of Anti-depressant	GF role : present
Use of Benzodiazepines	GF role : lowest
SIPS : positive	GF role : highest
SIPS : negative	QIDS total
SIPS : disorganized	BIS : total
SIPS : general	BIS : attention
SIPS : total	BIS : activity
BPSS current : total	BIS : plan
BPSS current : severity	PQ-B total
BPSS current : frequency	PQ-B threshold
BPSS current : T1 threshold	WHOQOL-BREF : total
BPSS current : T2 threshold	WHOQOL-BREF : overall
BPSS current : T3 threshold	WHOQOL-BREF : physical
BPSS current : T4 threshold	WHOQOL-BREF : psychological
BPSS current : threshold total	WHOQOL-BREF : social relationship
BPSS current : positive mania	WHOQOL-BREF : environment
BPSS current : negative mania	STAI-T
BPSS past : total	STAI-S
BPSS past : severity	CD-RISC : total
BPSS past : frequency	CTQ : total
BPSS past : T1 threshold	CTQ : Physical abuse
BPSS past : T2 threshold	CTQ : Sexual abuse
BPSS past : T3 threshold	CTQ : Emotional abuse
BPSS past : T4 threshold	CTQ : Emotional neglect
BPSS past : threshold total	CTQ : Physical neglect
BPSS past : positive mania	CSM total
BPSS past : negative mania	SPAQ : GSS total
PANSS : total	SPAQ : sleep
PANSS : positive	SPAQ : social
PANSS : negative	SPAQ : general mood
PANSS : general	SPAQ : weight
HAM-D : total	SPAQ : energy
HAM-D : anhedonia/retardation	SPAQ : appetite
HAM-D : guilt/agitation	SPAQ : SAD

Table 3-2. List of baseline 158 proteomic features included in models

Uniprot accession number	Gene names	Peptide	Uniprot accession number	Gene names	Peptide	Uniprot accession number	Gene names	Peptide
Q9UGM5	FETUB	AIFYMNNPSR	P02774	VTDB	HLSLLTTLSNR	P61626	LYSC	WESGYNTR
Q9UGM5	FETUB	LVVLPFPK	Q08380	LG3BP	ASHEEVEGLVEK	O00187	MASP2	WPEPVFGR
P05452	TETN	LDTLAQEVALLK	P12259	FA5	SQHLDNFSNQIGK	P11226	MBL2	FQASVATPR
P21333	FLNA	AGVAPLQVK	P12259	FA5	NFFNPPIISR	P11226	MBL2	EEAFLGITDEK
P02748	CO9	LSPIYNLVPVK	P04275	VWF	VIVIPVGIGPHANLK	P11226	MBL2	TEGQFVDLTGNR
P10643	CO7	AASGTQNNVLR	P04275	VWF	EAPDLVLQR	P40925	MDHC	GEFVTTVQQR
P06396	GELS	HVVPNEVVVQR	P04275	VWF	VTVFPIGIGDR	O60462	NRP2	DGDSESADLLGK
P06396	GELS	EGGQTAPASTR	P04275	VWF	ILAGPAGDSNVVK	P19652	A1AG2	TEDTIFLR
P06396	GELS	TGAQELLR	P01042	KNG1	TVGSDTFYSFK	P02763	A1AG1	SDVVYTDWK
P06727	APOA4	LEPYADQLR	P01042	KNG1	QVVAGLNFR	Q96PD5	PGRP2	TFTLLDPK
P06727	APOA4	ISASAEELR	P55058	PLTP	AGALQLLLVGDK	Q96PD5	PGRP2	WGAAPYR
P06727	APOA4	LAPLAEDVR	P05154	IPSP	TLYLADTFPTNFR	P14618	KPYM	LDIDSPPITAR
P10645	CMGA	ELQDLALQGAK	P05154	IPSP	EDQYHYLLDR	P62937	PPIA	VSFELFADK
P10645	CMGA	SGELEQEEER	P05154	IPSP	FSIEGSYQLEK	P62937	PPIA	FEDENFILK
P07225	PROS	VYFAGFPR	P00748	FA12	TEQAAVAR	P32119	PRDX2	QITVNDLPVGR
P07225	PROS	QLAVLDK	P00748	FA12	VVGGLVALR	P22891	PROZ	APDLQDLPWQVK
P01009	A1AT	SVLGQLGITK	Q04756	HGFA	VANYVDWINDR	P22891	PROZ	DFAEHLLIPR
P01009	A1AT	AVLTIDEK	P03952	KLKB1	DSVTGTLPK	O00391	QSOX1	DTGAALLAESR
P02647	APOA1	DYVSQFEGSALGK	P03952	KLKB1	TGAVSGHSLK	P61224	RAP1B	LVVLGSGGVGK
P02647	APOA1	LLDNWDSVTSTFSK	P03952	KLKB1	YSPGGTPTAIK	P06702	S10A9	LTWASHEK
P02647	APOA1	QGLLPVLESFK	P00742	FA10	TGIVSGFGR	P14151	LYAM1	DNYTDLVAIQNK
P02652	APOA2	SPELQAEAK	Q13103	SPP24	VNSQSLSPYLFR	P02741	CRP	AFVFPK
P02652	APOA2	EQLTPLIK	Q13103	SPP24	VSAQQVQGVHAR	P02741	CRP	ESDTSYVSLK
P04114	APOB	SVSLPSLDPASAK	P20742	PZP	ATVLNYLPK	P04003	C4BPA	QSSSYSFFK
P04114	APOB	TEVIPPLIENR	P20742	PZP	SSGSLLNNAIK	P20851	C4BPB	ALLAFQESK
P04114	APOB	GVISIPR	P20742	PZP	YGAATFTR	P01011	AACT	ADLSGITGAR
P05090	APOD	IPTTFENGR	P04406	G3P	VIPELNGK	P01011	AACT	NLAVSQVVHK
P05090	APOD	NILTSNNIDVK	Q16610	ECM1	NVALVSGDTENAK	P08185	CBG	AQLLQGLGFNLTER
P08519	APOA	GTLSTTITGR	P02654	APOC1	EFGNTLEDK	P08185	CBG	ITQDAQLK
P23560	BDNF	QYFYETK	P18428	LBP	LAEGFPLPLLK	P05546	HEP2	SVNDLYIQK
P00450	CERU	ALYLQYTDETFR	P43251	BTD	LSSGLVTAALYGR	P05546	HEP2	QFPILLDFK
P00450	CERU	DIFTGLIGPMK	Q9UHG3	PCYOX	YQSHDYAFSSVEK	P05546	HEP2	TLEAQLTPR
P00450	CERU	DDEEFIESNK	P06276	CHLE	NIAAFGGNPK	P04278	SHBG	IALGGLLFPASNLR
P09341	GROA	NIQSVNVK	P06276	CHLE	IFFPGVSEFGK	P04278	SHBG	QAEISASAPTSLR
P02790	HEMO	QGHNSVFLIK	P06276	CHLE	FWTSFFPK	P04278	SHBG	LDVDQALNR
P02790	HEMO	GGYTLVSGYPK	P26927	HGFL	SPLNDFQVLR	P37802	TAGL2	QMEQISQFLQAAER
P02790	HEMO	LHIMAGR	P04217	A1BG	HQFLLTGDTQGR	P37802	TAGL2	TLMNLGGLAVAR
Q14624	ITIH4	FAHTVVTSR	P04217	A1BG	SGLSTGWTQLSK	P02787	TRFE	DGAGDVAFVK
Q14624	ITIH4	ILDDLSPR	P04217	A1BG	ATWSGAVLAGR	P02787	TRFE	EGYYGYTGAFR
Q14624	ITIH4	GSEMVVAGK	P01019	ANGT	DPTFIPAPIQAK	Q15582	BGH3	GDELADSALEIFK
Q14624	ITIH4	ETLFSVMPGLK	P01019	ANGT	ALQDQLVLVAAK	Q9Y490	TLN1	GLAGAVSELLR
P03971	MIS	ELSVDLR	P61769	B2MG	VNHVTLSQPK	Q9Y490	TLN1	TLAESALQLLYTAK
P03971	MIS	YGNHVVLLLK	P01024	CO3	LVAYYTLIGASGQR	P22105	TENX	LQVVPVAANQR
P14174	MIF	LLCGLLAER	P01024	CO3	IHWESASLLR	P22105	TENX	ILISGLEPSTPYR
P80511	S10AC	ELANTIK	O00533	NCHL1	VIAVNEVGR	P22105	TENX	TLSPVLESPR
P02743	SAMP	VFVFPR	P49747	COMP	DTDLDGFPDEK	P67936	TPM4	TIDDLEEK
P02743	SAMP	VGEYSLYIGR	P00488	F13A	VEYVIGR	P18206	VINC	ELLPVLISAMK
P02743	SAMP	QGYFVEAQPK	P00488	F13A	GTYIPVPIVSELQSGK	P04004	VTNC	FEDGVLDPDYPR
P08254	MMP3	DLVFIFK	Q9NPH3	IL1AP	VAFPLEVVQK	P04004	VTNC	QPQFISR
P01023	A2MG	TEHPFTVEEFVLPK	Q06033	ITIH3	EVSFDVELPK	O75083	WDR1	DYSGQGVVK
P01023	A2MG	QGIPFFGQVR	Q06033	ITIH3	GISMLNK	P63104	1433Z	SVTEQGAELSNEER
P51884	LUM	ILGPLSYSK	Q06033	ITIH3	DYIFGNYIER	P63104	1433Z	FLIPNASQAESK
Q86UX7	URP2	LTQLYEQAR	P04180	LCAT	SSGLVSNAPGVQIR			

Table 3-3. Summary of prediction model

Model	Aim	Outcome	Target outcome to be predicted	Data modalities
Model 1. Prediction model of transition to psychiatric disorder				
Model 1a	Predict transition of prodromal stage to major psychiatric disorders using baseline clinical data	14 Transition 76 NT group	Transition	110 clinical variable
Model 1b	Predict transition of prodromal stage to major psychiatric disorders using baseline proteomic data		Transition	158 peptides
Model 1c	Predict transition of prodromal stage to major psychiatric disorders using baseline clinical and proteomic data		Transition	110 clinical variable+158 peptides
Model 2. Prediction model of disease course				
Model 2a	Predict disease course using baseline clinical data	39 recovered 44 not-recovered group	Symptom recover	110 clinical variable
Model 2b	Predict disease course using baseline proteomic data		Symptom recover	158 peptides
Model 2c	Predict disease course using baseline clinical and proteomic data		Symptom recover	110 clinical variable+158 peptides
Model 2d	Predict disease course using baseline clinical and proteomic data		Symptom recover	clinical variable (SF > 0.8 Model 1a) + peptides (SF >0.8 Model 1b)

Table 3-4. Comparison of baseline demographical characteristics between CHR-T vs. CHR-NT

	Missing, n (%)	CHR-T (n = 14)	CHR-NT (n = 76)	Statistics[a]	*p*-value
Age, years, mean ± SD	0 (0.0)	22.1 ± 4.8	21.6 ± 3.8	516.5	0.866
Gender, male, n (%)	0 (0.0)	7 (50.0)	33 (43.4)	0.2	0.772
Body mass Index, kg/m^2, mean ± SD	0 (0.0)	22.9 ± 4.0	22.3 ± 3.5	502.5	0.747
Education, years, mean ± SD	0 (0.0)	13.6 ± 2.6	13.4 ± 1.8	501.5	0.736
Social economic status	0 (0.0)			FE	0.749
Unstable, n (%)		3 (21.4)	22 (28.9)		
Stable, n (%)		11 (78.6)	54 (71.1)		
Initial classification (Study 1)	0 (0.0)				
CHR-PRS, n (%)		2 (14.3)	27 (35.5)	FE	0.211
CHR-BD, n (%)		**12 (85.7)**	**28 (36.8)**	**11.4**	**0.001**
CHR-MDD, n (%)		0 (0.0)	2 (2.6)	FE	>0.999
CHR-PL, n (%)		0 (0.0)	4 (5.3)	FE	>0.999
CHR-Extended, n (%)		0 (0.0)	15 (19.7)	FE	0.115
LPA classification (Study2)	0 (0.0)			FE	0.573
LS, n (%)		4 (28.6)	24 (31.6)		
DPD, n (%)		5 (35.7)	35 (46.1)		
MD, n (%)		5 (35.7)	17 (22.4)		
Duration of follow up, month , mean ± SD	**0 (0.0)**	**23.4 ± 9.5**	**17.1 ± 10.0**	**314.5**	**0.014**
Psychiatric family history, n (%)	0 (0.0)	4 (28.6)	36 (47.4)	1.7	0.248
History of treatment	0 (0.0)			4.2	0.223
No, n (%)	0 (0.0)	1 (7.1)	13 (17.1)		
Medication-present, n (%)	0 (0.0)	9 (64.3)	44 (57.9)		
Medication-past, n (%)	0 (0.0)	4 (28.6)	9 (11.8)		
Psychosocial intervention, n (%)	0 (0.0)	0 (0.0)	10 (13.2)		
Use of psychotropic medication, n (%)	0 (0.0)	9 (64.3)	5 (35.7)	0.2	0.772
Use of antipsychotics, n (%)	0 (0.0)	5 (38.5)	30 (40.0)	0.0	>0.999
Use of Mood stabilizers, n (%)	0 (0.0)	7 (53.8)	24 (32.4)	FE	0.208
Use of Anti-depressant, n (%)	0 (0.0)	1 (7.7)	13 (17.3)	FE	0.683
Use of Benzodiazepines, n (%)	0 (0.0)	5 (38.5)	23 (31.1)	FE	0.749

Significant findings at *p*-value<0.050 are in **bold**

[a] Significant differences between the CHR-T and CHR-NT were examined using the chi-square test and Fisher's exact test for categorical variables, and Mann-Whitney's U test for continuous variables.

Table 3-5. Comparison of baseline symptom characteristics between CHR-T vs. CHR-NT

	Missing, n (%)	CHR-T (n = 14)	CHR-NT (n = 76)	Statistics[a]	*p*-value
SIPS : positive, mean ± SD	0 (0.0)	4.4 ± 3.8	6.0 ± 5.1	450.0	0.365
SIPS : negative, mean ± SD	0 (0.0)	8.6 ± 6.7	9.7 ± 6.6	471.5	0.506
SIPS : disorganized, mean ± SD	0 (0.0)	2.6 ± 3.1	2.9 ± 2.8	461.0	0.425
SIPS : general, mean ± SD	0 (0.0)	6.7 ± 4.0	6.8 ± 3.9	527.5	0.962
SIPS : total, mean ± SD	0 (0.0)	22.4 ± 14.6	25.6 ± 14.0	430.0	0.260
BPSS current : total, mean ± SD	0 (0.0)	23.6 ± 16.4	16.7 ± 16.6	378.0	0.086
BPSS current : severity, mean ± SD	**0 (0.0)**	**12.6 ± 8.8**	**7.6 ± 7.4**	**334.5**	**0.026**
BPSS current : frequency, mean ± SD	0 (0.0)	11.0 ± 9.1	9.0 ± 9.8	436.0	0.284
BPSS current : T1 threshold, mean ± SD	**0 (0.0)**	**0.6 ± 0.8**	**0.1 ± 0.3**	**325.0**	**<0.001**
BPSS current : T2 threshold, mean ± SD	0 (0.0)	0.6 ± 0.6	0.8 ± 1.2	497.0	0.681
BPSS current : T3 threshold, mean ± SD	0 (0.0)	0.1 ± 0.3	0.0 ± 0.1	501.0	0.288
BPSS current : T4 threshold, mean ± SD	0 (0.0)	0.8 ± 1.1	0.5 ± 0.9	433.5	0.188
BPSS current : threshold total, mean ± SD	0 (0.0)	2.1 ± 1.7	1.3 ± 1.7	367.0	0.055
BPSS current : positive mania, mean ± SD	0 (0.0)	3.6 ± 4.4	2.2 ± 3.8	415.0	0.150
BPSS current : negative mania, mean ± SD	**0 (0.0)**	**8.9 ± 5.7**	**5.5 ± 4.5**	**337.0**	**0.028**
BPSS past : total, mean ± SD	0 (0.0)	33.9 ± 26.4	28.2 ± 28.0	456.5	0.403
BPSS past : severity, mean ± SD	0 (0.0)	16.9 ± 13.9	13.1 ± 13.2	435.0	0.282
BPSS past : frequency, mean ± SD	0 (0.0)	17.0 ± 13.5	15.1 ± 15.6	462.0	0.438
BPSS past : T1 threshold, mean ± SD	**0 (0.0)**	**0.7 ± 1.1**	**0.2 ± 0.5**	**372.0**	**0.008**
BPSS past : T2 threshold, mean ± SD	0 (0.0)	0.9 ± 1.0	1.0 ± 1.3	516.5	0.858
BPSS past : T3 threshold, mean ± SD	0 (0.0)	0.1 ± 0.4	0.2 ± 0.6	502.0	0.605
BPSS past : T4 threshold, mean ± SD	0 (0.0)	1.6 ± 2.1	1.4 ± 2.3	489.5	0.613
BPSS past : threshold total, mean ± SD	0 (0.0)	3.3 ± 3.3	2.7 ± 3.1	460.0	0.418
BPSS past : positive mania, mean ± SD	0 (0.0)	7.0 ± 7.0	5.5 ± 6.5	452.0	0.361
BPSS past : negative mania, mean ± SD	0 (0.0)	9.9 ± 8.0	7.6 ± 7.3	430.0	0.256
PANSS : total, mean ± SD	0 (0.0)	51.9 ± 12.0	52.0 ± 12.8	524.0	0.932
PANSS : positive, mean ± SD	0 (0.0)	10.9 ± 3.0	10.4 ± 2.5	498.5	0.711
PANSS : negative, mean ± SD	0 (0.0)	11.4 ± 4.3	12.4 ± 5.3	473.0	0.515
PANSS : general, mean ± SD	0 (0.0)	29.7 ± 7.1	29.3 ± 7.4	522.5	0.919
HDRS : total, mean ± SD	0 (0.0)	12.8 ± 7.0	11.6 ± 6.5	490.5	0.649
HDRS : anhedonia/retardation, mean ± SD	0 (0.0)	4.9 ± 2.5	4.1 ± 2.2	479.5	0.559
HDRS : guilt/agitation, mean ± SD	0 (0.0)	3.0 ± 2.3	3.4 ± 2.3	477.5	0.546
HDRS : somatic, mean ± SD	0 (0.0)	1.6 ± 1.9	1.5 ± 1.6	515.0	0.853
HDRS : insomnia, mean ± SD	0 (0.0)	1.7 ± 1.3	1.8 ± 1.9	486.5	0.606
HDRS : appetite, mean ± SD	0 (0.0)	0.7 ± 1.0	0.5 ± 0.9	471.0	0.444
HAM-A total, mean ± SD	0 (0.0)	10.4 ± 6.2	10.6 ± 6.6	495.5	0.689
YMRS : total, mean ± SD	0 (0.0)	3.1 ± 2.9	2.5 ± 3.2	421.5	0.207
YMRS : psychotic mania, mean ± SD	0 (0.0)	1.2 ± 1.8	1.1 ± 2.0	496.0	0.653
YMRS : irritable mania, mean ± SD	0 (0.0)	1.6 ± 1.4	1.2 ± 1.6	422.5	0.189
YMRS : core feature of mania, mean ± SD	0 (0.0)	0.4 ± 0.5	0.3 ± 0.6	452.5	0.285

Significant findings at *p*-value<0.050 are in **bold**

[a] Significant differences between the CHR-T and CHR-NT were examined using the chi-square test and Fisher's exact test for categorical variables, and Mann-Whitney's U test for continuous variables.

BRIAN : total, mean ± SD	1 (1.1)	49.4 ± 12.3	52.9 ± 11.6	443.0	0.361
BRIAN : sleep, mean ± SD	1 (1.1)	17.6 ± 8.2	17.8 ± 12.0	458.5	0.458
BRIAN : activity, mean ± SD	1 (1.1)	11.1 ± 4.7	12.1 ± 3.8	443.0	0.359
BRIAN : social, mean ± SD	1 (1.1)	8.8 ± 2.4	10.0 ± 3.2	395.5	0.145
BRIAN : eating, mean ± SD	1 (1.1)	9.1 ± 2.9	9.8 ± 3.1	447.0	0.382
BRIAN : chronotype, mean ± SD	1 (1.1)	6.5 ± 1.5	6.8 ± 1.7	458.0	0.447
CGI-S, mean ± SD	0 (0.0)	3.5 ± 0.7	3.5 ± 0.8	523.5	0.957
GAF : present, mean ± SD	0 (0.0)	57.7 ± 11.3	57.5 ± 9.1	504.0	0.760
GAF : past one year, mean ± SD	0 (0.0)	70.4 ± 12.4	73.0 ± 12.8	483.5	0.595
GAF : before symptom onset, mean ± SD	0 (0.0)	82.3 ± 5.6	80.1 ± 11.3	512.5	0.832
GF social : present, mean ± SD	0 (0.0)	6.1 ± 1.2	6.0 ± 1.4	493.0	0.659
GF social : lowest, mean ± SD	0 (0.0)	5.7 ± 1.1	5.7 ± 1.4	515.0	0.850
GF social : highest, mean ± SD	0 (0.0)	7.4 ± 1.1	7.2 ± 1.4	488.5	0.626
GF role : present, mean ± SD	0 (0.0)	6.5 ± 1.4	6.2 ± 1.3	447.0	0.333
GF role : lowest, mean ± SD	0 (0.0)	5.9 ± 1.6	5.8 ± 1.2	495.0	0.675
GF role : highest, mean ± SD	0 (0.0)	7.3 ± 1.4	7.5 ± 1.3	512.0	0.819
QIDS total, mean ± SD	3 (3.3)	12.1 ± 7.5	13.4 ± 6.4	473.5	0.670
BIS : total, mean ± SD	3 (3.3)	67.7 ± 10.5	71.3 ± 10.7	399.5	0.201
BIS : attentional, mean ± SD	3 (3.3)	18.0 ± 6.3	19.2 ± 4.8	482.0	0.742
BIS : motor, mean ± SD	3 (3.3)	30.8 ± 9.2	31.3 ± 6.3	445.0	0.451
BIS : non-planning, mean ± SD	3 (3.3)	18.9 ± 6.2	20.9 ± 5.3	445.5	0.454
PQ-B total, mean ± SD	5 (5.6)	21.4 ± 19.0	26.9 ± 18.4	382.5	0.301
PQ-B threshold, mean ± SD	5 (5.6)	0.9 ± 0.3	0.8 ± 0.4	432.5	0.682
WHOQOL-BREF : total, mean ± SD	5 (5.6)	68.9 ± 14.8	67.2 ± 14.5	443.0	0.765
WHOQOL-BREF : overall, mean ± SD	5 (5.6)	5.2 ± 1.3	5.0 ± 1.3	429.5	0.630
WHOQOL-BREF : physical, mean ± SD	5 (5.6)	17.9 ± 5.6	18.0 ± 4.8	463.5	0.959
WHOQOL-BREF : psychological, mean ± SD	5 (5.6)	12.9 ± 4.0	13.1 ± 4.3	455.0	0.877
WHOQOL-BREF : social relationship, mean ± SD	5 (5.6)	9.2 ± 2.8	8.0 ± 2.5	369.0	0.227
WHOQOL-BREF : environment, mean ± SD	5 (5.6)	23.7 ± 5.4	23.1 ± 5.3	437.5	0.714
STAI-T, mean ± SD	5 (5.6)	60.2 ± 9.2	60.6 ± 10.8	441.0	0.747
STAI-S, mean ± SD	5 (5.6)	55.8 ± 7.3	58.6 ± 11.9	385.5	0.319
CD-RISC : total, mean ± SD	5 (5.6)	47.7 ± 14.7	39.4 ± 17.8	330.0	0.093
CTQ : total, mean ± SD	5 (5.6)	66.5 ± 13.6	64.4 ± 12.6	420.0	0.614
CTQ : physical abuse, mean ± SD	5 (5.6)	9.9 ± 5.8	9.4 ± 4.3	448.0	0.870
CTQ : sexual abuse, mean ± SD	5 (5.6)	7.0 ± 2.8	5.7 ± 1.8	344.0	0.063
CTQ : emotional abuse, mean ± SD	5 (5.6)	12.1 ± 5.9	11.8 ± 5.4	452.0	0.910
CTQ : emotional neglect, mean ± SD	5 (5.6)	19.4 ± 5.6	20.1 ± 5.6	440.0	0.795
CTQ : physical neglect, mean ± SD	5 (5.6)	10.5 ± 3.3	10.3 ± 3.1	442.5	0.817
CSM total, mean ± SD	5 (5.6)	28.8 ± 5.6	27.5 ± 7.3	410.5	0.534
SPAQ : GSS total, mean ± SD	6 (6.7)	11.8 ± 6.2	11.0 ± 6.0	424.5	0.648
SPAQ : sleep, mean ± SD	6 (6.7)	2.1 ± 1.1	1.9 ± 1.2	410.5	0.511
SPAQ : social, mean ± SD	6 (6.7)	1.8 ± 1.1	1.8 ± 1.2	430.0	0.672
SPAQ : general mood, mean ± SD	6 (6.7)	2.0 ± 1.2	2.0 ± 1.3	459.0	0.985
SPAQ : weight, mean ± SD	6 (6.7)	1.9 ± 1.3	1.6 ± 1.1	401.5	0.409
SPAQ : energy, mean ± SD	6 (6.7)	2.2 ± 1.4	2.0 ± 1.3	430.5	0.684
SPAQ : appetite, mean ± SD	6 (6.7)	1.8 ± 0.8	1.6 ± 1.1	375.0	0.214
SPAQ	6 (6.7)			0.2	0.768
No, n (%)		6 (46.2)	37 (52.1)		
SAD+S-SAD, n (%)		7 (53.8)	34 (47.9)		

Table3-6[a]. Comparison of baseline proteomic data between CHR-T vs. CHR-NT

Uniprot No.	Gene names	Peptide	Mean difference (CHR-T vs. CHR-NT)	Mann-Whiteny's U	*p*-value	*q*-value
Over-expressed in CHR-T (CHR-T > CHR-NT)						
P05154	IPSP	FSIEGSYQLEK	0.15	311.0	**0.013**	0.345
P05154	IPSP	TLYLADTFPTNFR	0.38	313.5	**0.014**	0.345
Over-expressed in CHR-NT (CHR-T < CHR-NT)						
Q16610	ECM1	NVALVSGDTENAK	-0.65	279.5	**0.004**	0.345
P00488	F13A	GTYIPVPIVSELQSGK	-0.59	303.5	**0.010**	0.345
P04004	VTNC	FEDGVLDPDYPR	-0.55	322.0	**0.018**	0.345
P19652	A1AG2	TEDTIFLR	-1.05	322.5	**0.019**	0.345
P02790	HEMO	GGYTLVSGYPK	-0.53	327.0	**0.022**	0.345
P20851	C4BPB	ALLAFQESK	-0.54	328.5	**0.022**	0.345
P00488	F13A	VEYVIGR	-0.64	329.5	**0.023**	0.345
P05452	TETN	LDTLAQEVALLK	-0.62	331.5	**0.025**	0.345
P62937	PPIA	FEDENFILK	1.05	335.0	**0.027**	0.345
P02647	APOA1	QGLLPVLESFK	-0.41	337.0	**0.029**	0.345
P10643	CO7	AASGTQNNVLR	-0.60	338.5	**0.030**	0.345
P04114	APOB	TEVIPPLIENR	-0.42	341.0	**0.033**	0.345
P04217	A1BG	HQFLLTGDTQGR	-0.46	345.0	**0.037**	0.345
P06276	CHLE	IFFPGVSEFGK	-1.02	346.5	**0.038**	0.345
P20742	PZP	SSGSLLNNAIK	-0.83	353.0	**0.046**	0.345
Q9NPH3	IL1AP	VAFPLEVVQK	-0.49	355.0	**0.048**	0.345

q-value means adjusted *p*-value for multiple comparison by using Benjamini-Hochberg FDR. Proteins significant findings at *p*-value<0.050 and *q*-value<0.050 are in **bold**

[a] Table 3-6 Summarize statistically significant peptides among 158 peptides when analyzed by Mann-Whitney's U test

Table 3-7A. Unique models of Model 1a based on the combinations of features selected

Model 1a			
Unique Model	Number of features	Frequency	Model probability
1	2	5	0.367
2	3	3	0.272
3	1	7	0.258
4	10	4	0.025
5	6	2	0.019
6	0	1	0.017
7	11	2	0.017
8	7	2	0.017
9	9	2	0.005
10	13	1	0.001
11	15	2	0.001
12	14	1	0.001
13	20	3	2.27.E-06
14	19	1	1.10.E-06
15	21	2	9.14.E-07
16	24	3	1.68.E-08
17	23	1	1.58.E-08
18	28	24	9.85.E-09
19	25	2	3.73.E-09
20	28	1	2.16.E-09
21	29	3	3.30.E-10
22	26	1	3.25.E-10
23	30	17	2.72.E-10
24	27	3	1.95.E-10
25	29	1	1.48.E-10
26	31	3	3.64.E-12
27	32	3	1.64.E-13
Total	100		1.0

Table 3-7B. Unique models of Model 1b based on the combinations of features selected

Model 1b			
Unique Model	Number of features	Frequency	Model probability
1	1	2	0.398
2	0	1	0.390
3	2	1	0.096
4	9	3	0.040
5	6	3	0.036
6	4	1	0.015
7	8	2	0.012
8	5	1	0.008
9	11	1	0.003
10	17	3	0.001
11	18	3	0.001
12	16	5	0.001
13	13	1	3.78.E-04
14	14	1	1.79.E-04
15	19	1	9.54.E-05
16	20	2	6.11.E-05
17	22	1	1.11.E-06
18	23	1	2.15.E-07
19	24	1	3.83.E-08
20	25	1	6.65.E-09
21	28	10	2.00.E-09
22	30	23	1.50.E-09
23	26	1	1.03.E-09
24	31	21	4.02.E-10
25	27	1	1.40.E-10
26	29	2	1.19.E-10
27	32	7	9.44.E-12
Total		100	1.0

Table 3-7C. Unique models of Model 1c based on the combinations of features selected

Model 1c			
Unique Model	Number of features	Frequency	Model probability
1	2	5	0.513
2	1	7	0.361
3	4	2	0.081
4	0	1	0.024
5	8	3	0.011
6	6	1	0.009
7	11	2	0.001
8	13	2	1.58E-04
9	15	1	1.01E-05
10	18	1	1.35E-07
11	24	3	9.26E-09
12	22	1	1.59E-10
13	25	5	1.51E-10
14	24	1	1.58E-11
15	27	1	5.95E-12
16	24	1	5.85E-12
17	28	3	3.95E-12
18	31	4	2.05E-13
19	30	3	5.34E-14
20	31	4	1.35E-14
21	32	3	1.20E-14
22	32	6	4.00E-15
23	34	12	4.11E-16
24	33	1	1.92E-16
25	35	7	2.99E-17
26	35	13	2.91E-17
27	34	3	2.27E-17
28	35	3	8.07E-19
29	36	1	2.69E-20
Total		100	1.0

Table3-8. Features included in prediction models (Model1)

Features	Selection fraction	Average coefficient
Model 1a (Clinical variables)		
Initial classification (CHR-BD)	0.92	1.966
Duration of follow up	0.87	0.536
Psychiatric family history	0.84	-0.285
History of treatment (Medication-past)	0.82	1.121
Use of Mood stabilizers	0.84	1.291
BPSS current : T1 threshold	0.99	1.463
HDRS : anhedonia/retardation	0.80	1.075
WHOQOL-BREF : social relationship	0.84	0.150
CD-RISC : total	0.80	0.287
Model 1b (Proteomic variables)		
FETUB_AIFYMNNPSR	0.97	-0.712
CO9_LSPIYNLVPVK	0.94	0.934
CMGA_SGELEQEEER	0.86	-0.751
A1AT_SVLGQLGITK	0.89	1.147
ITIH4_FAHTVVTSR	0.85	0.207
MIS_ELSVDLR	0.91	0.558
MIF_LLCGLLAER	0.86	0.201
SAMP_VFVFPR	0.84	0.647
VWF_ILAGPAGDSNVVK	0.83	-0.575
HGFA_VANYVDWINDR	0.95	-0.512
A1BG_ATWSGAVLAGR	0.85	0.164
F13A_GTYIPVPIVSELQSGK	0.96	-1.155
MASP2_WPEPVFGR	0.83	0.560
MDHC_GEFVTTVQQR	0.91	0.413
PPIA_FEDENFILK	0.96	0.368
Model 1c (Mixed)		
Initial classification (CHR-BP)	0.92	1.021
History of treatment (Medication-past)	0.81	1.414
Use of Mood stabilizers	0.84	0.883
Duration of follow up	0.87	0.440
BPSS current : T1 threshold	0.99	1.190
HGFA_VANYVDWINDR	0.85	-0.397
ECM1_NVALVSGDTENAK	0.85	-0.344

Table3-9. Comparison of baseline demographical characteristics between Recovered vs. Not-recovered group

	Missing, n (%)	Recovered (n = 39)	Not-recovered (n = 44)	Statistics[a]	*p*-value
Age, years, mean ± SD	0 (0.0)	21.8 ± 3.7	21.5 ± 4.2	0.4	0.671
Gender, male, n (%)	**0 (0.0)**	**10 (25.6)**	**25 (56.8)**	**8.2**	**0.007**
Body mass Index, kg/m^2, mean ± SD	0 (0.0)	22.2 ± 3.5	22.4 ± 3.4	-0.2	0.854
Education, years, mean ± SD	0 (0.0)	13.6 ± 2.1	13.3 ± 1.9	0.7	0.503
Social economic status	0 (0.0)			2.5	0.147
Unstable, n (%)		8 (20.5)	16 (36.4)		
Stable, n (%)		31 (79.5)	28 (63.6)		
Initial classification (Study 1)	0 (0.0)				
CHR-PRS, n (%)		11 (28.2)	13 (29.5)	0.0	>0.999
CHR-BD, n (%)		21 (53.8)	18 (40.9)	1.4	0.275
CHR-MDD, n (%)		1 (2.6)	1 (2.3)	FE	>0.999
CHR-PL, n (%)		0 (0.0)	4 (9.1)	FE	0.119
CHR-Extended, n (%)		6 (15.4)	8 (18.2)	0.1	0.777
LPA classification (Study 2)	0 (0.0)			2.2	0.344
LS, n (%)		13 (33.3)	11 (25.0)		
DPD, n (%)		14 (35.9)	23 (52.3)		
MD, n (%)		12 (30.8)	10 (22.7)		
Duration of follow up, month, mean ± SD	**0 (0.0)**	**22.5 ± 8.1**	**16.6 ± 9.4**	**3.1**	**0.003**
Psychiatric family history, n (%)	0 (0.0)	14 (35.9)	23 (52.3)	2.2	0.185
History of treatment	0 (0.0)			4.4	0.241
No, n (%)		4 (10.3)	9 (20.5)		
Medication-present, n (%)		24 (61.5)	28 (63.6)		
Medication-past, n (%)		8 (20.5)	3 (6.8)		
Psychosocial intervention, n (%)		3 (7.7)	4 (9.1)		
Use of psychotropic medication, n (%)	0 (0.0)	24 (61.5)	28 (63.6)	0.0	>0.999
Use of antipsychotics, n (%)	0 (0.0)	17 (43.6)	18 (40.9)	0.1	0.827
Use of Mood stabilizers, n (%)	0 (0.0)	11 (28.2)	20 (45.5)	2.6	0.118
Use of Anti-depressant, n (%)	0 (0.0)	7 (17.9)	8 (18.2)	0.0	>0.999
Use of Benzodiazepines, n (%)	0 (0.0)	12 (30.8)	16 (36.4)	0.3	0.647

Significant findings at *p*-value<0.050 are in **bold**

[a] Significant differences between the Recovered and Not recovered were examined using the chi-square test and Fisher's exact test for categorical variables, and Student's independent test for continuous variables.

Table3-10. Comparison of baseline symptom characteristics between Recovered vs. Not-recovered group

	Missing, n (%)	Recovered (n = 39)	Not-recovered (n = 44)	Statistics[a]	*p*-value	*q*-value
SIPS : positive	**0 (0.0)**	**4.3 ± 4.9**	**6.7 ± 4.5**	**-2.2**	**0.029**	0.258
SIPS : negative	0 (0.0)	8.2 ± 6.9	10.2 ± 6.1	-1.4	0.168	0.419
SIPS : disorganized	0 (0.0)	2.6 ± 3.0	2.7 ± 2.2	-0.1	0.915	0.947
SIPS : general	0 (0.0)	6.3 ± 3.5	7.5 ± 4.0	-1.4	0.157	0.419
SIPS : total	0 (0.0)	21.6 ± 13.5	27.0 ± 13.9	-1.8	0.076	0.329
BPSS current : total	0 (0.0)	18.2 ± 17.5	19.0 ± 16.5	-0.2	0.817	0.920
BPSS current : severity	0 (0.0)	8.7 ± 7.6	9.1 ± 8.2	-0.3	0.788	0.911
BPSS current : frequency	0 (0.0)	9.5 ± 10.7	9.9 ± 9.2	-0.2	0.857	0.930
BPSS current : T1 threshold	0 (0.0)	0.1 ± 0.3	0.2 ± 0.5	-1.1	0.259	0.549
BPSS current : T2 threshold	0 (0.0)	0.8 ± 1.1	0.8 ± 1.1	-0.2	0.843	0.926
BPSS current : T3 threshold	0 (0.0)	0.0 ± 0.2	0.0 ± 0.2	0.1	0.932	0.953
BPSS current : T4 threshold	0 (0.0)	0.5 ± 1.0	0.6 ± 1.0	-0.3	0.732	0.890
BPSS current : threshold total	0 (0.0)	1.4 ± 1.6	1.6 ± 1.8	-0.6	0.549	0.810
BPSS current : positive mania	0 (0.0)	2.4 ± 3.6	2.8 ± 4.4	-0.5	0.604	0.810
BPSS current : negative mania	0 (0.0)	6.3 ± 4.8	6.3 ± 5.1	0.0	0.992	0.992
BPSS past : total	0 (0.0)	28.4 ± 27.6	32.3 ± 28.9	-0.6	0.533	0.810
BPSS past : severity	0 (0.0)	13.2 ± 11.8	15.4 ± 14.8	-0.7	0.453	0.741
BPSS past : frequency	0 (0.0)	15.3 ± 16.3	16.9 ± 15.0	-0.5	0.628	0.810
BPSS past : T1 threshold	0 (0.0)	0.2 ± 0.4	0.3 ± 0.8	-1.0	0.311	0.615
BPSS past : T2 threshold	0 (0.0)	1.0 ± 1.3	1.1 ± 1.3	-0.2	0.880	0.936
BPSS past : T3 threshold	0 (0.0)	0.1 ± 0.3	0.2 ± 0.6	-1.5	0.142	0.419
BPSS past : T4 threshold	0 (0.0)	1.3 ± 1.9	1.7 ± 2.7	-0.8	0.435	0.741
BPSS past : threshold total	0 (0.0)	2.6 ± 2.8	3.3 ± 3.4	-1.1	0.287	0.581
BPSS past : positive mania	0 (0.0)	5.5 ± 6.1	6.6 ± 7.1	-0.7	0.458	0.741
BPSS past : negative mania	0 (0.0)	7.7 ± 6.4	8.8 ± 8.5	-0.7	0.495	0.787
PANSS : total	0 (0.0)	50.5 ± 14.0	53.0 ± 11.4	-0.9	0.378	0.676
PANSS : positive	0 (0.0)	10.1 ± 2.9	10.8 ± 2.4	-1.2	0.236	0.525
PANSS : negative	0 (0.0)	11.4 ± 5.6	12.4 ± 4.2	-0.9	0.380	0.676
PANSS : general	0 (0.0)	29.0 ± 7.8	29.8 ± 7.3	-0.5	0.614	0.810
HDRS : total	0 (0.0)	10.6 ± 5.6	13.0 ± 7.1	-1.7	0.085	0.329
HDRS : anhedonia/retardation	0 (0.0)	3.7 ± 2.3	4.6 ± 2.2	-1.8	0.070	0.329
HDRS : guilt/agitation	0 (0.0)	3.3 ± 2.2	3.6 ± 2.3	-0.5	0.599	0.810
HDRS : somatic	0 (0.0)	1.3 ± 1.6	1.7 ± 1.8	-1.1	0.258	0.549
HDRS : insomnia	0 (0.0)	1.4 ± 1.6	2.1 ± 2.0	-1.6	0.117	0.381
HDRS : appetite	0 (0.0)	0.4 ± 0.8	0.6 ± 0.9	-1.2	0.232	0.525
HAM-A total	**0 (0.0)**	**8.9 ± 4.9**	**11.8 ± 6.5**	**-2.3**	**0.026**	0.258
YMRS : total	0 (0.0)	2.5 ± 3.2	2.9 ± 3.4	-0.5	0.608	0.810
YMRS : psychotic mania	0 (0.0)	1.1 ± 1.9	1.3 ± 2.1	-0.4	0.655	0.833
YMRS : irritable mania	0 (0.0)	1.2 ± 1.8	1.4 ± 1.5	-0.4	0.719	0.889
YMRS : core feature of mania	0 (0.0)	0.3 ± 0.6	0.3 ± 0.6	-0.3	0.759	0.896

Significant findings at *p*-value<0.050 are in **bold**

[a] Significant differences between the Recovered and Not recovered were examined using the chi-square test and Fisher's exact test for categorical variables, and Student's independent test for continuous variables.

BRIAN : total	**1 (1.2)**	**48.5 ± 10.9**	**55.0 ± 12.1**	**-2.5**	**0.013**	0.193
BRIAN : sleep	1 (1.2)	17.8 ± 13.2	18.2 ± 10.5	-0.1	0.885	0.936
BRIAN : activity	1 (1.2)	11.0 ± 4.0	12.6 ± 4.0	-1.8	0.073	0.329
BRIAN : social	1 (1.2)	9.2 ± 2.9	10.0 ± 3.0	-1.4	0.173	0.419
BRIAN : eating	**1 (1.2)**	**8.9 ± 2.5**	**10.3 ± 3.3**	**-2.1**	**0.039**	0.289
BRIAN : chronotype	1 (1.2)	6.4 ± 1.5	6.9 ± 1.7	-1.4	0.160	0.419
CGI-S : baseline	0 (0.0)	3.4 ± 0.7	3.7 ± 0.8	-1.6	0.103	0.381
CGI-S : last-follow up	**0 (0.0)**	**1.6 ± 0.5**	**3.6 ± 0.7**	**-14.9**	**<0.001**	**<0.001**
Transition to major psychiatric disorder,	0 (0.0)	5 (12.8)	8 (18.2)	0.5	0.558	0.810
GAF : present	**0 (0.0)**	**60.3 ± 9.2**	**55.5 ± 9.3**	**2.4**	**0.020**	0.254
GAF : past one year	0 (0.0)	75.6 ± 11.9	70.8 ± 12.9	1.8	0.084	0.329
GAF : before symptom onset	0 (0.0)	81.9 ± 10.0	79.5 ± 11.8	1.0	0.335	0.648
GF social : present	0 (0.0)	6.2 ± 1.3	6.0 ± 1.3	0.8	0.412	0.719
GF social : lowest	0 (0.0)	5.8 ± 1.4	5.7 ± 1.3	0.6	0.575	0.810
GF social : highest	0 (0.0)	7.5 ± 1.3	7.2 ± 1.3	1.1	0.287	0.581
GF role : present	0 (0.0)	6.6 ± 1.2	6.1 ± 1.2	2.0	0.051	0.320
GF role : lowest	0 (0.0)	6.1 ± 1.2	5.8 ± 1.3	1.3	0.206	0.482
GF role : highest	0 (0.0)	7.7 ± 1.1	7.4 ± 1.2	1.4	0.174	0.419
QIDS total	3 (3.6)	11.9 ± 6.2	14.6 ± 6.3	-1.9	0.059	0.328
BIS : total	3 (3.6)	68.8 ± 11.2	72.7 ± 10.1	-1.6	0.110	0.381
BIS : attentional	3 (3.6)	18.6 ± 4.3	20.1 ± 4.1	-1.6	0.124	0.381
BIS : motor	3 (3.6)	30.2 ± 6.2	31.0 ± 5.6	-0.6	0.553	0.810
BIS : non-planning	3 (3.6)	20.1 ± 4.5	21.7 ± 4.5	-1.6	0.112	0.381
PQ-B total	3 (3.6)	21.5 ± 18.8	29.7 ± 17.9	-2.0	0.050	0.320
PQ-B threshold,	**3 (3.6)**	**27 (73.0)**	**41 (95.3)**	**7.8**	**0.010**	0.178
WHOQOL-BREF : total	**3 (3.6)**	**72.1 ± 14.9**	**63.8 ± 12.9**	**2.7**	**0.009**	0.178
WHOQOL-BREF : overall	3 (3.6)	5.3 ± 1.2	4.8 ± 1.3	1.8	0.070	0.329
WHOQOL-BREF : physical	**3 (3.6)**	**19.2 ± 5.6**	**16.8 ± 4.0**	**2.2**	**0.028**	0.258
WHOQOL-BREF : psychological	**3 (3.6)**	**14.4 ± 4.0**	**11.9 ± 4.0**	**2.8**	**0.007**	0.178
WHOQOL-BREF : social relationship	3 (3.6)	8.7 ± 2.6	7.9 ± 2.6	1.4	0.163	0.419
WHOQOL-BREF : environment	3 (3.6)	24.5 ± 5.5	22.4 ± 4.8	1.8	0.078	0.329
STAI-T	3 (3.6)	58.4 ± 11.0	62.1 ± 10.1	-1.6	0.123	0.381
STAI-S	3 (3.6)	55.2 ± 12.3	60.2 ± 10.2	-2.0	0.054	0.320
CD-RISC : total	**3 (3.6)**	**45.1 ± 17.8**	**36.8 ± 16.8**	**2.1**	**0.037**	0.289
CTQ : total	3 (3.6)	64.1 ± 14.2	64.7 ± 11.4	-0.2	0.827	0.920
CTQ : physical abuse	3 (3.6)	9.7 ± 4.5	9.2 ± 4.8	0.5	0.621	0.810
CTQ : sexual abuse	3 (3.6)	6.1 ± 2.3	5.8 ± 1.9	0.6	0.571	0.810
CTQ : emotional abuse	3 (3.6)	11.5 ± 6.2	11.8 ± 4.8	-0.3	0.765	0.896
CTQ : emotional neglect	3 (3.6)	19.1 ± 5.9	20.8 ± 5.0	-1.4	0.162	0.419
CTQ : physical neglect	3 (3.6)	10.1 ± 2.9	10.3 ± 3.3	-0.4	0.673	0.844
CSM total	**3 (3.6)**	**30.4 ± 7.2**	**25.5 ± 6.2**	**3.3**	**0.002**	0.089
SPAQ : GSS total	3 (3.6)	10.8 ± 5.7	11.3 ± 6.2	-0.3	0.740	0.890
SPAQ : sleep	3 (3.6)	1.9 ± 1.1	2.0 ± 1.2	-0.1	0.894	0.936
SPAQ : social	3 (3.6)	1.6 ± 1.1	1.9 ± 1.2	-0.9	0.367	0.676
SPAQ : general mood	3 (3.6)	1.9 ± 1.2	2.1 ± 1.4	-0.8	0.446	0.741
SPAQ : weight	3 (3.6)	1.6 ± 1.0	1.7 ± 1.2	0.0	0.992	0.992

SPAQ : energy	3 (3.6)	2.0 ± 1.3	2.1 ± 1.3	-0.2	0.812	0.920
SPAQ : appetite	3 (3.6)	1.7 ± 1.0	1.6 ± 1.1	0.5	0.612	0.810
SPAQ	3 (3.6)			0.8	0.380	0.676
No,		21 (56.8)	20 (46.5)			
SAD+S-SAD,		16 (43.2)	23 (53.5)			

Table3-11[a]. Comparison of baseline proteomic data between Recovered vs. Not-recovered group

Uniprot No.	Gene names	Peptide	Mean difference (Recovered vs. Not recovered)	Student's t	*p*-value	*q*-value
Over-expressed in Not-recovered (Recovered > Not-recovered)						
P00748	FA12	TEQAAVAR	-0.44	-3.3	**0.001**	0.079
P00748	FA12	VVGGLVALR	-0.48	-3.5	**0.001**	0.079
P61769	B2MG	VNHVTLSQPK	-0.22	-2.4	**0.017**	0.632
P04003	C4BPA	QSSSYSFFK	-0.41	-2.3	**0.023**	0.632
P02743	SAMP	VGEYSLYIGR	-0.40	-2.3	**0.024**	0.632
P02743	SAMP	QGYFVEAQPK	-0.33	-2.3	**0.024**	0.632
Q13103	SPP24	VSAQQVQGVHAR	-0.40	-2.1	**0.040**	0.795
P06702	S10A9	LTWASHEK	-0.42	-2.1	**0.041**	0.795

q-value means adjusted *p*-value for multiple comparison by using Benjamini-Hochberg FDR. Proteins significant findings at *p*-value<0.050 and *q*-value<0.050 are in **bold**

[a] Table 3-11 Summarize statistically significant peptides among 158 peptides when analyzed by Student's independent t-test

Table 3-12A. Unique models of Model 2a based on the combinations of features selected

Model 2a			
Unique Model	Number of features	Frequency	Model probability
1	0	1	0.443
2	3	3	0.165
3	10	3	0.111
4	9	3	0.098
5	8	5	0.073
6	5	2	0.042
7	12	3	0.038
8	11	1	0.020
9	7	1	0.004
10	16	4	0.003
11	14	1	0.002
12	15	1	0.001
13	18	1	7.24E-05
14	20	3	2.59E-05
15	22	1	4.72E-07
16	23	1	1.07E-07
17	24	1	2.22E-08
18	26	1	2.88E-10
19	29	3	8.39E-13
20	31	1	1.50E-15
21	34	1	3.58E-20
22	36	6	3.00E-21
23	35	1	1.23E-21
24	37	1	2.29E-23
25	39	2	2.45E-27
26	43	11	3.62E-31
27	42	1	3.61E-31
28	41	1	1.07E-32
29	43	3	6.57E-34
30	42	2	4.77E-35
31	44	5	3.93E-37
32	44	6	3.95E-38
33	44	7	3.68E-39
34	45	2	1.18E-41
35	46	2	1.18E-41
36	44	1	6.80E-42
37	45	4	1.07E-42
38	45	4	1.31E-44
Total		100	1.0

Table 3-12B. Unique models of Model 2b based on the combinations of features selected

Model 2b			
Unique Model	Number of features	Frequency	Model probability
1	2	4	0.283
2	1	4	0.275
3	3	3	0.195
4	0	1	0.091
5	5	2	0.050
6	8	3	0.039
7	9	2	0.032
8	6	1	0.019
9	7	1	0.010
10	12	3	0.004
11	13	1	0.001
12	15	2	4.27E-04
13	14	1	3.86E-04
14	16	1	9.83E-05
15	18	2	1.65E-05
16	20	2	1.65E-06
17	21	1	3.29E-07
18	23	1	8.60E-09
19	26	2	3.65E-11
20	27	2	1.26E-11
21	30	1	2.03E-15
22	32	4	1.94E-16
23	31	1	1.61E-16
24	33	1	4.99E-18
25	37	2	5.60E-25
26	38	1	4.19E-27
27	39	1	3.14E-29
28	40	1	1.52E-31
29	41	1	1.64E-33
30	41	1	4.57E-34
31	42	1	7.94E-37
32	43	5	2.67E-38
33	44	1	4.85E-42
34	45	1	9.34E-46
35	46	2	1.87E-49
36	47	2	7.92E-54
37	48	10	7.80E-56
38	48	1	1.01E-57
39	48	1	3.36E-58
40	48	3	1.82E-58
41	49	4	2.27E-62
42	49	1	2.15E-62
43	49	2	4.41E-63
44	49	4	4.05E-63

45	49	4	1.18E-63
46	50	1	7.87E-69
47	50	1	6.83E-69
48	50	3	3.47E-69
Total		100	1.0

Table 3-12C. Unique models of Model 2c based on the combinations of features selected

Model 2c			
Unique Model	Number of features	Frequency	Model probability
1	10	5	0.235
2	1	2	0.219
3	11	2	0.185
4	0	1	0.160
5	2	1	0.076
6	3	1	0.060
7	8	1	0.037
8	5	1	0.014
9	9	2	0.006
10	14	1	0.003
11	7	1	0.003
12	8	1	0.002
13	18	1	8.72E-06
14	21	3	2.96E-06
15	20	1	5.71E-07
16	23	2	2.36E-07
17	24	2	1.42E-07
18	25	1	1.95E-08
19	26	2	1.03E-08
20	28	3	4.05E-10
21	30	1	8.98E-13
22	31	2	1.81E-13
23	34	4	3.32E-17
24	35	2	1.11E-18
25	36	2	3.33E-20
26	37	1	2.64E-22
27	39	2	3.73E-26
28	40	1	2.04E-27
29	40	2	1.42E-27
30	40	2	2.17E-28
31	41	5	3.55E-29
32	40	1	2.10E-29
33	41	2	4.05E-30
34	41	1	6.35E-31
35	41	1	2.99E-31
36	42	1	1.21E-32
37	42	1	3.97E-34
38	43	3	2.75E-35
39	43	1	1.55E-35
40	44	1	3.41E-39
41	45	2	5.64E-42
42	45	3	2.05E-42
43	46	1	1.29E-46
44	46	2	1.02E-46
45	47	5	2.42E-50

46	47	3	1.04E-50
47	47	4	7.96E-51
48	47	1	1.20E-51
49	48	6	2.38E-55
50	48	1	2.48E-56
51	48	2	1.78E-56
Total		100	1.0

Table3-13. Features included in prediction models (Model2)

Features	Selection fraction	Average coefficient
Model 2a (Clinical variables)		
Initial classification (CHR-Pluripotent)	0.81	2.275
Duration of follow up	0.93	-1.174
BPSS current : T3 threshold	0.82	-0.944
HAM-A	0.85	0.771
GAF : present	0.88	-0.312
WHOQOL-BREF : psychological	0.99	-0.387
CSM total	0.99	-1.258
PQ-B threshold	0.99	2.702
Model 2b (Proteomic variables)		
SAMP_QGYFVEAQPK	0.95	1.085
LUM_ILGPLSYSK	0.85	0.375
FA12_TEQAAVAR	0.86	0.728
FA12_VVGGLVALR	0.99	0.424
KLKB1_TGAVSGHSLK	0.88	-0.443
SPP24_VSAQQVQGVHAR	0.91	0.704
CHLE_FWTSFFPK	0.84	1.192
CBG_ITQDAQLK	0.81	-0.370
Model 2c (Mixed)		
History of treatment (Medication-past)	0.82	-0.695
Use of Mood stabilizers	0.80	0.230
Duration of follow up	0.96	-0.979
BPSS current : T1 threshold	0.81	0.386
PQ-B threshold	0.92	0.727
GAF : present	0.89	-0.612
GAF : past one year	0.81	-0.055
WHOQOL-BREF : psychological	0.97	-0.232
CSM total	0.99	-0.964
MIS_ELSVDLR	0.80	0.169
LUM_ILGPLSYSK	0.82	0.439
VWF_ILAGPAGDSNVVK	0.81	-0.673
FA12_TEQAAVAR	0.93	0.769
SPP24_VSAQQVQGVHAR	0.84	0.533

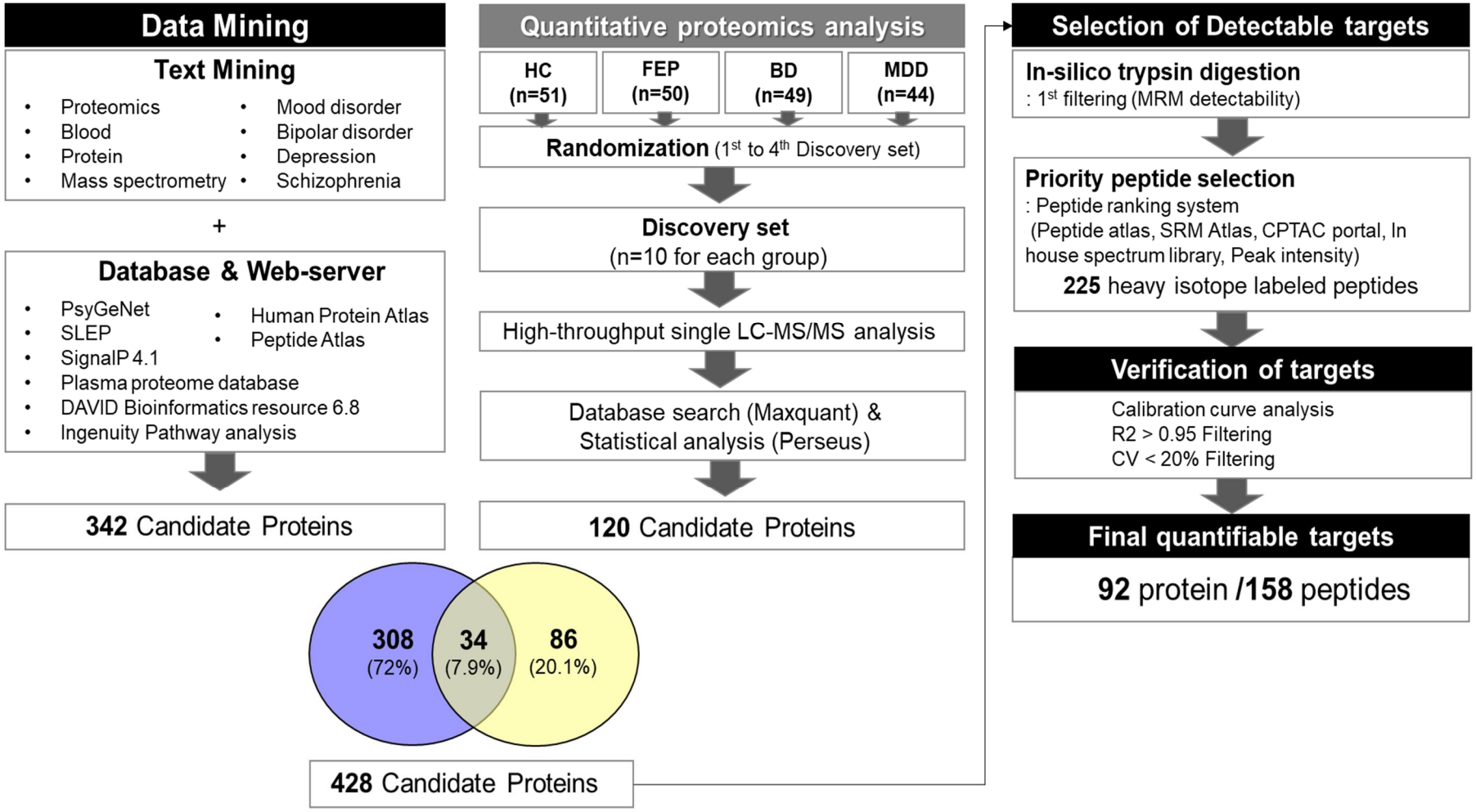


Figure 1-1. Determination of quantifiable targets for CHR of psychiatric disorders

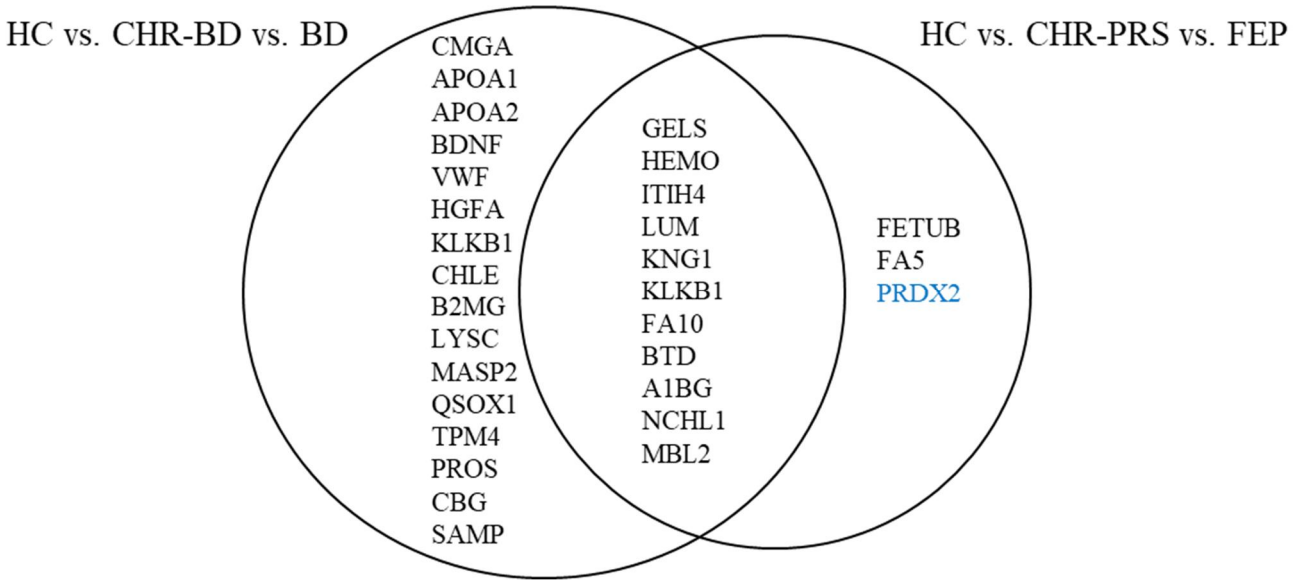


Figure 1-2. Summary of DEPs in CHR-BD[a] and CHR-PRS[b]

---

[a] Results of ANCOVA between HC, CHR-BD, and BD ($q < 0.05$)

[b] Results of ANCOVA between HC, CHR-PRS, and FEP ($q < 0.05$)

PRDX2 in blue color was significant in ANOVA between HC, CHR-BD, and BD ($q < 0.05$)

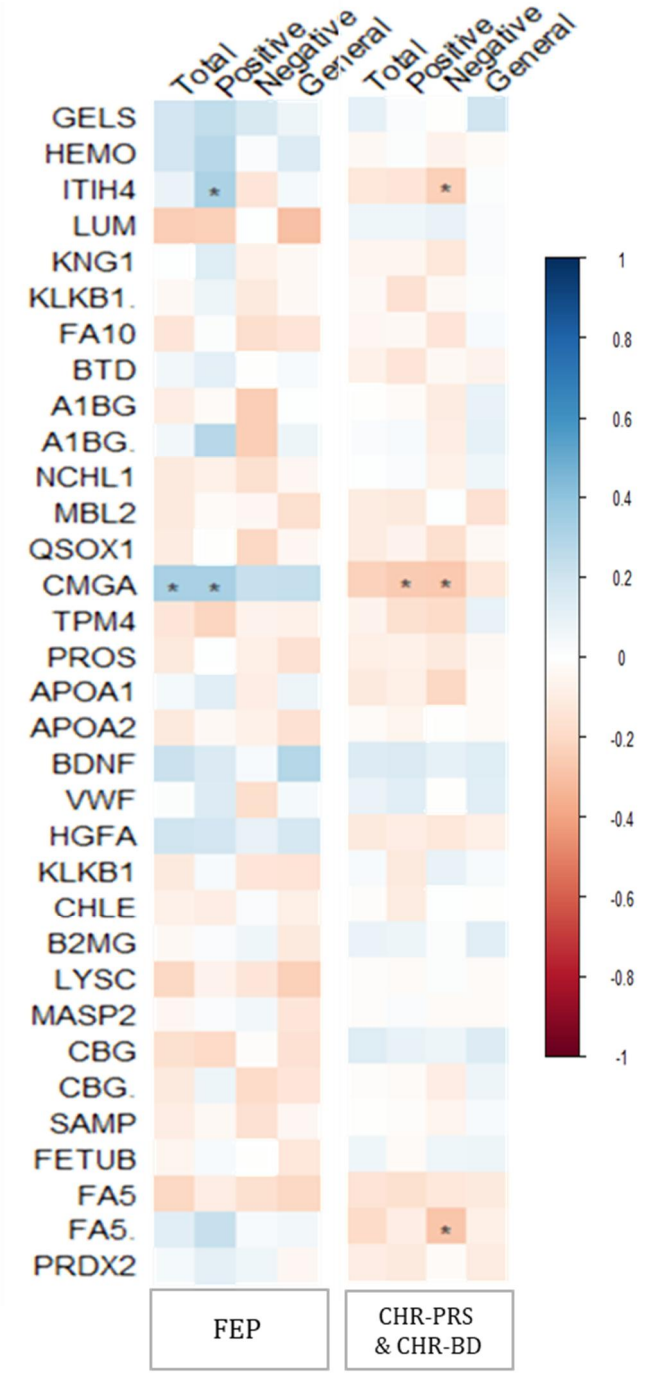


Figure 1-3. Correlation plot[a] showing Pearson's correlation analysis between clinical variable (PANSS) and DEPs

---

* p <0.05, ** p < 0.01, *** p < 0.001

[a] Positive and negative correlation coefficients are colored blue and red, respectively

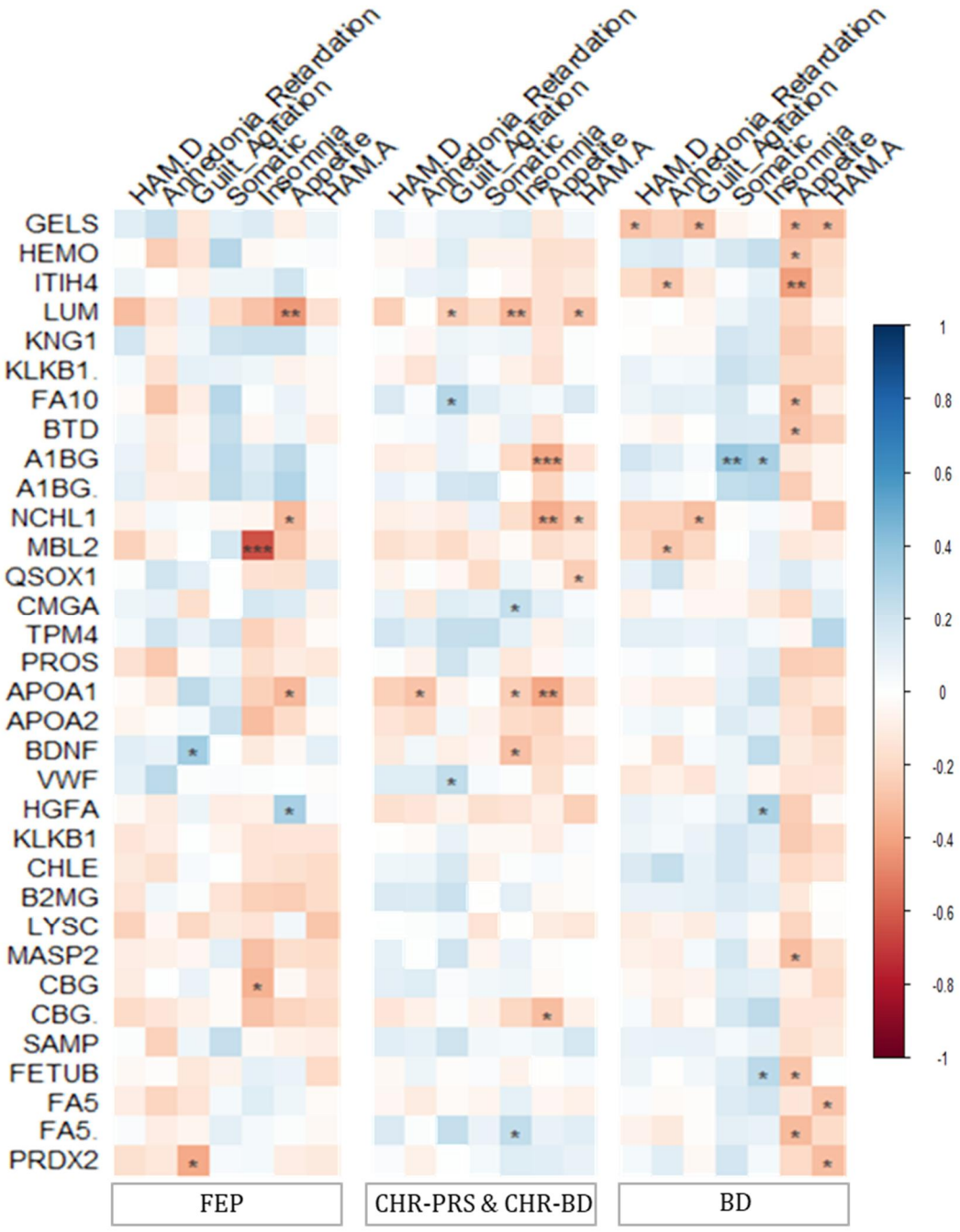


Figure 1-4. Correlation plot[a] showing Pearson's correlation analysis between clinical variable (HAM-D, HAM-A) and DEPs

---

* p <0.05, ** p < 0.01, *** p < 0.001

[a] Positive and negative correlation coefficients are colored blue and red, respectively

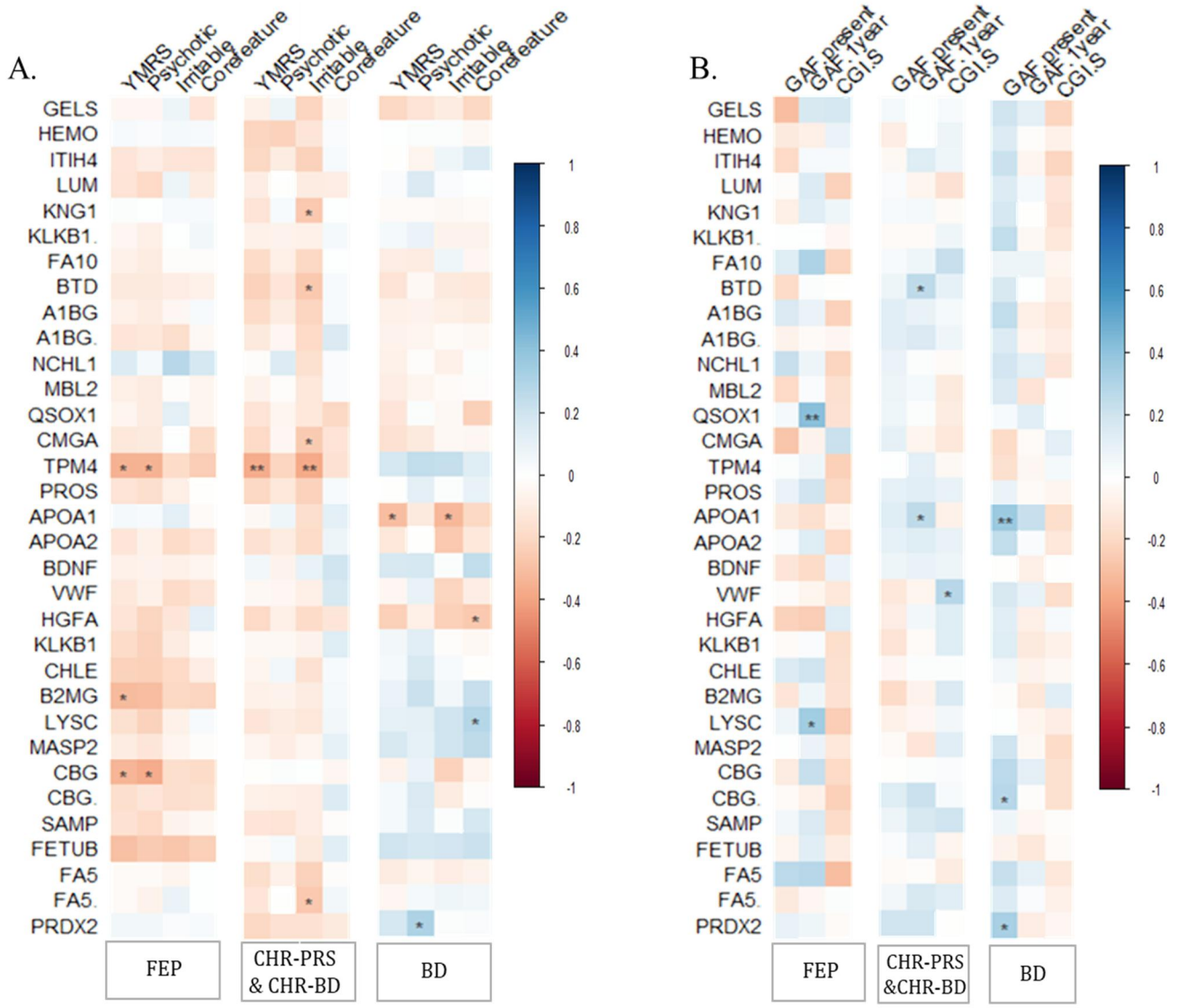


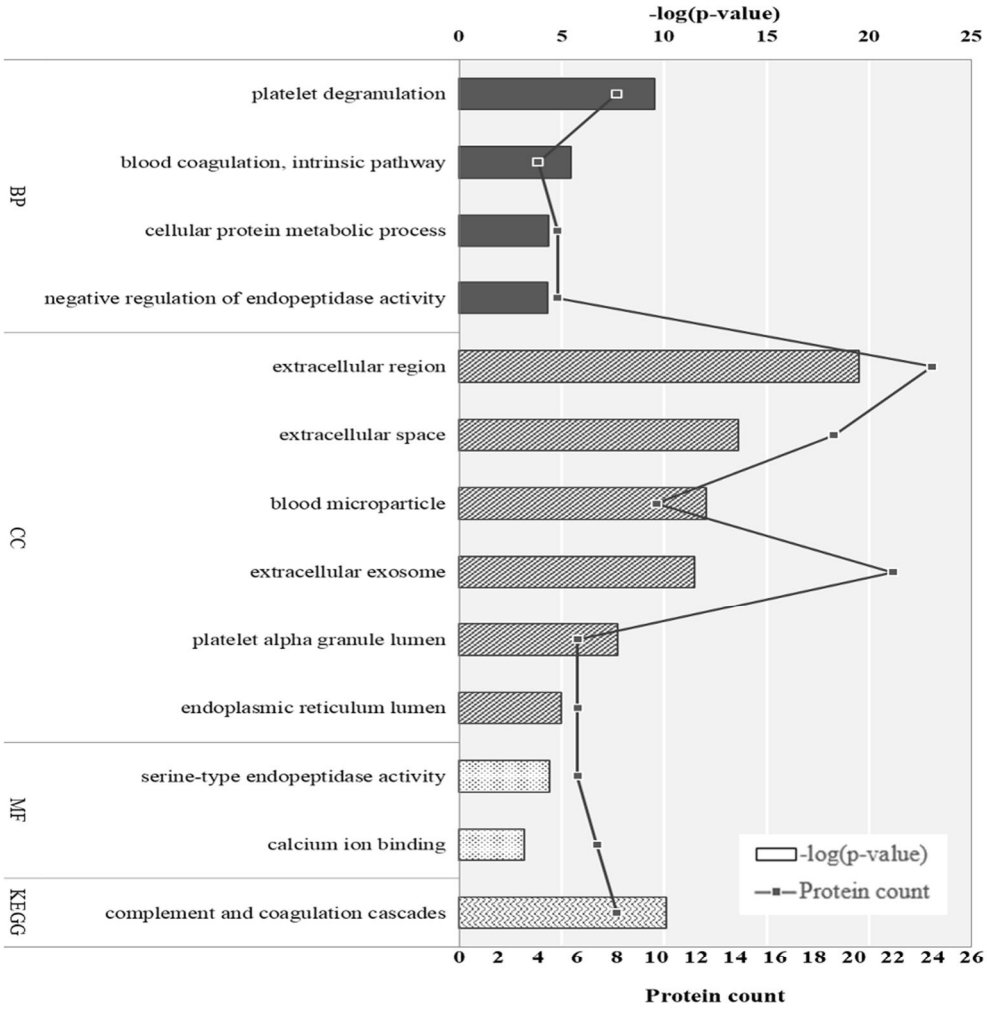
Figure 1-5. Correlation plot[a] showing Pearson's correlation analysis between clinical variable (A-YMRS, B-GAF, CGI) and DEPs

---

* p <0.05, ** p < 0.01, *** p < 0.001

[a] Positive and negative correlation coefficients are colored blue and red, respectively

A. Gene ontology (GO) analysis of the proteins that DEPs between groups (*DAVID Bioinformatics resources 6.8*)



B. Upstream regulator prediction analysis

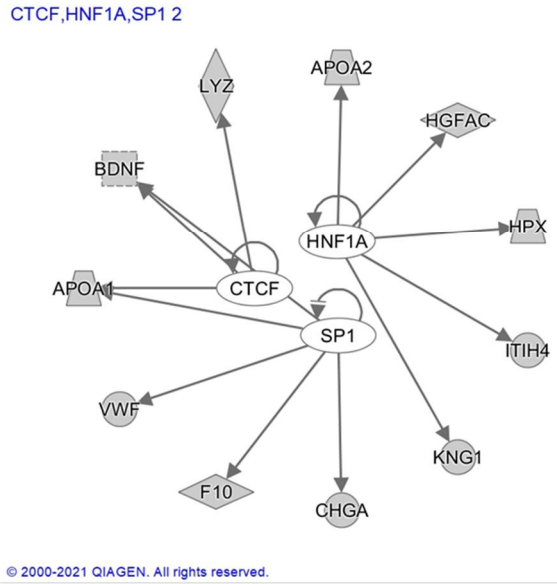


Figure 1-6. Bioinformatics analysis of the proteins that differentially expressed in CHR

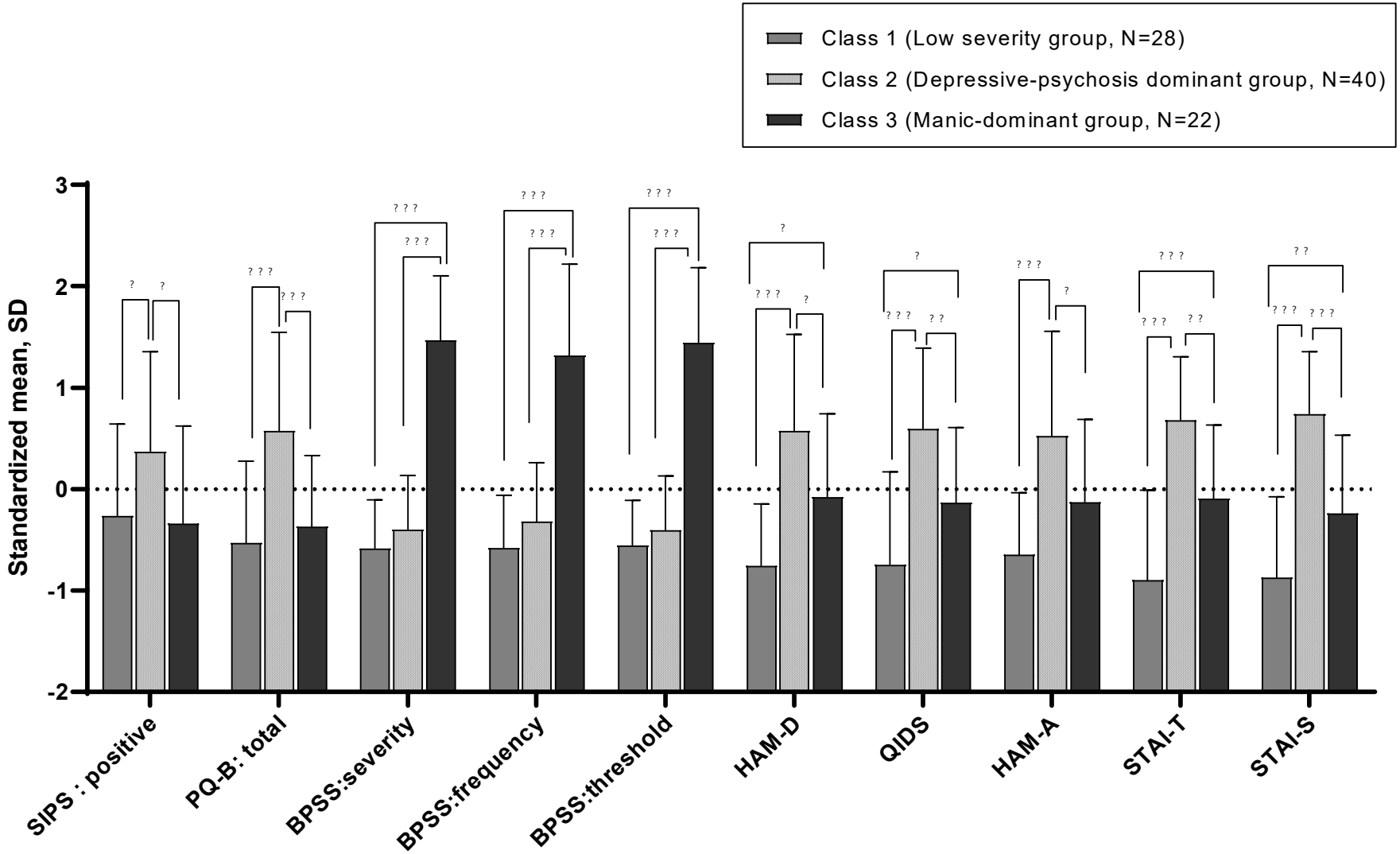


Figure 2-1. Mean and SD of clinical variables between Class 1 (Low severity group), Class 2 (Depressive-psychosis dominant group), and Class 3 (Manic-dominant group)
* $p < 0.05$, **$p < 0.01$, ***$p < 0.001$

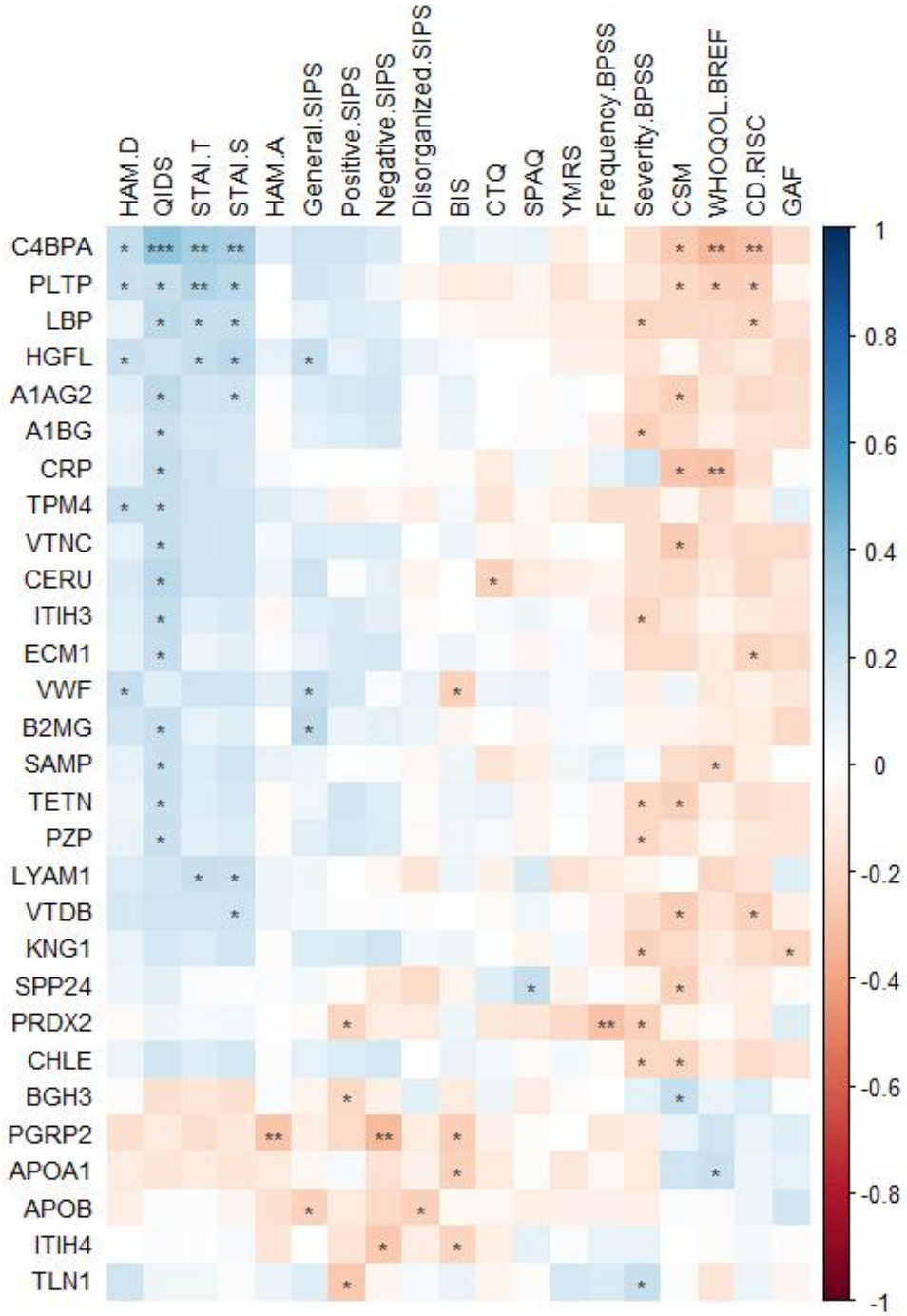


Figure 2-2. Correlation plot[a] showing Pearson's correlation analysis between 19 clinical variable and peptides [b]

---

* p <0.05, ** p < 0.01, *** p < 0.001

[a] Positive and negative correlation coefficients are colored blue and red, respectively

[b] Proteins that showed significant correlation with at least two of the 19 clinical variables were selected.

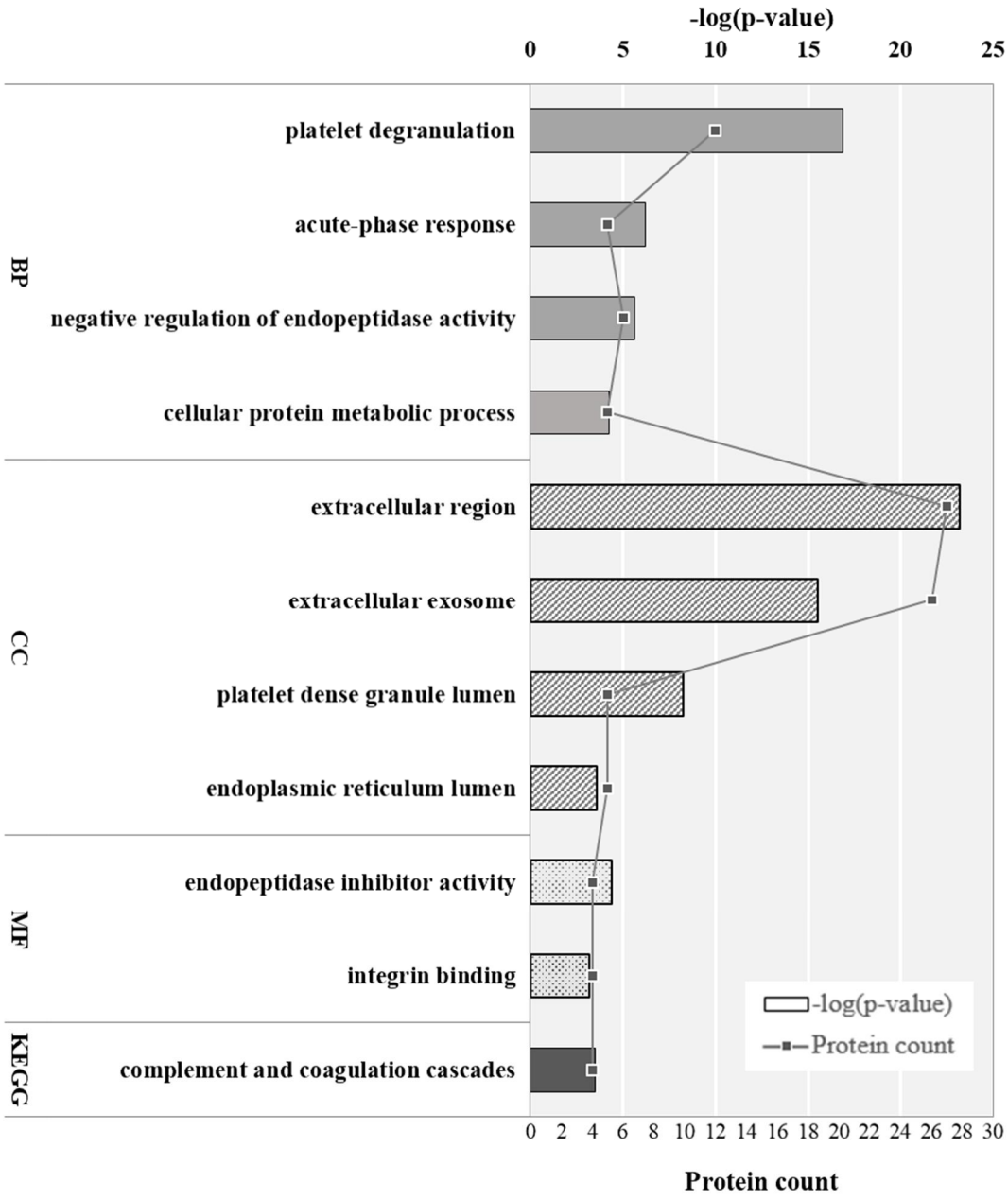


Figure 2-3. Bioinformatics analysis[a] of the proteins significantly reflect severity of symptoms

[a] GO analysis of the protein (DAVID Bioinformatics resources 6.8)

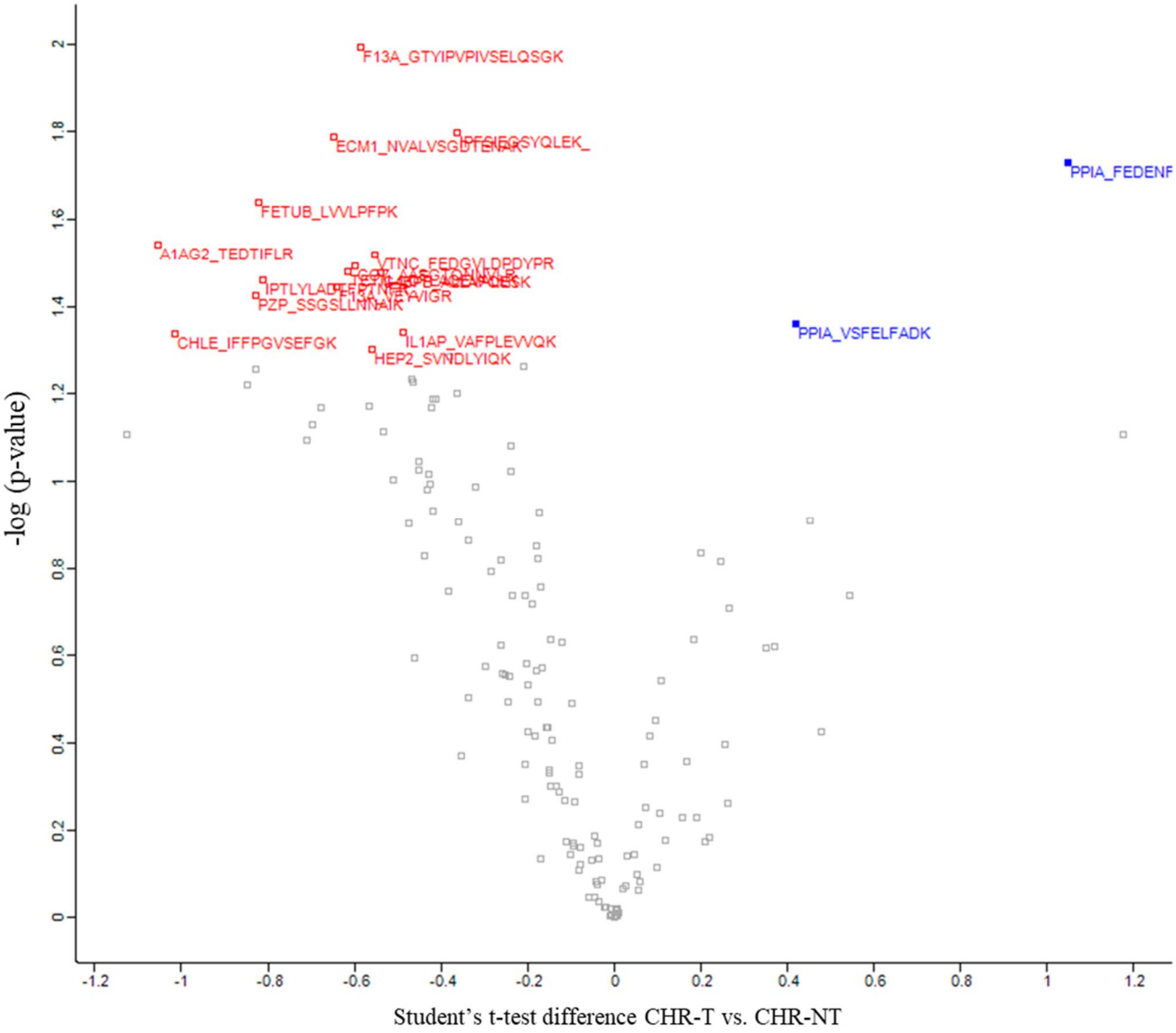


Figure 3-1. Volcano plot of *p*-values for DEPs between CHR-T and CHR-NT

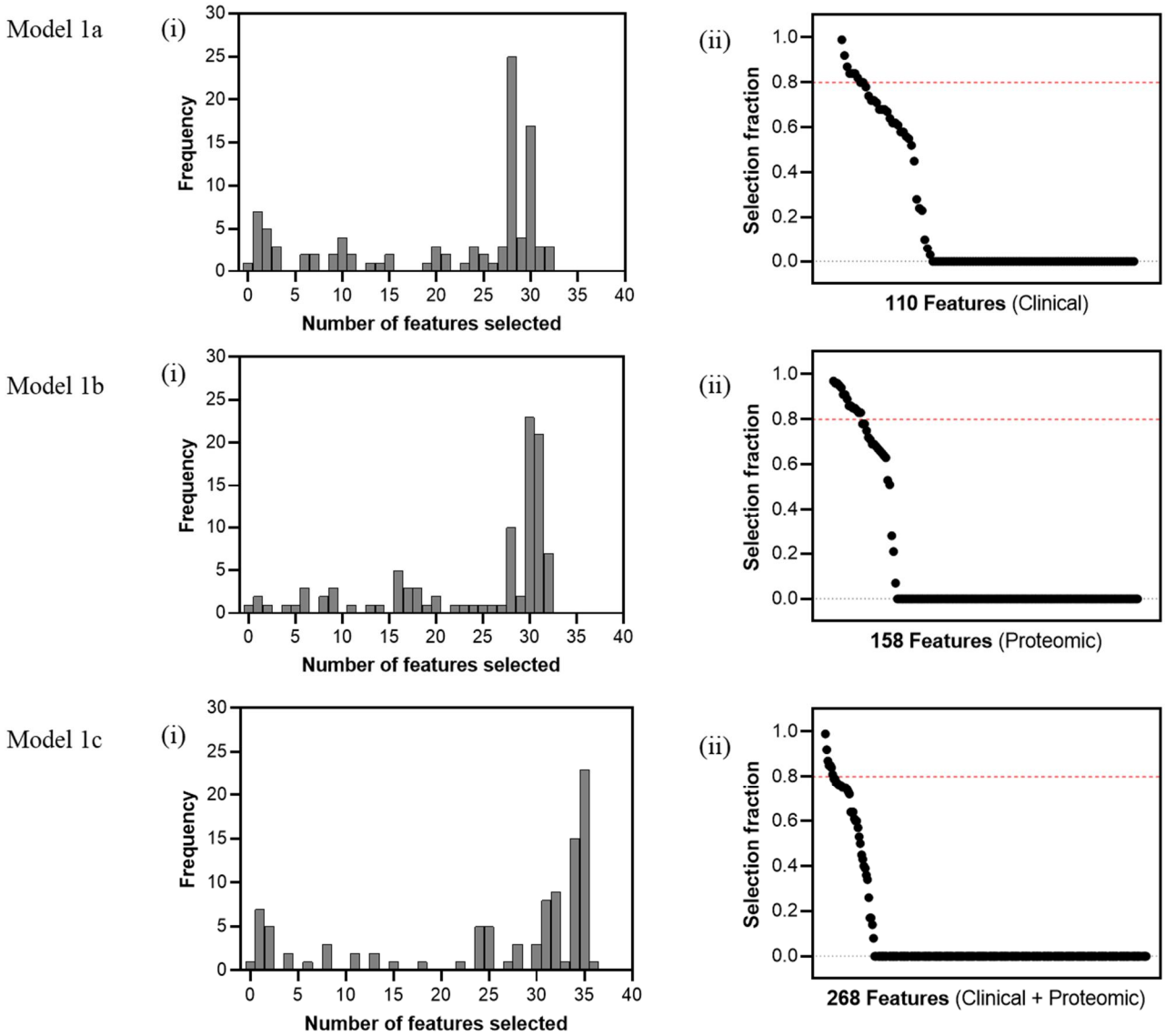


Figure 3-2. Feature selection[a] among 100 models analyzed from repeated application of LASSO regression with 10-fold CV on the training data (Dependent variable: Transition)

[a] Model 1a : Model selection including 110 clinical features, Model 1b : Model selection including 158 proteomic features, Model 1c : Model selection including 268 feature

(i) The number of features selected in each model

(ii) Selection fractions of each feature

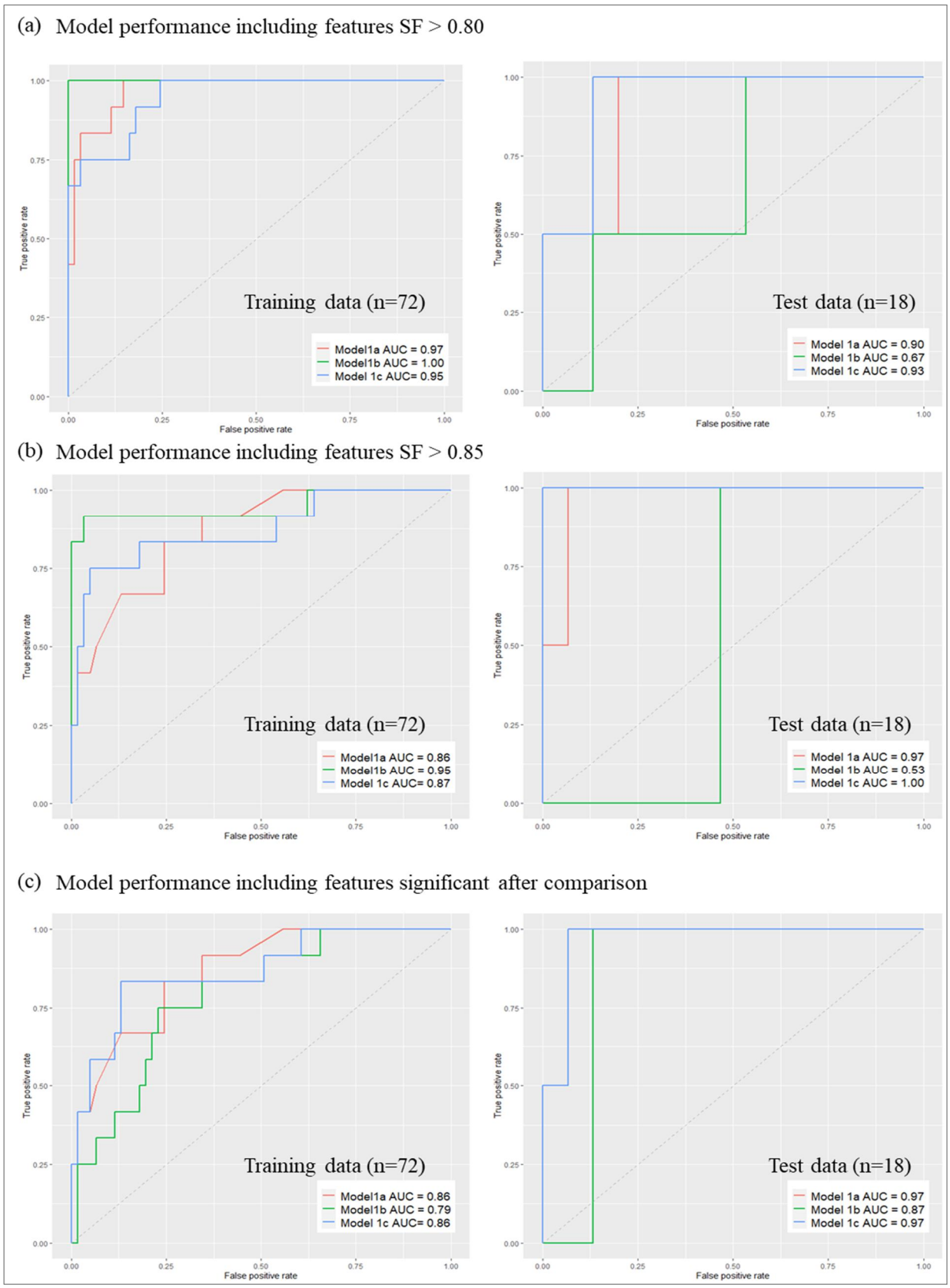


Figure 3-3. ROC curves showing model performance in predicting the probability of Transition (CHR-T).

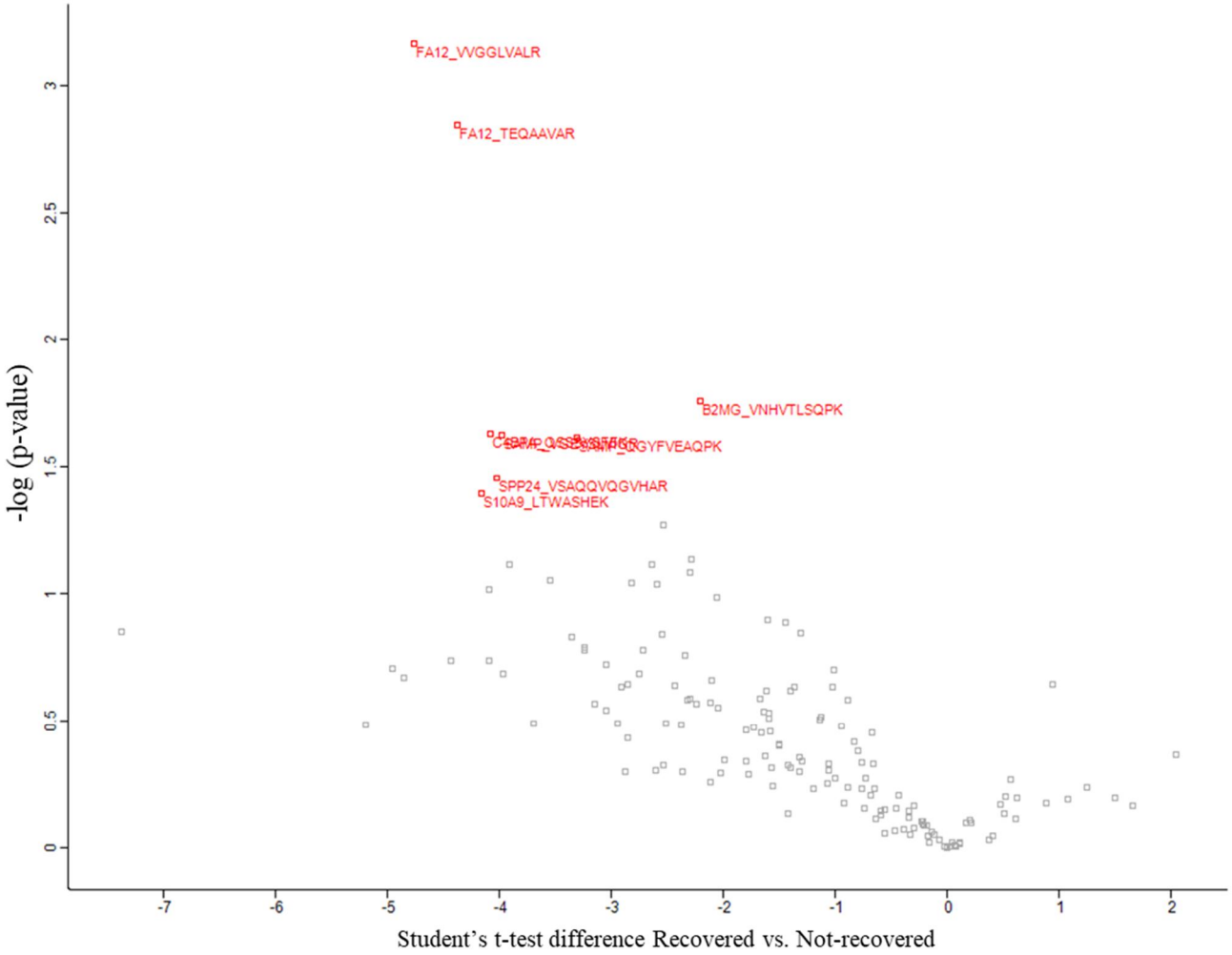


Figure 3-4. Volcano plot of *p*-values for DEPs between Recovered and Not-recovered group

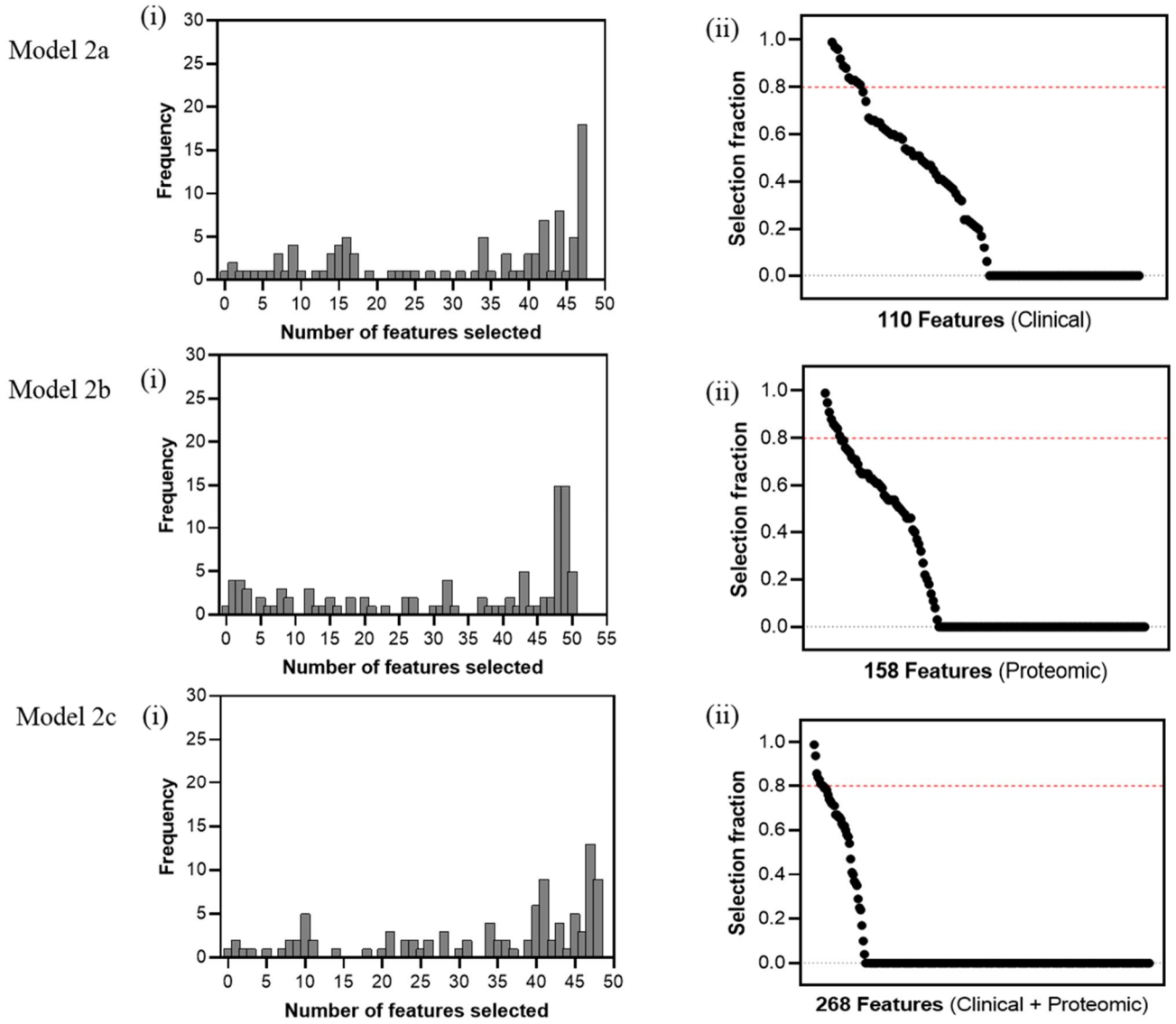


Figure 3-5. Feature selection[a] among 100 models analyzed from repeated application of LASSO regression with 10-fold CV on the training data (Dependent variable: Symptom recover)

[a] Model 1a : Model selection including 110 clinical features, Model 1b : Model selection including 158 proteomic features, Model 1c : Model selection including 268 feature

(i) The number of features selected in each model

(ii) Selection fractions of each feature

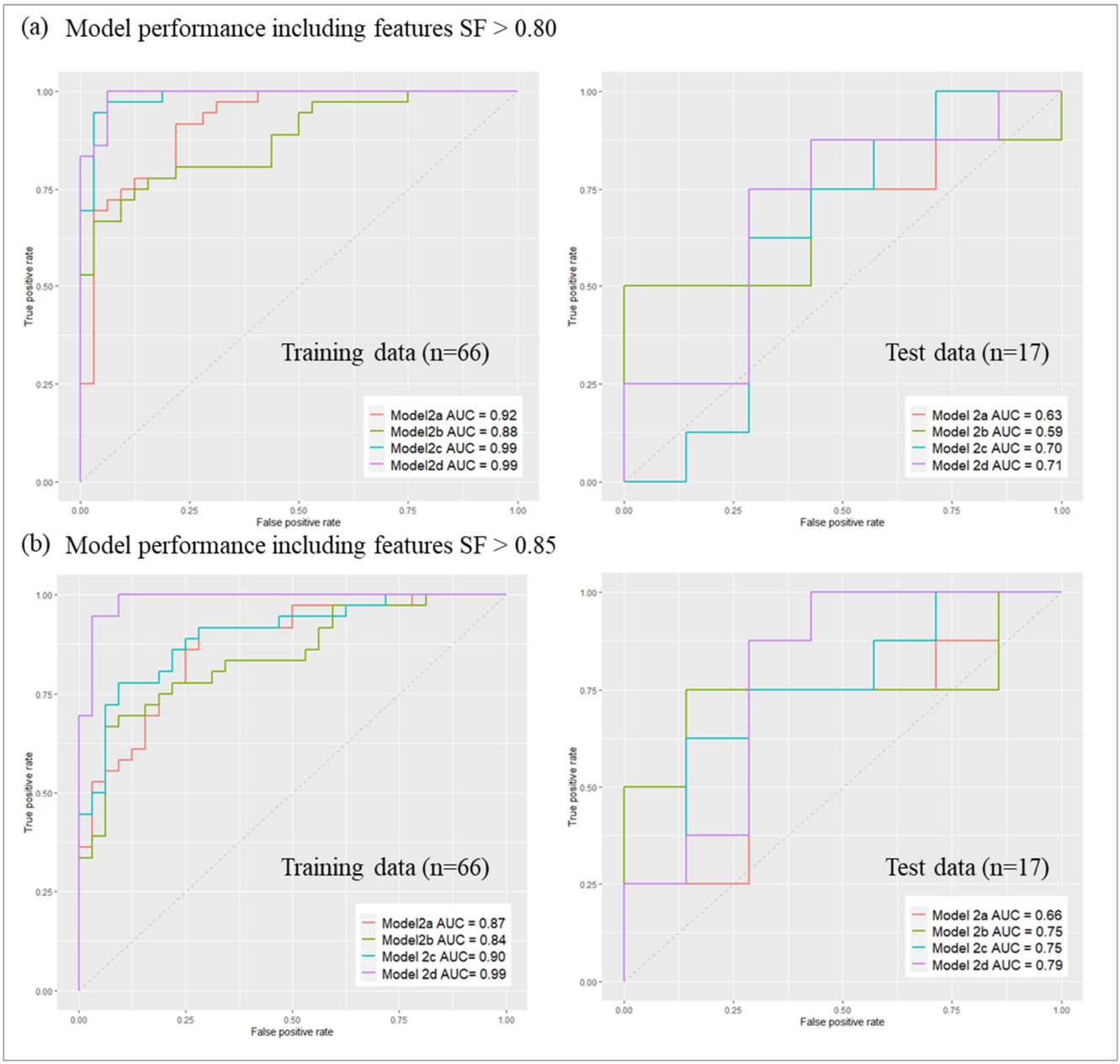


Figure 3-6. ROC curves showing model performance in predicting the probability of symptom recover.

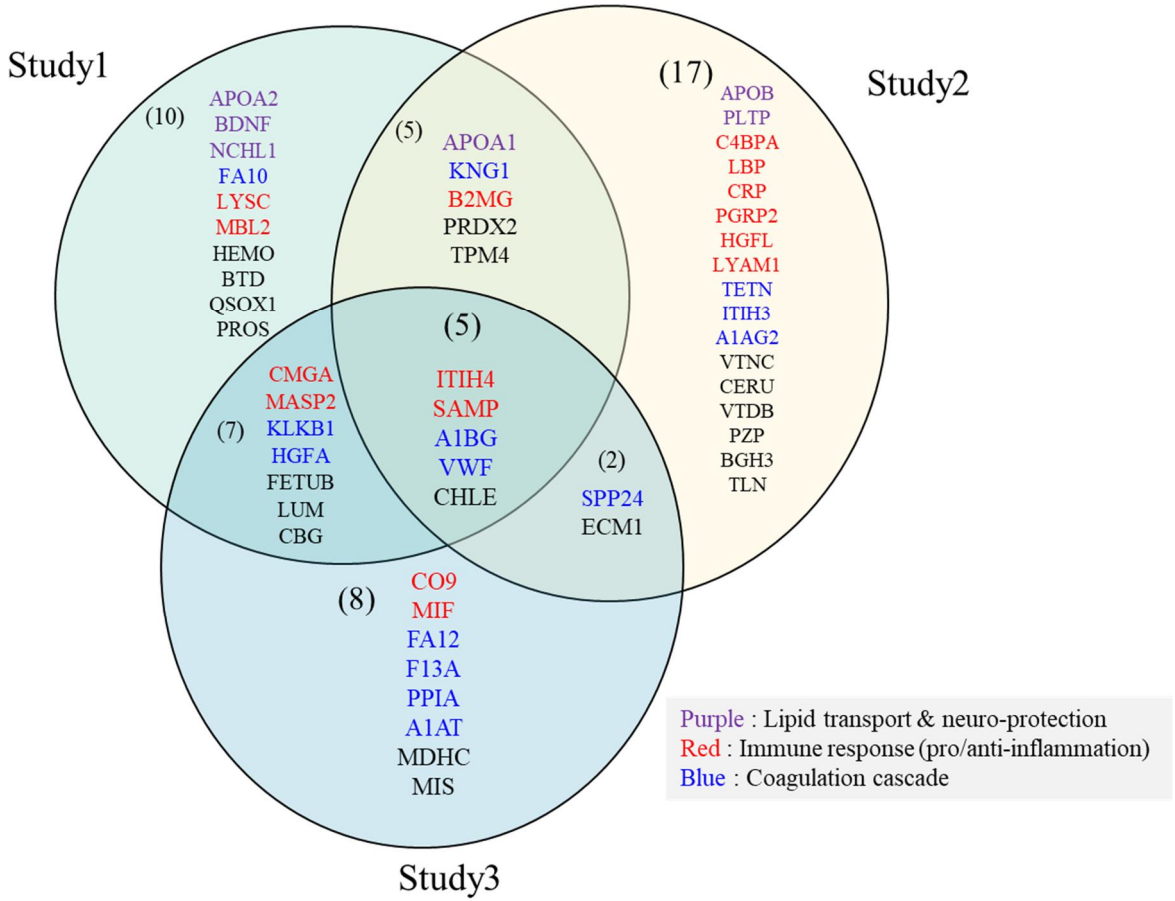


Figure 4. Summary[a] of the results of study 1-3

---

[a] Study 1 : Proteins differentially expressed in CHR (compared to HC, FEP or BD)

Study 2 : Proteins significantly related to at least 2 symptom severity

Study 3 : Proteins selected in LASSO regression, could reflect Transition or Symptom recovery

참고 문헌

1. Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012;380(9859):2197-223.

2. Go D-S, Kim Y-E, Paik J-W, Roh S, Yoon S-J. A comparison of disease burden and the government budget for mental health in Korea. Journal of Mental Health. 2020:1-8.

3. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2163-96.

4. Woods SW, Carlson JP, McGlashan TH. DSM-5 and the 'Psychosis Risk Syndrome': The DSM-5 proposal is better than DSM-IV. Psychosis. 2010;2(3):187-90.

5. Fusar-Poli P, Schultze-Lutter F, Cappucciati M, Rutigliano G, Bonoldi I, Stahl D, et al. The Dark Side of the Moon: Meta-analytical Impact of Recruitment Strategies on Risk Enrichment in the Clinical High Risk State for Psychosis. Schizophr Bull. 2016;42(3):732-43.

6. Fusar-Poli P, Nelson B, Valmaggia L, Yung AR, McGuire PK. Comorbid depressive and anxiety disorders in 509 individuals with an at-risk mental state: impact on psychopathology and transition to psychosis. Schizophr Bull. 2014;40(1):120-31.

7. Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, et al. Identification of a Common Neurobiological Substrate for Mental Illness. JAMA Psychiatry. 2015;72(4):305-15.

8. Guest PC. Proteomic Methods in Neuropsychiatric Research. 1st ed. 2017.. ed: Cham : Springer International Publishing : Imprint: Springer; 2017.

9. Janiri D, Moser DA, Doucet GE, Luber MJ, Rasgon A, Lee WH, et al. Shared Neural Phenotypes for Mood and Anxiety Disorders: A Meta-analysis of 226 Task-Related Functional Imaging Studies. JAMA Psychiatry. 2020;77(2):172-9.

10. Lisa M. McTeague, Ph.D. ,, Julia Huemer, M.D. ,, David M. Carreon, M.D. ,, Ying Jiang, M.D. ,, Simon B. Eickhoff, Dr.Med. ,, Amit Etkin, M.D., Ph.D. Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. American Journal of Psychiatry. 2017;174(7):676-85.

11. McGorry PD, Purcell R, Hickie IB, Yung AR, Pantelis C, Jackson HJ. Clinical staging: a heuristic model for psychiatry and youth mental health. Medical Journal of Australia. 2007;187(S7):S40-S2.

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

13. Ren J, Zhao G, Sun X, Liu H, Jiang P, Chen J, et al. Identification of plasma biomarkers for distinguishing bipolar depression from major depressive disorder by iTRAQ-coupled LC-MS/MS and bioinformatics analysis. Psychoneuroendocrinology. 2017;86:17-24.

14. Rodrigues-Amorim D, Rivera-Baltanás T, Vallejo-Curto MdC, Rodriguez-Jamardo C, de las Heras E, Barreiro-Villar C, et al. Proteomics in Schizophrenia: A Gateway to Discover Potential Biomarkers of Psychoneuroimmune Pathways. Frontiers in Psychiatry. 2019;10(885).

15. Li Y, Zhou K, Zhang Z, Sun L, Yang J, Zhang M, et al. Label-free quantitative proteomic analysis reveals dysfunction of complement pathway

in peripheral blood of schizophrenia patients: evidence for the immune hypothesis of schizophrenia. Mol Biosyst. 2012;8(10):2664-71.

16. Schmidt A, Cappucciati M, Radua J, Rutigliano G, Rocchetti M, Dell'Osso L, et al. Improving Prognostic Accuracy in Subjects at Clinical High Risk for Psychosis: Systematic Review of Predictive Models and Meta-analytical Sequential Testing Simulation. Schizophr Bull. 2017;43(2):375-88.

17. Lin A, Reniers RLEP, Wood SJ. Clinical staging in severe mental disorder: evidence from neurocognition and neuroimaging. British Journal of Psychiatry. 2013;202(s54):s11-s7.

18. Sartori JM, Reckziegel R, Passos IC, Czepielewski LS, Fijtman A, Sodré LA, et al. Volumetric brain magnetic resonance imaging predicts functioning in bipolar disorder: A machine learning approach. Journal of Psychiatric Research. 2018;103:237-43.

19. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma Proteome Profiling to Assess Human Health and Disease. Cell systems. 2016;2(3):185-95.

20. Geyer PE, Holdt LM, Teupser D, Mann M. Revisiting biomarker discovery by plasma proteomics. Mol Syst Biol. 2017;13(9):942-.

21. Moos T, Høyer PE. Detection of plasma proteins in CNS neurons: conspicuous influence of tissue-processing parameters and the utilization of serum for blocking nonspecific reactions. J Histochem Cytochem. 1996;44(6):591-603.

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

23. Shin D, Rhee SJ, Lee J, Yeo I, Do M, Joo EJ, et al. Quantitative

Proteomic Approach for Discriminating Major Depressive Disorder and Bipolar Disorder by Multiple Reaction Monitoring-Mass Spectrometry. J Proteome Res. 2021;20(6):3188-203.

24. Lee TY, Lee J, Lee HJ, Lee Y, Rhee SJ, Park DY, et al. Study Protocol for a Prospective Longitudinal Cohort Study to Identify Proteomic Predictors of Pluripotent Risk for Mental Illness: The Seoul Pluripotent Risk for Mental Illness Study. Front Psychiatry. 2020;11:340.

25. Hartmann JA, Yuen HP, McGorry PD, Yung AR, Lin A, Wood SJ, et al. Declining transition rates to psychotic disorder in "ultra-high risk" clients: Investigation of a dilution effect. Schizophrenia Research. 2016;170(1):130-6.

26. Yung AR, Yuen HP, Berger G, Francey S, Hung T-C, Nelson B, et al. Declining Transition Rate in Ultra High Risk (Prodromal) Services: Dilution or Reduction of Risk? Schizophr Bull. 2007;33(3):673-81.

27. Hartmann JA, Nelson B, Spooner R, Paul Amminger G, Chanen A, Davey CG, et al. Broad clinical high-risk mental state (CHARMS): Methodology of a cohort study validating criteria for pluripotent risk. Early Interv Psychiatry. 2019;13(3):379-86.

28. Tandy J. Miller, Ph.D. ,, Thomas H. McGlashan, M.D. ,, Joanna Lifshey Rosen, Psy.D. ,, Lubna Somjee, Ph.D. ,, Philip J. Markovich, B.A. ,, Kelly Stein, Ph.D. , and, et al. Prospective Diagnosis of the Initial Prodrome for Schizophrenia Based on the Structured Interview for Prodromal Syndromes: Preliminary Evidence of Interrater Reliability and Predictive Validity. American Journal of Psychiatry. 2002;159(5):863-5.

29. Miller TJ, McGlashan TH, Rosen JL, Cadenhead K, Cannon T, Ventura J, et al. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive

validity, interrater reliability, and training to reliability. Schizophr Bull. 2003;29(4):703-15.

30. Correll CU, Olvet DM, Auther AM, Hauser M, Kishimoto T, Carrión RE, et al. The Bipolar Prodrome Symptom Interview and Scale-Prospective (BPSS-P): description and validation in a psychiatric sample and healthy controls. Bipolar Disorders. 2014;16(5):505-22.

31. Fusar-Poli P, De Micheli A, Rocchetti M, Cappucciati M, Ramella-Cravaro V, Rutigliano G, et al. Semistructured Interview for Bipolar At Risk States (SIBARS). Psychiatry Research. 2018;264:302-9.

32. Hartmann JA, Nelson B, Spooner R, Paul Amminger G, Chanen A, Davey CG, et al. Broad clinical high-risk mental state (CHARMS): Methodology of a cohort study validating criteria for pluripotent risk. Early Intervention in Psychiatry. 2019;13(3):379-86.

33. Saraf G, Moazen-Zadeh E, Pinto JV, Ziafat K, Torres IJ, Kesavan M, et al. Early intervention for people at high risk of developing bipolar disorder: a systematic review of clinical trials. Lancet Psychiatry. 2021;8(1):64-75.

34. Howie LD, Pastor PN, Lukacs SL. Use of medication prescribed for emotional or behavioral difficulties among children aged 6-17 years in the United States, 2011-2012. NCHS Data Brief. 2014(148):1-8.

35. Duffy A, Horrocks J, Doucette S, Keown-Stoneman C, Grof P, Andreazza A, et al. Immunological and neurotrophic markers of risk status and illness development in high-risk youth: understanding the neurobiological underpinnings of bipolar disorder. International Journal of Bipolar Disorders. 2014;2(1):4.

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

37. 이중서, 배승오, 안용민, 박두병, 노경선, 신현균, et al. 한국판 Hamilton 우울증 평가 척도의 신뢰도, 타당도 연구. 신경정신의학. 2005;44(4):456-65.

38. Park HW, Lee, E.H., Kim, D.K., Yu, B.H., Lee, D.S., Kim, J.H. The factor structure of the Korean hamilton depression rating scale (K-HDRS): A confirmatory factor analysis. J Korean Neuropsychiatr Assoc. 2009;48(1):21-8.

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

40. 정희연, 조현상, 주연호, 신현균, 이중서, 황석현, et al. Young 조증 평가 척도의 한국판 표준화 연구. 신경정신의학. 2003;42(2):263-9.

41. Mohammadi Z, Pourshahbaz A, Poshtmashhadi M, Dolatshahi B, Barati F, Zarei M. Psychometric Properties of the Young Mania Rating Scale as a Mania Severity Measure in Patients With Bipolar I Disorder. Practice in Clinical Psychology. 2018;6(3):175-82.

42. Thompson E. Hamilton Rating Scale for Anxiety (HAM-A). Occupational Medicine. 2015;65(7):601-.

43. 김창윤. 정신장애 평가도구: 하나의학사; 2001. 278 p.

44. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull. 1987;13(2):261-76.

45. 이중서, 안용민, 신현균, 안석균, 주연호, 김승현, et al. 한국판 양성 및 음성증후군 척도(Positive and Negative Syndrome Scale)의 신뢰도와 타당도. 신경정신의학. 2001;40(6):1090-105.

46. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont). 2007;4(7):28-37.

47. Yi JS CM, Lee BW, Lee SI. Reliability and Validity of the

Korean Version of Global Assessment of Functioning Scale. Korean J Schizophr Res. 2003;6(2):118-24.

48. 정명훈. 한국어판 총괄기능평가척도(Global Assessment of Functioning Scale)의 신뢰도와 타당도. 춘천: 翰林大學校 大學院; 2003.

49. First MB. Structured Clinical Interview for DSM-IV Axis I Disorders. Biometrics Research Department. 1997.

50. Hahn OS, Ahn JH, Song SH, Cho M, Kim J-K, Bae JN. Development of Korean version of Structured Clinical Interview Schedule for DSM-IV Axis I Disorder: Interrater reliability. J Korean Neuropsychiatr Assoc. 2000;39:362-72.

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

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

53. Mongan D, Föcking M, Healy C, Susai SR, Heurich M, Wynne K, et al. Development of Proteomic Prediction Models for Transition to Psychotic Disorder in the Clinical High-Risk State and Psychotic Experiences in Adolescence. JAMA Psychiatry. 2021;78(1):77-90.

54. Khoury R, Nasrallah HA. Inflammatory biomarkers in individuals at clinical high risk for psychosis (CHR-P): State or trait? Schizophr Res. 2018;199:31-8.

55. Mantere O, Saarela M, Kieseppä T, Raij T, Mäntylä T, Lindgren M, et al. Anti-neuronal anti-bodies in patients with early psychosis. Schizophr Res. 2018;192:404-7.

56. Kim H, Rhee SJ, Lee H, Han D, Lee TY, Kim M, et al. Identification of altered protein expression in major depressive disorder and bipolar disorder patients using liquid chromatography-tandem mass spectrometry. Psychiatry Res. 2021;299:113850.

57. Lee H, Rhee SJ, Kim J, Lee Y, Kim H, Lee J, et al. Predictive protein markers for depression severity in mood disorders: A preliminary trans-diagnostic approach study. J Psychiatr Res. 2021;142:63-72.

58. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26(7):966-8.

59. McTeague LM, Goodkind MS, Etkin A. Transdiagnostic impairment of cognitive control in mental illness. J Psychiatr Res. 2016;83:37-46.

60. Tanaka T, Biancotto A, Moaddel R, Moore AZ, Gonzalez-Freire M, Aon MA, et al. Plasma proteomic signature of age in healthy humans. Aging Cell. 2018;17(5):e12799.

61. Miike K, Aoki M, Yamashita R, Takegawa Y, Saya H, Miike T, et al. Proteome profiling reveals gender differences in the composition of human serum. PROTEOMICS. 2010;10(14):2678-91.

62. Huang C-C, Luo Q, Palaniyappan L, Yang AC, Hung C-C, Chou K-H, et al. Transdiagnostic and Illness-Specific Functional Dysconnectivity Across Schizophrenia, Bipolar Disorder, and Major Depressive Disorder. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2020;5(5):542-53.

63. R Development core team. R: A language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.

64. Thomas EA, Dean B, Pavey G, Sutcliffe JG. Increased CNS levels of apolipoprotein D in schizophrenic and bipolar subjects: implications for the pathophysiology of psychiatric disorders. Proc Natl Acad Sci U S A. 2001;98(7):4066-71.

65. Song YR, Wu B, Yang YT, Chen J, Zhang LJ, Zhang ZW, et al. Specific alterations in plasma proteins during depressed, manic, and euthymic states of bipolar disorder. Braz J Med Biol Res. 2015;48(11):973-82.

66. Zhou X, Barrett TB, Kelsoe JR. Promoter variant in the GRK3 gene associated with bipolar disorder alters gene expression. Biol Psychiatry. 2008;64(2):104-10.

67. Sams DS, Nardone S, Getselter D, Raz D, Tal M, Rayi PR, et al. Neuronal CTCF Is Necessary for Basal and Experience-Dependent Gene Regulation, Memory Formation, and Genomic Structure of BDNF and Arc. Cell Rep. 2016;17(9):2418-30.

68. Willis M, Leitner I, Jellinger KA, Marksteiner J. Chromogranin peptides in brain diseases. J Neural Transm (Vienna). 2011;118(5):727-35.

69. Knutsvik G, Collett K, Arnes J, Akslen LA, Stefansson IM. QSOX1 expression is associated with aggressive tumor features and reduced survival in breast carcinomas. Mod Pathol. 2016;29(12):1485-91.

70. Hoirisch-Clapauch S, Nardi AE, Gris JC, Brenner B. Coagulation and mental disorders. Rambam Maimonides Med J. 2014;5(4):e0036.

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

72. Foldager L, Köhler O, Steffensen R, Thiel S, Kristensen AS, Jensenius JC, et al. Bipolar and panic disorders may be associated with hereditary defects in the innate immune system. J Affect Disord. 2014;164:148-54.

73. Fernandes BS, Gama CS, Ceresér KM, Yatham LN, Fries GR, Colpo G, et al. Brain-derived neurotrophic factor as a state-marker of mood episodes in bipolar disorders: a systematic review and meta-regression analysis. J Psychiatr Res. 2011;45(8):995-1004.

74. Calabrò M, Mandelli L, Crisafulli C, Lee SJ, Jun TY, Wang SM, et al. Neuroplasticity, Neurotransmission and Brain-Related Genes in Major Depression and Bipolar Disorder: Focus on Treatment Outcomes in an Asiatic Sample. Adv Ther. 2018;35(10):1656-70.

75. Vieta E, Gastó C, Osaba MM, Otero A, Nieto E, Pintor L, et al. Cortisol-binding globulin levels in bipolar disorder. Eur Psychiatry. 1997;12(1):11-5.

76. Spinardi L, Witke W. Gelsolin and diseases. Subcell Biochem. 2007;45:55-69.

77. Frye MA, Nassan M, Jenkins GD, Kung S, Veldic M, Palmer BA, et al. Feasibility of investigating differential proteomic expression in depression: implications for biomarker development in mood disorders. Transl Psychiatry. 2015;5:e689.

78. Russo AJ. Decreased Serum Hepatocyte Growth Factor (HGF) in Individuals with Bipolar Disorder Normalizes after Zinc and Anti-oxidant Therapy. Nutr Metab Insights. 2010;3:49-55.

79. Lee S-Y, Wang T-Y, Lu R-B, Wang L-J, Li S-C, Tu C-Y, et al. Identification of potential plasma protein biomarkers for bipolar II disorder: a preliminary/exploratory study. Scientific Reports. 2021;11(1):9452.

80. Tramontina JF, Andreazza AC, Kauer-Sant'anna M, Stertz L, Goi J, Chiarani F, et al. Brain-derived neurotrophic factor serum levels before and after treatment for acute mania. Neurosci Lett. 2009;452(2):111-3.

81. Fernandes BS, Gama CS, Maria Ceresér K, Yatham LN, Fries GR, Colpo G, et al. Brain-derived neurotrophic factor as a state-marker of mood episodes in bipolar disorders: A systematic review and meta-regression analysis. Journal of Psychiatric Research. 2011;45(8):995-1004.

82. Passos IC, Mwangi B, Vieta E, Berk M, Kapczinski F. Areas of controversy in neuroprogression in bipolar disorder. Acta Psychiatrica Scandinavica. 2016;134(2):91-103.

83. Berk M, Kapczinski F, Andreazza AC, Dean OM, Giorlando F, Maes M, et al. Pathways underlying neuroprogression in bipolar disorder: Focus on inflammation, oxidative stress and neurotrophic factors. Neuroscience & Biobehavioral Reviews. 2011;35(3):804-17.

84. Tein JY, Coxe S, Cham H. Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. Struct Equ Modeling. 2013;20(4):640-57.

85. Jung MH, Jang JH, Kang DH, Choi JS, Shin NY, Kim HS, et al. The reliability and validity of the korean version of the structured interview for prodromal syndrome. Psychiatry Investig. 2010;7(4):257-63.

86. Loewy RL, Bearden CE, Johnson JK, Raine A, Cannon TD. The prodromal questionnaire (PQ): preliminary validation of a self-report screening measure for prodromal and psychotic syndromes. Schizophr Res. 2005;79(1):117-25.

87. Preti A, Raballo A, Kotzalidis GD, Scanu R, Muratore T, Gabbrielli M, et al. Quick Identification of the Risk of Psychosis: The Italian Version of the Prodromal Questionnaire-Brief. Clin Pract Epidemiol Ment Health.

2018;14:120-31.

88. Jang YE, Lee TY, Hur J-W, Kwon JS. Validation of the Korean Version of the Prodromal Questionnaire-Brief Version in Non-Help-Seeking Individuals. Psychiatry Investig. 2019;16(2):109-14.

89. Van Meter A, Guinart D, Bashir A, Sareen A, Cornblatt BA, Auther A, et al. Bipolar Prodrome Symptom Scale - Abbreviated Screen for Patients: Description and validation. Journal of Affective Disorders. 2019;249:357-65.

90. 홍진표, 전대종, 임아영, 박수빈, 박승진. 한국어판 자기보고형 간이우울증상평가척도(QIDS-SR)의 신뢰도 및 타당도. 우울조울병. 2013;11(1):44-50.

91. Spielberger CD. State-Trait anxiety inventory. The Corsini encyclopedia of psychology. 2010:1-.

92. Lee Y-J BS-Y, Choi S, Lee H-K, Kim B-W, Kim W, et al. . Korean State-Trait Anxiety Inventory Application Study in Middle and High School Students. J Korean Neuropsychiatr Assoc. 2008;47:471-80.

93. Kim D, Park SC, Yang H, Oh DH. Reliability and validity of the korean version of the childhood trauma questionnaire-short form for psychiatric outpatients. Psychiatry Investig. 2011;8(4):305-11.

94. Bernstein DP, Fink L, Handelsman L, Foote J, Lovejoy M, Wenzel K, et al. Initial reliability and validity of a new retrospective measure of child abuse and neglect. The American journal of psychiatry. 1994;151(8):1132-6.

95. Lee S-R, Lee W-H, Park J-S, Kim S-M, Kim J-W, Shim J-H. The Study on Reliability and Validity of Korean Version of the Barratt Impulsiveness Scale-11-Revised in Nonclinical Adult Subjects. J Korean Neuropsychiatr Assoc. 2012;51(6):378-86.

96. Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. J Clin Psychol. 1995;51(6):768-74.

97. Cho C-H, Jung S-Y, Kapczinski F, Rosa AR, Lee H-J. Validation of the Korean Version of the Biological Rhythms Interview of Assessment in Neuropsychiatry. Psychiatry Investig. 2018;15(12):1115-20.

98. Chul-Hyun C, Seo-Yeon J, Flávio K, Adriane RR, Heon-Jeong L. Validation of the Korean Version of the Biological Rhythms Interview of Assessment in Neuropsychiatry. PSYCHIATRY INVESTIGATION. 2018;15(12):1115-20.

99. 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.

100. 이형영, 윤진상, 국승희. 아침형-저녁형의 측정을 위한 한국어판 조합척도의 교차 타당화. 신경정신의학. 1999;38(2):297-305.

101. Baek JH, Kim JS, Huh I, Lee K, Park JH, Park T, et al. Prevalence, behavioral manifestations and associated individual and climatic factors of seasonality in the Korean general population. Comprehensive Psychiatry. 2015;57:148-54.

102. 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. Archives of General Psychiatry. 1984;41(1):72-80.

103. Kasper S, Wehr TA, Bartko JJ, Gaist PA, Rosenthal NE. Epidemiological Findings of Seasonal Changes in Mood and Behavior: A Telephone Survey of Montgomery County, Maryland. Archives of General Psychiatry. 1989;46(9):823-33.

104. Min SK, Kim KI, Lee CI, Jung YC, Suh SY, Kim DK. Development of

the Korean versions of WHO Quality of Life scale and WHOQOL-BREF. Quality of Life Research. 2002;11(6):593-600.

105. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med. 1998;28(3):551-8.

106. Baek H-S, Lee K-U, Joo E-J, Lee M-Y, Choi K-S. Reliability and Validity of the Korean Version of the Connor-Davidson Resilience Scale. Psychiatry Investig. 2010;7(2):109-15.

107. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety. 2003;18(2):76-82.

108. Cornblatt BA, Auther AM, Niendam T, Smith CW, Zinberg J, Bearden CE, et al. Preliminary Findings for Two New Measures of Social and Role Functioning in the Prodromal Phase of Schizophrenia. Schizophr Bull. 2007;33(3):688-702.

109. Nylund-Gibson K, Choi AY. Ten frequently asked questions about latent class analysis. Translational Issues in Psychological Science. 2018;4(4):440-61.

110. Scott J, Leboyer M, Hickie I, Berk M, Kapczinski F, Frank E, et al. Clinical staging in psychiatry: a cross-cutting model of diagnosis with heuristic and practical value. Br J Psychiatry. 2013;202(4):243-5.

111. McGorry PD, Nelson B, Goldstone S, Yung AR. Clinical staging: a heuristic and practical strategy for new research and better health and social outcomes for psychotic and related mood disorders. Can J Psychiatry. 2010;55(8):486-97.

112. Nieman DH, McGorry PD. Detection and treatment of at-risk mental state for developing a first psychosis: making up the balance. The Lancet

Psychiatry. 2015;2(9):825-34.

113. McGorry P, Keshavan M, Goldstone S, Amminger P, Allott K, Berk M, et al. Biomarkers and clinical staging in psychiatry. World psychiatry : official journal of the World Psychiatric Association (WPA). 2014;13(3):211-23.

114. Xu HB, Zhang RF, Luo D, Zhou Y, Wang Y, Fang L, et al. Comparative proteomic analysis of plasma from major depressive patients: identification of proteins associated with lipid metabolism and immunoregulation. Int J Neuropsychopharmacol. 2012;15(10):1413-25.

115. Domenici E, Wille DR, Tozzi F, Prokopenko I, Miller S, McKeown A, et al. Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections. PLoS One. 2010;5(2):e9166.

116. Jaros JAJ, Martins-de-Souza D, Rahmoune H, Rothermundt M, Leweke FM, Guest PC, et al. Protein phosphorylation patterns in serum from schizophrenia patients and healthy controls. Journal of Proteomics. 2012;76:43-55.

117. Lee MY, Kim EY, Kim SH, Cho K-C, Ha K, Kim KP, et al. Discovery of serum protein biomarkers in drug-free patients with major depressive disorder. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2016;69:60-8.

118. Schwarz E, Guest PC, Rahmoune H, Martins-de-Souza D, Niebuhr DW, Weber NS, et al. Identification of a blood-based biological signature in subjects with psychiatric disorders prior to clinical manifestation. World J Biol Psychiatry. 2012;13(8):627-32.

119. Raivisto T, Heikkinen AM, Silbereisen A, Kovanen L, Ruokonen H, Tervahartiala T, et al. Regulation of Salivary Peptidoglycan Recognition Protein 1 in Adolescents. JDR Clin Trans Res. 2020;5(4):332-41.

120. Arentsen T, Qian Y, Gkotzis S, Femenia T, Wang T, Udekwu K, et al. The bacterial peptidoglycan-sensing molecule Pglyrp2 modulates brain development and behavior. Mol Psychiatry. 2017;22(2):257-66.

121. Han SYS, Cooper JD, Ozcan S, Rustogi N, Penninx BWJH, Bahn S. Integrating proteomic, sociodemographic and clinical data to predict future depression diagnosis in subthreshold symptomatic individuals. Translational Psychiatry. 2019;9(1):277.

122. Cooper JD, Han SYS, Tomasik J, Ozcan S, Rustogi N, van Beveren NJM, et al. Multimodel inference for biomarker development: an application to schizophrenia. Translational Psychiatry. 2019;9(1):83.

123. Burnham KP, Anderson DR, Huyvaert KP. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behavioral Ecology and Sociobiology. 2011;65(1):23-35.

124. Burnham KP, Anderson DR. Multimodel Inference:Understanding AIC and BIC in Model Selection. Sociological Methods & Research. 2004;33(2):261-304.

125. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. In: Parzen E, Tanabe K, Kitagawa G, editors. Selected Papers of Hirotugu Akaike. New York, NY: Springer New York; 1998. p. 199-213.

126. HURVICH CM, TSAI C-L. Regression and time series model selection in small samples. Biometrika. 1989;76(2):297-307.

127. Sugiura N. Further analysts of the data by akaike' s information criterion and the finite corrections. Communications in Statistics - Theory and Methods. 1978;7(1):13-26.

128. Burnham KPA, D. R. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. second edn ed: Springer; 2002.

129. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics. 2005;21(20):3940-1.

130. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29-36.

131. Perez M, Farrant J. Immune reactions and mental disorders. Psychological Medicine. 1988;18(1):11-3.

Abstract

A Study on Clinical variables and Blood Proteins related to Diagnosis, reflect Severity, and Progress of High-Risk State of Psychiatric Disorders

Hyunju Lee
Department of Medicine (Clinical Medical Science)
The Graduate School
Seoul National University

Introduction: Early diagnosis and treatment of major mental disorders are very important in improving the prognosis and quality of life of young age. The existing diagnostic system has limitations in diagnosing high-risk patients with non-specific symptoms, and it is necessary to find biological markers that can reflect the disease status of high-risk groups. Therefore, in this study, blood proteins were quantitatively analyzed for high-risk groups of major mental illness, and protein markers reflecting the state of high-risk groups and disease progress were explored in various ways.

Method: 90 clinical high-risk subjects recruited from 7 medical

institutions in Seoul Clinical evaluation and blood collection were performed on the same day, and blood proteins were quantified using 158 target peptides with multiple reaction monitoring-mass spectrometry. The analyzed protein expression in the high-risk group was compared with that in the patient group and the normal control group. Protein markers reflecting the high-risk group status were explored(Study1), In addition, through latent profile analysis, heterogeneous characteristics of high-risk were categorized, and correlation analysis was performed to search for proteins reflecting symptom severity(Study2). 10-crossfold LASSO regression was repeated 100 times for each model, features predict transition to major psychiatric disorders, and recovery of symptoms were selected. Predictive performance of the models were measured by area under the receiver operating characteristics (Study3).

Results: Through study 1, Twenty-seven protein markers reflecting high-risk conditions were identified. Most of proteins were reported in previous studies related to psychiatric disorders. In study 2, clinical high-risk groups were categorized into three subgroups, and it was different from categorized group in study 1. Twenty-nine protein markers with significant correlation with two or more of the 19 clinical indicators could be selected. In study 3, clinical variables (9 markers in clinical models, 5 markers in mixed models) and protein (15 markers in proteomic models, 2 markers in mixed models) were selected to predict the transition to psychiatric disorders. The performance of the mixed model (Model 1c) was showed excellent performance in the test data (AUC 0.90-0.97). On the other hand, clinical variables (8 markers in clinical models, 9 markers mixed models) and protein variables (8 markers in clinical models, 5

markers in mixed models) were selected to predict symptom recovery of symptom. However, the performance of the prediction model was fair at AUC 0.70−0.75 in the test data. Protein markers discovered through study 1−3 were integrated and most of the significant proteins were related to the blood coagulation cascade, immune response, lipid metabolism, and antioxidant response pathways reported in previous study.

Conclusion: Through this study, it was confirmed that meaningful biological alterations reported in the patient group were also observed in the patients with high−risk state. In addition, it was possible to identify common protein markers that could reflect the severity of symptoms, and predict the progression of disease. However, in order to validate our results and confirm the stability of the discovered protein markers, it will be necessary to evaluate the longitudinal variation of the protein markers in the future. Nevertheless, the protein and clinical indicators discovered in this study are expected to be used as useful markers for early diagnosis of the conditions of high−risk states in the future and predicting the course of the disease.

Keywords : Major psychiatric disorder, High−risk state, Transition, Progress, Protein, Predictor

Student Number: 2020−32637