



Extracellular vesicle biomarkers for complement dysfunction in schizophrenia

©Ting Xue,^{1,2} Wenxin Liu,³ Lijun Wang,^{1,2} Yuan Shi,^{1,2} Ying Hu,⁴ Jing Yang,⁵ Guiming Li,⁵ Hongna Huang^{1,2} and ⊚Donghong Cui^{1,2,6}

Schizophrenia, a complex neuropsychiatric disorder, frequently experiences a high rate of misdiagnosis due to subjective symptom assessment. Consequently, there is an urgent need for innovative and objective diagnostic tools. In this study, we used cutting-edge extracellular vesicles' (EVs) proteome profiling and XGBoost-based machine learning to develop new markers and personalized discrimination scores for schizophrenia diagnosis and prediction of treatment response. We analysed plasma and plasma-derived EVs from 343 participants, including 100 individuals with chronic schizophrenia, 34 first-episode and drug-naïve patients, 35 individuals with bipolar disorder, 25 individuals with major depressive disorder and 149 age- and sex-matched healthy controls.

Our innovative approach uncovered EVs-based complement changes in patients, specific to their disease-type and status. The EV-based biomarkers outperformed their plasma counterparts, accurately distinguishing schizophrenia individuals from healthy controls with an area under curve (AUC) of 0.895, 83.5% accuracy, 85.3% sensitivity and 82.0% specificity. Moreover, they effectively differentiated schizophrenia from bipolar disorder and major depressive disorder, with AUCs of 0.966 and 0.893, respectively. The personalized discrimination scores provided a personalized diagnostic index for schizophrenia and exhibited a significant association with patients' antipsychotic treatment response in the follow-up cohort. Overall, our study represents a significant advancement in the field of neuropsychiatric disorders, demonstrating the potential of EV-based biomarkers in guiding personalized diagnosis and treatment of schizophrenia.

- 1 Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai 201108, China
- 2 Shanghai Key Laboratory of Psychotic Disorders, Brain Health Institute, Shanghai Mental Health Center, Shanghai 201108, China
- 3 College of Life Sciences, Shanghai Normal University, Shanghai 200234, China
- 4 Shenzhi Department, Fourth Affiliated Hospital of Xinjiang Medical University, Urumqi 830000, China
- 5 Department of Hematology, Tongji Hospital, Frontier Science Center for Stem Cell Research, Shanghai Key Laboratory of Signaling and Disease Research, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
- 6 Brain Science and Technology Research Center, Shanghai Jiao Tong University, Shanghai 200240, China

Correspondence to: Donghong Cui Shanghai Mental Health Center Shanghai Jiao Tong University School of Medicine Shanghai 201108, China E-mail: manyucc@126.com

Correspondence may also be addressed to: Ting Xue

E-mail: xueting221314@126.com

Keywords: extracellular vesicle; proteomics; machine learning; schizophrenia; antipsychotic response

Introduction

Schizophrenia is a severe psychiatric disorder, significantly affecting individuals' well-being and imposing a substantial societal health burden. Its diagnosis primarily relies on subjective assessments of clinical symptomatology. However, this approach is prone to misdiagnosis due to symptom overlap with other severe mental disorders like bipolar disorder (BD) and major depressive disorder (MDD). In addition, patients' subjective feelings and expressions can easily affect clinicians' assessments. Epidemiologic studies have revealed a high misdiagnosis rate of nearly 25% in clinical practice.² Meanwhile, there is high heterogeneity of response to antipsychotic drugs among patients.3 These clinical issues lead to delayed or inappropriate treatment for schizophrenia. Consequently, reliable and objective biomarkers that can guide clinicians towards precise diagnoses and treatment is urgently needed. As an emerging form of liquid biopsy, the billions of extracellular vesicles (EVs) in peripheral circulation offer a valuable and noninvasive resource.4

EVs are bilayer membrane-enclosed nanoparticles secreted from cells and tissues, primarily divided into microvesicles (150-1000 nm) and exosomes (30-150 nm) depending on their size and biogenesis. They play critical roles in transporting cell-derived biomolecules for intracellular signalling and cell-to-cell communication.5 EVs have recently been implicated in the pathogenesis of neurodegenerative diseases.⁶ They are enriched in proteins, some of which appear to change in amount during the pathogenic processes of neurodegenerative diseases. 7-10 Pharmaceutical treatment could also alter the protein compositions of EVs. 11,12 Therefore, mounting evidence suggests that EV proteins are effective diagnostic and prognostic biomarkers in neurological diseases, including Parkinson's disease and Alzheimer's disease. 13-15

Currently, research on EV-derived proteins in schizophrenia is limited and lacks comparative analyses with other psychiatric disorders exhibiting similar symptoms. 16-18 Previous studies primarily detected individual EV proteins associated with mitochondrial activity and insulin signalling by electrochemiluminescence or ELISA assay. 16-18 The EV proteome of schizophrenia remains largely unknown, which could provide more comprehensive biomarkers and insights into disease mechanisms. Furthermore, a potent approach is essential for training these biomarkers, yielding robust models for disease diagnosis and prediction. EXtreme Gradient Boosting (XGBoost) has been demonstrated as a powerful machine learning algorithm for generating accurate classifications and predictions, even with small sample sizes. It is an ensemble method that utilizes gradient-boosted trees and can handle high dimensional and sparse data.¹⁹ Herein, we conducted mass spectrometry-based analysis of the EV proteome and applied XGBoost-based machine learning to develop new EV-based biomarkers and yield personalized discrimination scores (PDS) for schizophrenia diagnosis and prediction of antipsychotic responses (Supplementary Fig. 1A).

We isolated plasma and plasma-derived EVs from 343 participants, including individuals with schizophrenia, BD, MDD and healthy controls. First, we examined the EV proteome of schizophrenia in a sex- and age- matched case-control cohort. Second, potential EV-based biomarkers were validated in another schizophrenia cohort. Plasma proteins were used as a comparison to EV proteins. Third, we applied XGBoost-based machine learning to train and test EV-based biomarkers in schizophrenia, BD and MDD cohorts. Finally, we used the optimized schizophrenia-specific biomarkers to build PDS for diagnosis and evaluating antipsychotic response at an individual level (Supplementary Fig. 1).

Our study successfully identified a panel of EV-based biomarkers that achieved high area under the curve (AUC), accuracy, specificity and sensitivity in distinguishing patients with schizophrenia from healthy control subjects, outperforming their plasma counterparts. Moreover, these EV-based biomarkers effectively discriminated schizophrenia from BD and MDD. Notably, the PDS we constructed exhibits high efficacy in individual-level schizophrenia diagnosis and antipsychotic response prediction, which holds significant clinical value. Our discovery provides a novel and objective approach for schizophrenia detection and treatment evaluation, providing valuable insights into the pathophysiology of schizophrenia.

Materials and methods

Study design

The study aimed to develop novel EV-based protein biomarkers and PDS for specific diagnosis of schizophrenia and prediction of antipsychotic response. We analysed plasma and plasma-derived EV samples from individuals with schizophrenia, BD, MDD and healthy controls. Participants were recruited from Shanghai Mental Health Center (SMHC) between 2017 and 2020. Consensus diagnosis of these illnesses was made by two experienced psychiatrists according to the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV). For longitudinal samples of schizophrenia, patients were treated with antipsychotic medication for ~3 months. Treatment response was defined as a reduction rate of $\geq 25\%^{20}$ in the Positive and Negative Syndrome Scale (PANSS).²¹ All participants or their relatives provided informed consent and the study was approved by the Institutional Review Board of SMHC. The experiments were conducted in accordance with the Declaration of Helsinki. Participants who met the inclusion and exclusion criteria and were willing to provide a minimum of 10 ml of blood were included. To ensure comparability between groups, participants were matched on age and sex. Ultimately, this study included a total of 100 individuals with chronic schizophrenia, 34 first-episode and drug-naïve (FEDN) patients, 35 individuals with BD, 25 individuals with MDD and 149 healthy controls. This sample size was sufficient for analyses. All investigators involved in EV isolation and biomarker quantification were blinded until all measurements were made and the dataset was blocked for analysis.

Isolation and characterization of extracellular vesicles

Plasma-derived EVs were isolated by sequential ultracentrifugation.^{22,23} Typical EV markers used were rabbit-anti-CD63 (Santa Cruz Biotechnology, SC-15363), mouse-anti-CD9 (Proteintech, 60232-1-Ig), and rabbit-anti-Flotillin-1 (Proteintech, 15571-1-AP). Albumin was visualized by Ponceau S staining. The size distribution and concentration of EVs were measured by nanoparticle tracking analysis (NTA) (Particle Metrix). Transmission electron microscopy (TEM) (FEI Company) with negative staining was performed to characterize EV morphology.

Analysis of extracellular vesicle proteome

EV proteins were extracted by sonication and RIPA lysis buffer. The filter aided sample preparation (FASP)²⁴ was performed to prepare peptides. A Synapt G2-Si quadrupole time-of-flight mass spectrometer (MS) equipped with ion mobility option (Waters Corporation) was used for sample analysis. Data were acquired in the HDMS^E

mode. MS raw spectra were processed in Waters Progenesis QI (QIP, version 3.0.2)²⁵ searched against the UniProt human proteomic database (version 2020/06). Data normalization was performed at peptide levels. Proteins were quantified using the TOP 3 method. 26,27 Proteins whose intensities exceeded the range of mean ± 2 sigma in each group were defined as outliers and removed. Only proteins identified in at least 50% samples in each group were retained for analysis. $\kappa\text{-nearest}$ neighbour ($\kappa\text{-NN}$) imputation was applied to impute the missing values. Three approaches were applied to identify differentially expressed proteins (DEPs): (i) Limma R package with Benjamini-Hochberg (BH) multiple correction; (ii) Samr R package with 1000 permutations and a false discovery rate (FDR) threshold of 0.05²⁸; and (iii) Students' t-test and BH multiple correction. Proteins with a fold change of $\geq |1.5|$ and an adjusted P-value <0.05 across all three methods were defined as DEPs. Biological function analyses were performed in g:profiler (https://biit.cs.ut.ee/gprofiler/).²⁹ Protein-protein interaction analysis was conducted in STRING (https://string-db.org/)30 and visualized with Cytoscape 3.6.1. mRNA expression data enriched in brain were obtained from The Human Protein Atlas (http://www. proteinatlas.org/).32 Weighted gene co-expression network analysis (WGCNA) package in R33 was used to build signed coexpression networks.

Quantification of complement DEPs in extracellular vesicles and plasma

Plasma-derived complement DEPs were quantified by ELISA assay. Complement C3 (C3, Cloud-Clone, SEA861Hu), Complement C4 (C4, Elabscience, E-EL-H6027), C4b-binding protein alpha chain (C4BPA, Cloud-Clone, SEB620Hu) and vitamin K-dependent protein S (PROS1, Cloud-Clone, SEB971Hu) were analysed. The C3 antibody recognizes the full-length of C3 and its fragments. The C4 antibody detects the full-length of the C4 protein.

EV-derived complement DEPs were quantified by Mesoscale Discovery electrochemiluminescence assays (MSD). Primary antibodies of C3 (21337-1-AP), C4 (22233-1-AP), C4BPA (11819-1-AP) and PROS1 (16910-1-AP) were used, which were all purchased from Proteintech Group, Inc. The C3 and C4 antibodies recognize the full-length of the proteins and their fragments. The C4 antibody does not differentiate between C4A and C4B.

Machine learning

Data processing and machine learning were performed in Python (version 3.7.3). The XGBoost package (version 0.90) was used for classification and prediction. Missing values were imputed with κ-NN using sklearn impute package. Features were selected using a decision tree. Cross-validation involved 5-fold training/test splits with 15 repetitions. Variables input into the XGBoost classifier were the normalized concentration of complement DEPs. Normalization was performed separately for each dataset and calculated as: (x – min) / (max - min). The XGBoost classifier was trained to predict disease probabilities, with the output being the predicted classes and their corresponding probabilities on a scale of 0 to 1. The probabilities were then transformed into z-scores, where a z-score > 0 indicates a higher likelihood of schizophrenia for the participant.

Statistical analysis

Normality of continuous variables was assessed by Shapiro-Wilk test. The Levene test was used to examine the homogeneity of variance. Two sample comparisons were performed with chi-square test, two-tailed Mann-Whitney U-test, Welch's test or Students' t-test, as appropriate. Smoking was adjusted as a covariate in MDD compared with healthy controls, and analysis of covariance (ANCOVA) was applied. Pearson or Spearman correlation test was used as appropriate, adjusted covariates as needed. P < 0.05 was considered statistically significant. Data were represented as mean ± standard deviation (SD) or displayed by box-and-whisker plots in which horizontal lines indicate the median. The rhombus indicates the mean. The hinges of the box denote the first and third quartiles above and below. The lower and upper whiskers represent 1.5 times the IQR (interquartile range). All statistical analyses were conducted in R package and visualized using ggplot2.

Results

Quality control of extracellular vesicles enriched by ultracentrifugation

In this study, we devised a step-wise approach to investigate plasma and plasma-derived EVs from 343 individuals, comprising both case and control groups. We isolated EVs using sequential ultracentrifugation and assessed their quality by canonical EVs markers (i.e. CD9, CD63 and Flotillin1) and morphology via NTA and TEM analysis (Supplementary Fig. 2A-C). To evaluate the potential contamination of EV samples by plasma proteins, we also examined the presence of albumin, a major plasma protein, and found it to be undetectable in EV samples (Supplementary Fig. 2A). Furthermore, we identified a significant correlation between EV numbers and their protein amount (Supplementary Fig. 2D, R = 0.79, P < 0.0001). Overall, our results suggest a relatively pure isolation of plasmaderived EVs using ultracentrifugation.

The extracellular vesicle proteome was differentially expressed in schizophrenia and healthy controls

EV proteome profiling was compared between age- and sex-matched schizophrenia patients (n = 20) and healthy controls (n = 28) (Supplementary Table 1). Patient-derived EVs exhibited a significantly higher count and a smaller size compared to controls' (all P < 0.05, Fig. 1A and B). Additionally, their EV numbers were associated with the disease course (Fig. 1C).

WGCNA clustered the total identified EV proteome (1262 proteins) into seven modules (Fig. 1D). Only the turquoise module showed a significant correlation with the disease state (R = 0.52, P < 0.001, Fig. 1E). The turquoise module also exhibits the highest proportion of DEPs compared to other modules, both in terms of the total proteins within the module and the total DEPs across all modules (Fig. 1F and G). For proteins to be considered as DEPs, they had to be identified by all three statistical methods (fold change \geq 1.5 or \leq 0.67, and $P_{adjusted} < 0.05)$ (Fig. 1H). Therefore, we proposed the turquoise module as the disease-modified module.

EV proteins in the turquoise module were generally highly expressed in patients compared to controls (ratio: 1.65 ± 0.68 , P < 0.001) (Fig. 1I) and they were primarily enriched in the complement network (Fig. 1J). Within this network, complement components such as C3, C4A, C4B, C4BPA and C4BPB were identified as DEPs. Additionally, C3, C4A and C4B were recognized as the hub genes within the network (Fig. 1K). Furthermore, through a literature review, we discovered that PROS1, a DEP in the turquoise module, forms a high-affinity complex with C4BP, 34-38 which has been proposed to have an important link with the complement system. Therefore, we described PROS1 as a protein with potential link

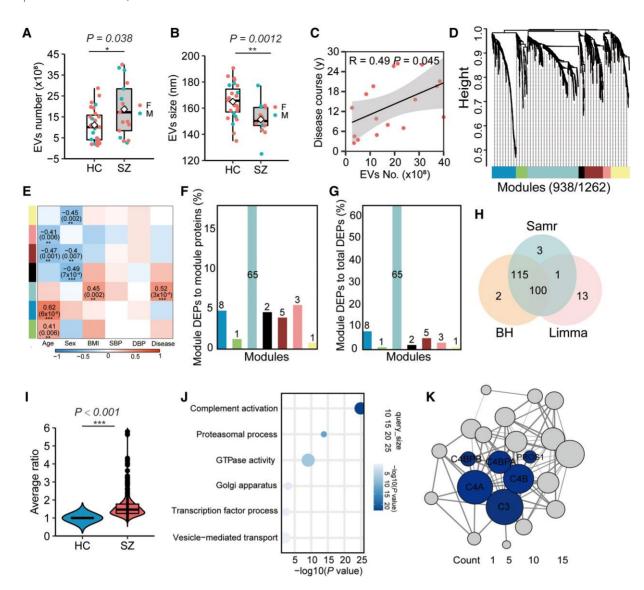


Figure 1 Proteome profiles of the plasma-derived EVs compared between patients with schizophrenia and healthy controls. Plasma-derived extracellular vesicles (EVs) were enriched by the ultracentrifugation strategy. (A and B) Comparison of EV number and EV size between patients with schizophrenia (SZ) and healthy controls (HC). No confounding covariates were adjusted. *P = 0.038, Cohen's d = 0.71; *P = 0.0012, Cohen's d = 1.07. (C) Pearson's correlation between the disease course and EV number in patients with schizophrenia. (D) Weighted gene co-expression network analysis (WGCNA) analysis of co-expression of EVs proteins in patients with schizophrenia and healthy controls. Each protein in the module has a correlation of ≥ 0.4 with the module eigengene. (E) Heat map showing the association of each WGCNA module with the phenotypes. Only the significant associations were shown (top number: R-value; bottom number: adjusted P-value). (F) The proportion of differentially expressed proteins (DEPs) within a module relative to the total proteins in that module. (G) The proportion of DEPs within a module relative to the total number of DEPs across all modules. Numbers on top of each column represent the DEPs numbers in this module. (H) DEPs defined by three statistical methods: (i) Samr R package was employed with 1000 permutations and a false discovery rate (FDR) threshold of 0.05; (ii) Limmar R package was used with Benjamini-Hochberg (BH) multiple correction; and (iii) Student's t-test and BH multiple correction was applied. (I) Distribution of protein expression ratio of EVs in the turquoise module compared between patients with schizophrenia and healthy controls. ***P < 0.001. (J) Biological network enriched in the turquoise module (protein-protein interaction confidence ≥0.7). Query size=proteins involved in the network. Adjusted P-value <0.05. (K) Network of complement activation in the turquoise module. Dark blue = DEPs defined by three statistical methods; Grey = non-DEPs in the network; circle size = numbers of proteins interacted with the node; line thickness = confidence of protein-protein interaction. SZ, n = 20; HC, n = 28.

with complement system and included it in the subsequent analysis. Later, we will refer to C3, C4A, C4B, C4BPA, C4BPB and PROS1 as complement DEPs.

Characterization of extracellular vesicle-derived complement DEPs in patients with schizophrenia

We found that all of the EV-derived complement DEPs were significantly elevated in patients with schizophrenia compared to controls (Fig. 2A-F). Furthermore, when compared to non-DEPs within the complement network, these complement DEPs exhibited the highest number of significant correlations with the disease course (P < 0.0015, Fig. 2G-I). Additionally, we discovered significant correlations between these complement DEPs and DEPs of NOE1, KIRR3 and LMO4. Notably, such correlation was exclusively observed in schizophrenia patients rather than controls (Fig. 2M). These three DEPs showed predominant mRNA expression in brain tissue according to the Human Protein Atlas database³² (https://

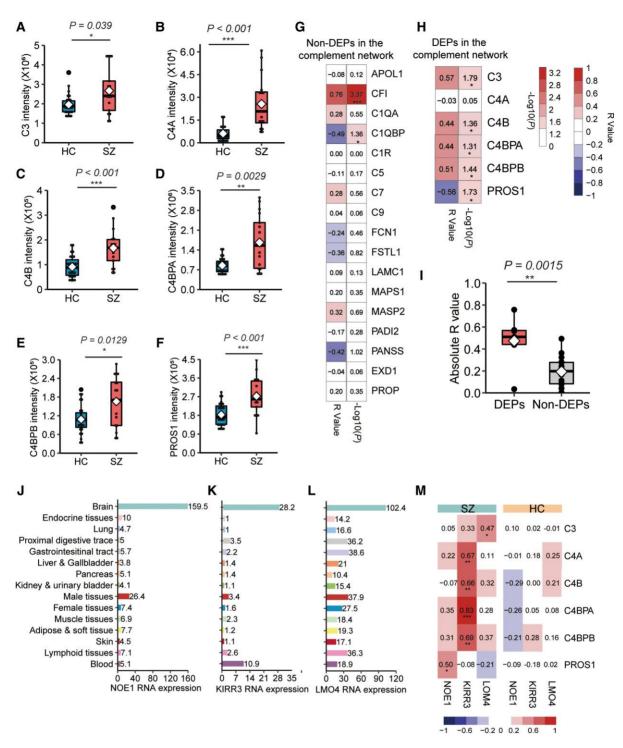


Figure 2 Characteristics of EV-derived complement DEPs in patients with schizophrenia. (A) Comparison of C3 expression between patients with schizophrenia (SZ) and healthy controls (HC). $^*P = 0.039$, Cohen's d = 0.78. (B) Comparison of C4A expression. $^{***P} < 0.001$, Cohen's d = 1.57. (C) Comparison of C4BPA expression. $^{**P} = 0.0029$, Cohen's d = 1.17. (E) Comparison of C4BPA expression. $^{**P} = 0.0029$, Cohen's d = 0.88. (F) Comparison of PROS1 expression. $^{***P} < 0.001$, Cohen's d = 1.20. Although the levels of body mass index (BMI) and diastolic blood pressure (DBP) were significantly different between patients and controls, they did not demonstrate significant correlations with the levels of extracellular vesicle (EV)-derived complement associated components within each independent group. Hence, they were not adjusted as covariates. (G and H) The heat map of correlation analysis between EV-derived non-differentially expressed proteins (DEPs) and DEPs in the complement network and the disease course. Significant correlation was defined as absolute R values ≥ 0.4 and P-value < 0.05 [$-\log 10$ (P) > 1.3]. Pearson or Spearman correlation was used as appropriate. (I) Box plot of the absolute R value of correlation analysis. $^{**P} = 0.0015$, Cohen's d = 1.67. (J-L) Expression levels of NOE1, KIRR3 and LMO4 mRNA in different tissues; their mRNA expression were mostly enriched in brain. Data were extracted from the Human Protein Atlas. (M) Heat map of linear correlation results between the EV-derived complement DEPs and brain enriched proteins of NOE1, KIRR3 and LMO4 in patients and healthy controls. The number in each box and the legend represent the R-value. $^{**P} < 0.05$; $^{**P} < 0.001$; $^{**P} < 0.001$.

www.proteinatlas.org) (Fig. 2J-L). Previous studies have reported the significance of these three DEPs in CNS development. 39-42 Taken together, we hypothesize that these complement DEPs may synergistically participate in regulating CNS development and play crucial roles in the progression of schizophrenia.

Validation of extracellular vesicle-derived complement DEPs in patients with schizophrenia

Next, we validated these EV-derived complement DEPs using MSD technology in another independent set of age- and sex-matched samples (Set 2), comprising 26 patients with schizophrenia and 26 healthy controls (Supplementary Table 1). Notably, C4A and C4B share high sequence similarity (≥95%) with only 10 different residues. Although we could differentiate between the two variants through their unique identified peptides in our MS analysis, there were no available antibodies to distinguish between them. Therefore, during the MSD validation, we measured the total level of C4 instead. To make a comparison with the MSD results, we summed the intensities of C4A and C4B detected by MS as an estimate of the overall C4 level (Fig. 3C). The unavailability of a suitable antibody for C4BPB also precluded its inclusion in analysis. Consequently, our validation focused solely on EV-derived C3, C4, C4BPA and PROS1. As a result, they were all significantly upregulated in patients compared to controls (Fig. 3B, D, F and H), which supports our proteomic discovery (Fig. 3A, C, E and G). Furthermore, to evaluate complement activation in EVs, we used neoepitope antibodies targeting C3a and C4a. We found increased levels of C3a and C4a in patients compared to controls, indicating enhanced complement activation in EVs under the illness (Fig. 3I-L).

Extracellular vesicle-derived complement DEPs outperform their counterparts from plasma as biomarkers

Plasma has been widely used as a convenient source for biomarker development in schizophrenia; however, the results have been inconsistent and inconclusive. 43 In this study, we compared the expression levels of C3, C4, C4BPA and PROS1 in both plasma and EVs between patients with schizophrenia and healthy controls. Our results showed that in plasma, none of these proteins displayed statistically significant or consistent difference across the two sample sets (Fig. 4). Conversely, EV-derived C3, C4, C4BPA and PROS1 exhibited better consistency between sets compared to their plasma counterparts, where they were all significantly overexpressed in patients relative to controls (Fig. 3). Therefore, we suggest that EV-derived complement DEPs outperform their plasma counterparts and represent a superior source of biomarkers for schizophrenia.

Machine learning to develop diagnostic biomarkers for schizophrenia

Next, based on the four EV-derived complement DEPs, we used XGBoost-based machine learning to establish models for classification of schizophrenia. We utilized the aforementioned two samples sets of schizophrenia patients and controls, testing all possible combinations of the four indexes ($\sum_{i=1}^{4} C_4^i = 15$). Five-fold training/ test splits and 15 repeats were applied for cross-validation. We found that combinations of two and more complement DEPs achieved an AUC of ≥84%. Notably, the combination of all four indexes demonstrated the highest AUC [0.895, 95% confidence

interval (CI) = 0.882-0.908], with an accuracy of 83.5%, a sensitivity of 85.3% and a specificity of 82.0% (Fig. 5A and Table 1) for distinguishing between schizophrenia and healthy controls.

Additionally, we assessed the performance of the four indexes in distinguishing FEDN schizophrenia patients and healthy controls in an age- and sex-matched case-control cohort (n = 34 each) (Supplementary Table 2). In this analysis, we did not observe any significant difference in the expression of EV-derived C3, C4, C4BPA and PROS1 between FEDN patients and controls (Supplementary Fig. 3A-E). Consequently, the four indexes were less effective in discriminating between FEDN patients and healthy controls, achieving only a 0.661 AUC, 62.4% accuracy, 64.9% sensitivity and 60.2% specificity (Supplementary Fig. 3F). However, the four indexes performed better in distinguishing between chronic and FEND patients, with a 0.881 AUC, 82.2% accuracy, 85.4% sensitivity and 78.7% specificity (Supplementary Fig. 3G).

Extracellular vesicle-based biomarkers were able to distinguish schizophrenia from bipolar disorder and major depressive disorder

Given the high rate of misdiagnosis of schizophrenia as BD and MDD, we investigated whether the EV-based four indexes could differentiate between these disorders. We recruited participants with BD (n = 35) or MDD (n = 26), and age- and sex- matched healthy controls (n = 35 and 25, respectively) (Supplementary Table 3). The expression of EV-derived C3, C4, C4BPA and PROS1 were measured by MSD. We found that their expression patterns were specific to each disorder (Table 2 and Supplementary Fig. 4). The EV-based four indexes yield a 0.966 AUC, 91.0% accuracy, 91.1% sensitivity and 91.2% specificity for discriminating participants with schizophrenia from those with BD (Fig. 5B and Table 1). When differentiating between schizophrenia and MDD, the four indexes achieved an 0.893 AUC, 84.0% accuracy, 80.5% sensitivity and 86.9% specificity (Fig. 5C and Table 1). Collectively, these data suggest that the EV-based biomarkers could effectively distinguish schizophrenia from BD and MDD.

The Personalized Discrimination Score could diagnose schizophrenia at the individual level

An individualized assessment tool holds significance in clinical practice. We further utilized the XGBoost algorithm to construct a PDS based on the EV-based four indexes for individual diagnosis (schizophrenia versus control). The PDS was assigned polarity in such a way that individuals with a positive score were predicted to belong to the schizophrenia group. We observed that PDS had a significant ability to differentiate individuals with schizophrenia from healthy controls (Fig. 6A). However, it was unable to distinguish healthy individuals from BD (Fig. 6B) or MDD (Fig. 6C). Therefore, we concluded that the discrimination score derived from the four indexes is highly sensitive and specific to schizophrenia, highlighting its potential for individual diagnosis of schizophrenia and differentiation from BD and MDD.

The Personalized Discrimination Score is available for prediction of antipsychotic response at the individual level

Next, to determine whether the PDS is related to the antipsychotic response of schizophrenia patients, we recruited a follow-up cohort that included symptomatic schizophrenia patients (n = 27). We collected clinical assessments and the levels of EV-derived C3, C4,

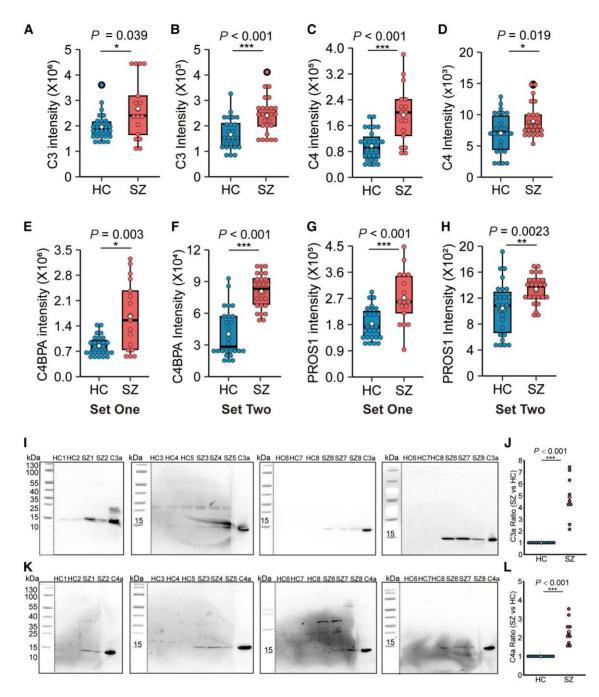


Figure 3 Potential biomarkers validated in EVs in patients with schizophrenia and healthy controls. The concentration of extracellular vesicle (EV)-derived complement differentially expressed proteins were detected by proteomics (Set 1) and Mesoscale Discovery technology (Set 2). (**A** and **B**) C3 amount in EVs. Set 1: $^*P = 0.039$, Cohen's d = 0.78. Set 2: $^*P < 0.001$, Cohen's d = 1.13. (**C** and **D**) C4 amount in EVs. Set 1: $^*P < 0.001$, Cohen's d = 1.36. C4 intensity was the combination of C4A and C4B intensity. Set 2: $^*P = 0.019$, Cohen's d = 0.67. (**E** and **F**) C4BPA amount in EVs. Set 1: $^*P = 0.003$, Cohen's d = 1.17. Set 2: $^*P = 0.001$, Cohen's d = 1.20. Set 2: $^*P = 0.0023$, Cohen's d = 0.90. Although the levels of body mass index or diastolic blood pressure were significantly different between patients (SZ) and controls (HC), they were not adjusted as covariates. (I-L) Western blot analysis of the activation of EV-derived Complement C3 and C4. Neoepitope-specific antibodies of C3a and C4a were used. Human serum purified proteins of C3a des Arg and C4a des Arg were used as the positive controls. The last lane of each membrane was the positive control. n = 11 for each group. SZ = schizophrenia; HC = healthy controls.

C4BPA and PROS1 at baseline and after an average of 3 months of antipsychotic treatment (Supplementary Table 4). We defined treatment response as a PANSS reduction rate of \geq 25% (Fig. 6D). We found that the responders had a significantly higher reduction rate in positive, negative and general psychological scores compared to the non-responders (P < 0.001, Fig. 6E–G). Additionally,

responders had a significantly higher baseline PDS than non-responders (P < 0.001, Fig. 6H). Furthermore, the baseline PDS was significantly associated with the reduction percentage in PANSS total score, as well as positive, negative and general psychological subscales (Fig. 6I–L). However, we did not observe such an association between the baseline PDS and the baseline PANSS total and

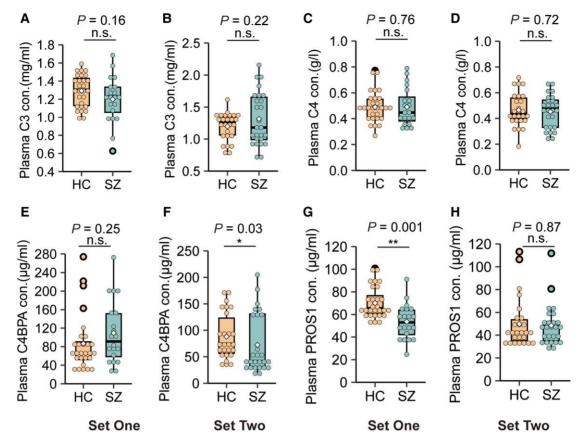


Figure 4 Potential biomarkers validated in plasma in patients with schizophrenia and healthy controls. The concentration of complement differentially expressed proteins in plasma were detected by ELISA. (A and B) Plasma concentration of C3. Set 1: P = 0.16, Cohen's d = 0.43. Set 2: P = 0.22, Cohen's d = 0.35. (C and D) Plasma concentration of C4. Set 1: P = 0.76, Cohen's d = 0.0005. Set 2: P = 0.72, Cohen's d = 0.098. (E and F) Plasma concentration of C4BPA. Set 1: P = 0.25, Cohen's d = 0.35. Set 2: P = 0.03, Cohen's d = 0.02. (G and H) Plasma concentration of PROS1. Set 1: P = 0.001, Cohen's D = 0.001,

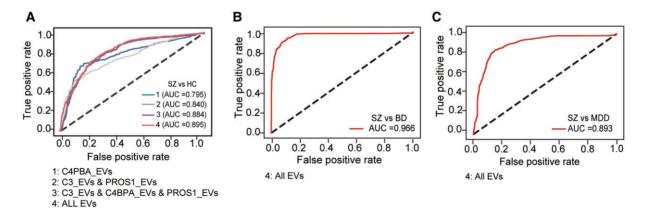


Figure 5 XGBoost-based machine learning to classify patients with schizophrenia from healthy controls, BD and MDD. (A) The area under the curve (AUC) of different combinations of four index biomarkers to discriminate patients with schizophrenia (SZ) and healthy controls (HC). (B and C) The XGBoost-based modelling to discriminate schizophrenia from bipolar disorder (BD) and major depressive disorder (MDD) using extracellular vesicle (EV)-based four indices.

subscale scores (Fig. 6M–P). In summary, the PDS developed from cross-sectional case–control datasets was significantly associated with individual treatment response in a longitude cohort, demonstrating its prognostic ability at an individual level.

Discussion

Through a combination of EV proteome profiling and machine learning, this study has successfully developed a novel EV-based protein biomarker panel, comprised of EV-derived C3, C4, C4PBA

Table 1 The AUC, accuracy, sensitivity, specificity, PPV and NPV of four index biomarkers for different discrimination

	SZ versus HC	SZ versus BD	SZ versus MDD
AUC	0.895	0.966	0.893
Accuracy	83.48	90.99	83.95
Sensitivity	85.32	91.08	80.51
Specificity	81.91	91.19	86.85
PPV	83.73	90.22	82.18
NPV	83.18	91.69	85.24

AUC = area under curve; BD = bipolar disorder; HC = healthy controls; MDD = major depressive disorder; NPV = negative predictive value; PPV = positive predictive value; SZ = schizophrenia.

and PROS1. This panel has proven to be highly effective to diagnose chronic schizophrenia from healthy controls, with an 0.895 AUC, 83.5% accuracy, 85.3% sensitivity and 82.0% specificity. It also accurately differentiates schizophrenia from BD and MDD. Furthermore, we have established a PDS based on the panel, which provides a personalized index for accurately discriminating schizophrenia and predict their treatment response.

In this study, we used a stepwise proteomics analysis to identify potential biomarkers for schizophrenia and conducted a validation in another independent sample set. Encouragingly, all of these EVs-derived biomarkers were successfully validated and showed significant overexpression in patients with schizophrenia. Most of these EV-derived biomarkers belong to the complement system, and our study revealed an increased complement activation mediated by EVs in schizophrenia patients. The complement system plays key roles in bridging innate and adaptive immunity under normal physiological conditions, 44,45 and its overall activity is tightly controlled. In the CNS, the complement system is involved in synaptic pruning, neurogenesis and migration modulation.⁴⁶ Mounting evidence have suggested that the aberrant complement system is a central pathway contributing to immune abnormalities and dysregulated neuronal development in schizophrenia. 43,47,48 Overall, our findings demonstrated the reliability of these EV-based biomarkers and they have biological significance in the pathogenesis of schizophrenia.

Previous studies on complement biomarkers for schizophrenia have primarily focused on plasma proteins, with little attention paid to EVs. However, the inconsistent and inconclusive findings in plasma have raised doubts about the efficacy of these proteins as diagnostic biomarkers.⁴³ Our study has unveiled a similar challenge, as two of the indexes failing to consistently replicate their results in plasma across different sample sets. Plasma proteins are susceptible to protease and enzymatic degradation in the bloodstream, which may account for their instability during preparation, leading to inconsistent findings. In contrast, EV proteins are encapsulated within a lipid membrane, protecting them from degradation and potentially rendering them more stable and resilient. In our study, we consistently observed elevated levels of EV-derived indexes in patients across different sample sets. Additionally, evidence suggest that EVs in the CNS can cross the blood-brain barrier (BBB) and enter the peripheral blood. 49-51 Our established EV-derived biomarkers demonstrate a significant positive relationship with brain-enriched proteins in patients with schizophrenia but not in healthy individuals. This finding suggests that these biomarkers may, to some extent, derived from the CNS in

Table 2 The different expression pattern of extracellular vesicles derived four indexes among SZ, BD and MDD patients compared to matched healthy controls

	SZ versus HC	BD versus HC	MDD versus HC
C3_EVs	Upregulated	Downregulated	Upregulated
C4_EVs	Upregulated	Downregulated	Not significantly changed
C4BPA_EVs	Upregulated	Not significantly	Not significantly changed
		changed	
PROS1_EVs	Upregulated	Downregulated	Not significantly changed

 $BD = bipolar\ disorder;\ EV = extracellular\ vesicle;\ HC = healthy\ controls;\ MDD = major\ depressive\ disorder;\ SZ = schizophrenia.$

pathological conditions, carry valuable information regarding brain pathology in schizophrenia. Similarly, previous studies have demonstrated increased levels of astrocytic-derived EVs in plasma for patients with Alzheimer's disease, ^{52,53} multiple sclerosis ⁵⁴ or early psychosis, ¹⁶ which may contribute to pathological synaptic loss and neuronal damage. Overall, we propose that EV-derived complement DEPs surpass their plasma counterparts in terms of their stability and potential to reflect brain pathology, making them superior candidates for schizophrenia biomarker development.

A clinically valuable biomarker should have the capability to provide informative data at the individual level. For this purpose, we developed PDS based on the EV-derived biomarkers, which enables specific diagnosis of schizophrenia in individual patients. An individual assigned a positive PDS was predicted to have schizophrenia. This scoring strategy is feasible for clinical transformation and holds value in complementing current subjective diagnostic criteria. Moreover, our PDS serves as a valuable tool for predicting antipsychotic treatment response, facilitating guidance of individualized treatment. For patients predicted to have a poor response, more comprehensive therapeutic strategies, such as psychotherapy and transcranial magnetic stimulation, could be implemented at the early stage of treatment. Importantly, we found that the PDS was significantly associated with the percentage reduction in PANSS total and subscale scores, rather than their baseline scores. These findings suggest that the PDS generated from the EV-derived biomarkers possesses robust predictive capability, unaffected by patients' baseline condition.

Our study has several limitations that warrant consideration. First, our findings are primarily based on a young and middle-aged (16-64 years) population, so caution should be exercised when generalizing the results to older age groups. Second, while XGboost has demonstrated its ability to generate reliable results in small sample size, the clinical utility of EV-based biomarkers in predicting and diagnosing schizophrenia still requires large-scale, prospective longitudinal studies that include multicentred cohorts. Third, in the longitudinal schizophrenia samples, all recruited individuals received treatment with second-generation antipsychotics, with 80% receiving monotherapy and the remainder receiving a combination of two antipsychotics. However, due to the limited sample size and the presence of combination therapy, further stratification of patients into specific drug categories was not feasible. Consequently, the potential influences of drug interactions and heterogeneity could not be completely excluded. Therefore, the PDS tool is primarily targeted towards second-generation

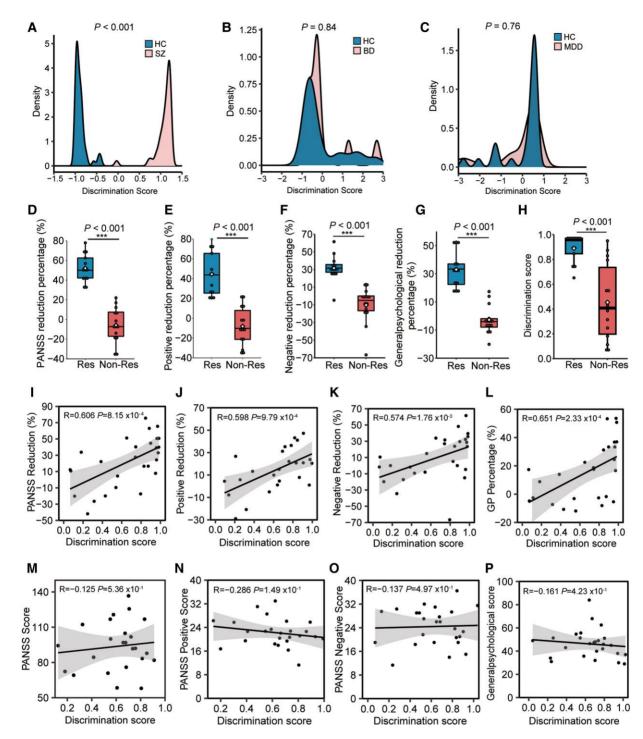


Figure 6 The personalized discrimination score to diagnosis and predict antipsychotic treatment response of patients with schizophrenia. (A–C) The Personalized Discrimination Score developed using the XGBoost-based machine learning strategy could significantly classify schizophrenia (SZ) from matched healthy controls (HC). P < 0.001. However, it could not significantly separate patients with bipolar disorder (BD) or patients with major depressive disorder (MDD) from their matched healthy controls. X-axis: the normalized discrimination score by z-score. P = 0.84, P = 0.76. (D–G) The comparison of reduction percentage of Positive and Negative Syndrome Scale (PANSS) total and subscale scores between the responders (Res) and non-responders (Non-Res). All ***P < 0.001. (H) The comparison of discrimination score between the responders and non-responders. **P < 0.001. (I–L) The Spearman correlation analysis between the baseline PDS and the reduction percentage of PANSS total and subscale scores. Baseline PANSS total and subscale scores were adjusted as the confounding covariates as needed, respectively. (M–P) The Spearman correlation analysis between the baseline PDS and the baseline PANSS total and subscale scores. Responders: n = 14; non-responders: n = 13.

antipsychotic medications, rather than specific drugs. Fourth, patients recruited for the followed-up study were symptomatic patients and the generalization and optimization of the PDS tool in FEDN and acute patients require further investigation.

In summary, we have developed novel EV-based biomarkers and a PDS tool for diagnosing schizophrenia, differentiating it from BD and MDD, and predicting antipsychotic response. The integration of symptom assessment and EV-based biomarkers has the

potential to enhance the accuracy of schizophrenia diagnosis and guide the development of complement-directed therapies.

Data availability

All data are provided within the paper or Supplementary material. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteo mexchange.org) via the iProX partner repository with the dataset identifier PXD040261. The code supporting the current study has not been deposited in a public repository as it does not contain newly generated software or custom code but is available from the corresponding author upon request.

Acknowledgements

We thank Yuanyuan Ma at Shenzhi Department of the Fourth Affiliated Hospital of Xinjiang Medical University, Xiaodan Wang at Shanghai Mental Health Center, and Kunxuan Zhu at Harbin University of Commerce for their scientific support.

Funding

This work was supported by National Key R&D Program of China (2017YFC0909200 to D.H.C.); Shanghai Sailing Program (17YF141630 to T.X.); National Natural Science Foundation of China (82271539, 81801324 to T.X., 82271544 to D.H.C. and 82373937 to J.Y.); Shanghai Natural Science Foundation (21ZR1455300 to T.X.); Shanghai Municipal Health Commission Clinical Specialization in Health Industry (20214Y0057 to T.X.).

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at Brain online.

References

- 1. Rossler W, Salize HJ, van Os J, Riecher-Rossler A. Size of burden of schizophrenia and psychotic disorders. Eur Neuropsychopharmacol.
- 2. Ayano G, Demelash S, Yohannes Z, et al. Misdiagnosis, detection rate, and associated factors of severe psychiatric disorders in specialized psychiatry centers in Ethiopia. Ann Gen Psychiatry.
- 3. Leucht S, Tardy M, Komossa K, et al. Antipsychotic drugs versus placebo for relapse prevention in schizophrenia: A systematic review and meta-analysis. Lancet. 2012;379:2063-2071.
- 4. Colombo M, Raposo G, Thery C. Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. Annu Rev Cell Dev Biol. 2014;30:255-289.
- 5. Yanez-Mo M, Siljander PR, Andreu Z, et al. Biological properties of extracellular vesicles and their physiological functions. J Extracell Vesicles. 2015;4:27066.
- 6. Xiao Y, Wang S-K, Zhang Y, et al. Role of extracellular vesicles in neurodegenerative diseases. Prog Neurobiol. 2021;201:102022.

- 7. Fevrier B, Vilette D, Archer F, et al. Cells release prions in association with exosomes. Proc Natl Acad Sci U S A. 2004:101: 9683-9688.
- 8. Guo M, Wang J, Zhao Y, et al. Microglial exosomes facilitate alpha-synuclein transmission in Parkinson's disease. Brain. 2020:143:1476-1497.
- Muraoka S, DeLeo AM, Sethi MK, et al. Proteomic and biological profiling of extracellular vesicles from Alzheimer's disease human brain tissues. Alzheimers Dement. 2020;16:896-907.
- 10. Ruan Z, Pathak D, Venkatesan Kalavai S, et al. Alzheimer's disease brain-derived extracellular vesicles spread tau pathology in interneurons. Brain. 2021;144:288-309.
- 11. Aubertin K, Silva AK, Luciani N, et al. Massive release of extracellular vesicles from cancer cells after photodynamic treatment or chemotherapy. Sci Rep. 2016;6:35376.
- 12. van Dommelen SM, van der Meel R, van Solinge WW, Coimbra M, Vader P, Schiffelers RM. Cetuximab treatment alters the content of extracellular vesicles released from tumor cells. Nanomedicine (Lond). 2016;11:881-890.
- 13. Dutta S, Hornung S, Kruayatidee A, et al. α--Synuclein in blood exosomes immunoprecipitated using neuronal and oligodendroglial markers distinguishes Parkinson's disease from multiple system atrophy. Acta Neuropathol. 2021;142:495-511.
- 14. Fiandaca MS, Kapogiannis D, Mapstone M, et al. Identification of preclinical Alzheimer's disease by a profile of pathogenic proteins in neurally derived blood exosomes: A case-control study. Alzheimers Dement. 2015;11:600-607.e1.
- 15. Kapogiannis D, Mustapic M, Shardell MD, et al. Association of extracellular vesicle biomarkers with Alzheimer disease in the Baltimore longitudinal study of aging. JAMA Neurol. 2019;76:
- 16. Goetzl EJ, Srihari VH, Guloksuz S, Ferrara M, Tek C, Heninger GR. Decreased mitochondrial electron transport proteins and increased complement mediators in plasma neural-derived exosomes of early psychosis. Transl Psychiatry. 2020;10:361.
- 17. Goetzl EJ, Srihari VH, Guloksuz S, Ferrara M, Tek C, Heninger GR. Neural cell-derived plasma exosome protein abnormalities implicate mitochondrial impairment in first episodes of psychosis. FASEB J. 2021;35:e21339.
- 18. Kapogiannis D, Dobrowolny H, Tran J, et al. Insulin-signaling abnormalities in drug-naive first-episode schizophrenia: Transduction protein analyses in extracellular vesicles of putative neuronal origin. Eur Psychiatry. 2019;62:124-129.
- 19. Chen TQ, Guestrin C. XGBoost: A Scalable Tree Boosting System. KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery; 2016:785-794.
- 20. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? Schizophr Res. 2005;79:231-238.
- 21. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull. 1987;13: 261-276.
- 22. Lobb RJ, Becker M, Wen SW, et al. Optimized exosome isolation protocol for cell culture supernatant and human plasma. J Extracell Vesicles. 2015;4:27031.
- 23. Thery C, Amigorena S, Raposo G, Clayton A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr Protoc Cell Biol. 2006; Chapter 3: Unit 3.22.
- 24. Kulak NA, Pichler G, Paron I, Nagaraj N, Mann M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods. 2014;11:319-324.
- 25. Siems SB, Jahn O, Eichel MA, et al. Proteome profile of peripheral myelin in healthy mice and in a neuropathy model. eLife. 2020;9: e51406.

- 26. Ahrne E, Molzahn L, Glatter T, Schmidt A. Critical assessment of proteome-wide label-free absolute abundance estimation strategies. Proteomics. 2013;13:2567-2578.
- 27. Silva JC, Gorenstein MV, Li G-Z, Vissers JPC, Geromanos SJ. Absolute quantification of proteins by LCMSE—A virtue of parallel MS acquisition. Mol Cell Proteomics. 2006:5:144-156.
- 28. Gao Q, Zhu HW, Dong LQ, et al. Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. Cell. 2019;179:561-577.e22.
- 29. Raudvere U, Kolberg L, Kuzmin I, et al. G:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019;47:W191-W198.
- 30. Szklarczyk D, Gable AL, Lyon D, et al. STRING V11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47:D607-D613.
- 31. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498-2504.
- 32. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledgebased human protein atlas. Nat Biotechnol. 2010;28:1248-1250.
- 33. Zhang B, Horvath S. A general framework for weighted gene coexpression network analysis. Stat Appl Genet Mol Biol. 2005;4:
- 34. Carlsson S, Dahlback B. Dependence on vitamin K-dependent protein S for eukaryotic cell secretion of the beta-chain of C4b-binding protein. J Biol Chem. 2010;285:32038-32046.
- 35. Furmaniak-Kazmierczak E, Hu CY, Esmon CT. Protein S enhances C4b binding protein interaction with neutrophils. Blood. 1993;81:405-411.
- 36. Kask L, Trouw LA, Dahlback B, Blom AM. The C4b-binding protein-protein S complex inhibits the phagocytosis of apoptotic cells. J Biol Chem. 2004;279:23869-23873.
- 37. Trouw LA, Nilsson SC, Goncalves I, Landberg G, Blom AM. C4b-binding protein binds to necrotic cells and DNA, limiting DNA release and inhibiting complement activation. J Exp Med. 2005;201:1937-1948.
- 38. Webb JH, Blom AM, Dahlback B. Vitamin K-dependent protein S localizing complement regulator C4b-binding protein to the surface of apoptotic cells. J Immunol. 2002;169:2580-2586.
- 39. Martin EA, Muralidhar S, Wang Z, et al. The intellectual disability gene Kirrel3 regulates target-specific mossy fiber synapse development in the hippocampus. Elife. 2015;4:e09395.
- 40. Taylor MR, Martin EA, Sinnen B, et al. Kirrel3-mediated synapse formation is attenuated by disease-associated missense variants. J Neurosci. 2020;40:5376-5388.

- 41. Moreno TA, Bronner-Fraser M. Noelins modulate the timing of neuronal differentiation during development. Dev Biol. 2005; 288:434-447.
- 42. Zaman T. Zhou X. Pandev NR, et al. LMO4 is essential for paraventricular hypothalamic neuronal activity and calcium channel expression to prevent hyperphagia. J Neurosci. 2014;34: 140-148.
- 43. Woo JJ, Pouget JG, Zai CC, Kennedy JL. The complement system in schizophrenia: Where are we now and what's next? Mol Psychiatry. 2020;25:114-130.
- 44. Killick J, Morisse G, Sieger D, Astier AL. Complement as a regulator of adaptive immunity. Semin Immunopathol. 2018;40:
- 45. West EE, Kolev M, Kemper C. Complement and the regulation of T cell responses. Annu Rev Immunol. 2018;36: 309-338
- 46. Dalakas MC, Alexopoulos H, Spaeth PJ. Complement in neurological disorders and emerging complement-targeted therapeutics. Nat Rev Neurol. 2020;16:601-617.
- 47. Beasley C, Shao L. Increased expression of early complement components in frontal cortex in schizophrenia. Schizophrenia Bull. 2017;43:S73.
- 48. Sekar A, Bialas AR, de Rivera H, et al. Schizophrenia risk from complex variation of complement component 4. Nature. 2016;
- 49. Morad G, Carman CV, Hagedorn EJ, et al. Tumor-derived extracellular vesicles breach the intact blood-brain barrier via transcytosis. ACS Nano. 2019;13:13853-13865.
- 50. Saint-Pol J, Gosselet F, Duban-Deweer S, Pottiez G, Karamanos Y. Targeting and crossing the blood-brain barrier with extracellular vesicles. Cells-Basel. 2020;9:851.
- 51. Tominaga N, Kosaka N, Ono M, et al. Brain metastatic cancer cells release microRNA-181c-containing extracellular vesicles capable of destructing blood-brain barrier. Nat Commun. 2015;
- 52. Goetzl EJ, Schwartz JB, Abner EL, Jicha GA, Kapogiannis D. High complement levels in astrocyte-derived exosomes of Alzheimer disease. Ann Neurol. 2018;83:544-552.
- 53. Nogueras-Ortiz CJ, Mahairaki V, Delgado-Peraza F, et al. Astrocyte- and neuron-derived extracellular vesicles from Alzheimer's disease patients effect complement-mediated neurotoxicity. Cells-Basel. 2020;9:1618.
- 54. Bhargava P, Nogueras-Ortiz C, Kim S, Delgado-Peraza F, Calabresi PA, Kapogiannis D. Synaptic and complement markers in extracellular vesicles in multiple sclerosis. Mult Scler. 2021;27:509-518.