ARTICLE   OPEN

Check for updates

Proteomic insights into mental health status: plasma markers in young adults

Alexey M. Afonin[1], Aino-Kaisa Piironen [1], Izaque de Sousa Maciel [1], Mariia Ivanova[1], Arto Alatalo[1], Alyce M. Whipp [2], Lea Pulkkinen[3], Richard J. Rose[4], Irene van Kamp[5], Jaakko Kaprio [2,6] and Katja M. Kanninen [1✉]

© The Author(s) 2024

Global emphasis on enhancing prevention and treatment strategies necessitates an increased understanding of the biological mechanisms of psychopathology. Plasma proteomics is a powerful tool that has been applied in the context of specific mental disorders for biomarker identification. The p-factor, also known as the "general psychopathology factor", is a concept in psychopathology suggesting that there is a common underlying factor that contributes to the development of various forms of mental disorders. It has been proposed that the p-factor can be used to understand the overall mental health status of an individual. Here, we aimed to discover plasma proteins associated with the p-factor in 775 young adults in the FinnTwin12 cohort. Using liquid chromatography–tandem mass spectrometry, 13 proteins with a significant connection with the p-factor were identified, 8 of which were linked to epidermal growth factor receptor (EGFR) signaling. This exploratory study provides new insight into biological alterations associated with mental health status in young adults.

INTRODUCTION

Mental health issues are increasingly becoming a major concern globally [1, 2]. In fact, the World Health Organization estimates that about one in every eight people across the globe suffer from a mental health disorder, making these disorders the primary cause of a reduced quality of life [1]. The recent COVID-19 pandemic has notably exacerbated mental health issues, particularly among young adults aged 18–29. During the pandemic, the number of young adults experiencing depression symptoms more than doubled in numerous European countries [3].

Despite the significant impact of mental health diseases on daily life and their considerable economic cost, these conditions often go undiagnosed and untreated [1, 3]. This highlights the pressing need for early detection of individuals at high risk for psychopathology, targeted preventative measures, and improvements in diagnostic procedures and treatments.

Although different mental disorders may have unique symptoms, they have been shown to share commonalities in terms of underlying biological, psychological, and social factors [4]. The p-factor, also known as the "general psychopathology factor," is a concept in psychopathology suggesting that there is a common underlying factor that contributes to the development of various forms of mental disorders [5, 6]. It has been proposed that this single latent factor can encapsulate individuals' proclivity to develop all forms of psychopathology included within the broad internalizing, externalizing, and thought disorder dimensions [7]. The p-factor is analogous to the general factor in intelligence (called the g-factor), which

summarizes the observation that individuals who do well on one type of cognitive test tend to do well on all other types of cognitive tests [5, 8]. Other factors, such as a general factor of personality (GFP) and a general factor of personality disorder (g-PD), have been previously shown to have a high correlation with the p-factor [9]. At the individual level, the p-factor reflects meaningful differences between persons on a single dimension that represents the tendency to experience psychiatric problems as persistent and comorbid; that is, high p-factor individuals experience difficulties in regulation/control when dealing with others, the environment, and the self [4, 5, 10].

Previous studies have shown the p-factor to be connected to brain functioning in adolescents, with higher p-factor scores associated with diminished activation of multiple brain zones during executive tasks [11]. Importantly, some studies have reported that the p-factor may be a stronger predictor of mental health outcomes than specific diagnoses of mental disorders [12]. A recent study showed that the p-factor was associated with poorer performance on the simple reaction time task and the inspection time task, with speed of processing being a common correlate of psychopathology factors [13]. Likewise, Pulkkinen [14] has shown that low emotion and behavior regulation observed as externalizing and internalizing problems in children are negatively associated with the executive functions of the forebrain for inhibition and updating (containing working memory and shifting). This suggests that the p-factor could be used to better understand the overall mental health status of an individual, rather than just focusing on individual diagnoses.

[1]A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland. [2]Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland. [3]Department of Psychology, University of Jyvaskyla, Jyvaskyla, Finland. [4]Department of Psychological & Brain Sciences, Indiana University, Bloomington, IN, USA. [5]Centre for Sustainability, Environment and Health, National Institute for Public Health and the Environment, Bilthoven, the Netherlands. [6]Department of Public Health, University of Helsinki, Helsinki, Finland. ✉email: katja.kanninen@uef.fi

Biomarker discovery has gained traction in recent years as researchers seek to uncover the biological underpinnings of mental health conditions [15, 16]. The development of "omics" technologies and state-of-the-art analytical methods have increased interest in the capabilities of plasma proteomics in biomarker discovery. LC-MS/MS-based proteomics provides a global snapshot of protein expression patterns that reflect physiological and pathological states [17], making comprehensive analysis of the plasma proteome possible [18]. This has enabled the simultaneous detection and quantification of thousands of proteins, expediting biomarker discovery efforts and reducing the time and resources required for this process. This holistic view of proteomics allows for the unbiased discovery of novel biomarkers, with less need for prior knowledge of target proteins. This feature is particularly important in cases where the biology of the process is not yet fully understood or when new, unforeseen biomarkers are needed for improved diagnostic or prognostic applications [19].

Proteomics approaches have been utilized for the identification of protein signatures associated with specific psychological disorders [20–22]. For example, several growth factors (BDNF, VEGF, NGF) and cytokines (IL-1β, IL-6, IFN-α) have been linked to depression [23]. Moreover, a recent multi-omics study reported reduced apolipoprotein levels and an increase in complement effector proteins in the plasma of schizophrenia (SCZ) patients [24]. However, proteomics analyses have not previously been combined to studies of the p-factor for identification of markers associated with overall mental health status.

The FinnTwin12 (FT12) cohort, a longitudinal study of Finnish twins born between 1983 and 1987, has a multitude of data and biological samples [25, 26]. As a valuable resource for exploring biological processes involved in mental health problems, we explored the connection between the p-factor, previously reported in [27], and plasma proteomics among young adults from this cohort.

METHODS

Cohort description
The FT12 cohort is a longitudinal population-based cohort of Finnish twins born 1983–1987 collected to investigate behavioral development and health habits [25, 26]. Initially, twins and their families were identified using the Finnish Central Population Registry, and questionnaire collection occurred for all participants in the cohort at ages 11/12, 14, 17, and 22. The baseline response rate was 87% ($N = 5600$ twins) and has remained high (response rate range: 85–90%). At age 14, a subset of the twins (from 1035 families) was more intensively studied, including psychiatric interviews and additional questionnaires (ages 14 and 22), as well as blood plasma samples (age 22). The "age 22" assessment wave of these more intensively studied twins involved 1347 individuals (mean age = 22.4 years, SD = 0.70; response rate 73.0%), 779 of whom attended in-person assessments and provided venous blood plasma samples. The blood samples were collected after overnight fasting, which involved abstaining from alcohol and tobacco since the night before sampling. Plasma was immediately extracted and stored at −80 °C [27].

p-factor calculation
In FT12, behavioral and emotional characteristics were measured at all data collection waves. The modified Multidimensional Peer Nomination Inventory (MPNI) measure aimed at observing individual differences in emotion and behavior regulation was used. It is an extension of the measure [28] used in the Jyväskylä Longitudinal Study of Personality and Social Development in which the development of the same individuals has been followed from age 8–50, with findings that low self-regulation is associated with social and psychological dysfunction [14]. The MPNI scale has been previously factor analyzed with three main factors termed Behavior Problems, Emotional Problems, and Adjustment [28].

The MPNI was collected in FT12 at ages 12, 14, and 17, from different raters (7 in total): parents (age 12), teachers (age 12 and 14), twin children themselves (age 14 and 17), and the child's co-twin (age 14 and 17). The measure includes subscales of the externalizing problem dimension:

aggression (6 items [for MPNI ages 12, 14, 17]), hyperactivity-impulsivity (7 items [MPNI ages 12, 14], 6 items [MPNI ages 17]), and inattention (4 items [MPNI ages 12, 14, 17]), as well as for the internalizing problem dimension; depression (5 items [MPNI ages 12, 14], 2 items [MPNI ages 17]), social anxiety (2 items [MPNI ages 12, 14], 3 items [MPNI ages 17]), and 1 item for victimization (MPNI ages 12, 14, 17). Each MPNI item (e.g., "Is restless, unable to sit still") has four response choices (from "not observed in the child" to "clearly observable in the child", scored 0–3 respectively). The MPNI p-factor score was created by combining all the items of the "externalizing" and "internalizing" dimensions together into a sum score, with at most 2 missing items allowed. Missing items were imputed based on the mean of the remaining items, with less than 3% of twins having missing items. A composite "combined" p-factor score was created using the p-factor scores of all seven of the abovementioned available MPNI ratings (Cronbach's alpha=0.76), because we know that ratings from different raters are not highly correlated, however, they can impart unique information [29–31]. Each of the seven scores were standardized as z scores, and then we took the mean of available scores. The p-factor for the FT12 cohort was previously created and analyzed in relation to metabolites in [27]. Eleven twins had no overall p-factor score, leaving 775 twins. Of them, 505 (65%) had been scored by all raters at all times, while 194 (25%) had only one rater value missing, the remaining 10% having scores from 2–4 raters. To examine the dimensionality of combining the seven individual p-factor scores, we performed a factor analysis on the subset of participants who had been rated on all measures. The factor analysis indicated one major factor, with the first eigenvalue associated with the first factor having a value well over one. A correlation analysis was performed for the newly calculated score with the sum scale based on all seven p-factors, showing a high correlation coefficient of 0.983. The composite "combined" p-factor score was used for the subsequent analyses.

The predictive power of the p-factor was tested using the data on the psychiatric interviews of the twins at age 22. Using a logistic regression model for MDD, p-factor score, adjusted for sex, predicts MDD reasonably well with an area under the receiver operating characteristic curve (ROC AUC) of 0.67.

High-abundance protein depletion
Albumin accounts for 50%, and the top 22 proteins account for 99% of plasma proteins by weight in human plasma samples [19]. Therefore, the depletion of high-abundant proteins is essential to the identification and analysis of low-abundant proteins. A commercial kit (High Select™ Top14 Abundant Protein Depletion Mini Spin Columns, cat. Number: A36370, ThermoScientific) was used to deplete the 14 most abundant proteins from plasma before the proteomic analyses. The depleted proteins were human serum albumin (HSA), albumin, IgG, IgA, IgM, IgD, IgE, kappa and lambda light chains, alpha-1-acidglycoprotein, alpha-1-antitrypsin, alpha-2-macro-globulin, apolipoprotein A1, fibrinogen, haptoglobin, and transferrin, according to manufacturer's manual. Briefly, 10 μL of total plasma was added to the mini spin columns and incubated for 10 min while rotating, followed by centrifugation of the columns ($1,000 \times g$) for 2 min. The filtrate was collected in 2 ml plastic tubes and stored at −20 °C until preparation for mass spectrometry proteomic analyses, which were performed at the Turku Proteomics Facility in Finland supported by Biocenter Finland.

Protein precipitation and digestion
The proteins of 786 depleted plasma samples were acetone precipitated and subjected to in-solution digestion according to standard protocol at the Turku Proteomics Facility, Turku, Finland (https://bioscience.fi/). After digestion, peptides were desalted with a Sep-Pak C18 96-well plate (Waters), evaporated to dryness, and stored at −20 °C.

Mass spectrometry analysis
Digested peptide samples were dissolved in 0.1% formic acid, and the peptide concentration was determined with a NanoDrop device. For data-independent acquisition (DIA) analysis, 500 ng of peptides were injected and analyzed in a random order, determined with the Excel rand() function. Wash runs were submitted between each sample to reduce potential carry-over of peptides. The Liquid Chromatography-Electrospray Ionization-Mass Spectrometry (LC-ESI-MS/MS) analysis was performed on a nanoflow HPLC system (Easy-nLC1000, Thermo Fisher Scientific) coupled to a Q Exactive HF mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with a nano-electrospray ionization source. Peptides were first loaded on a
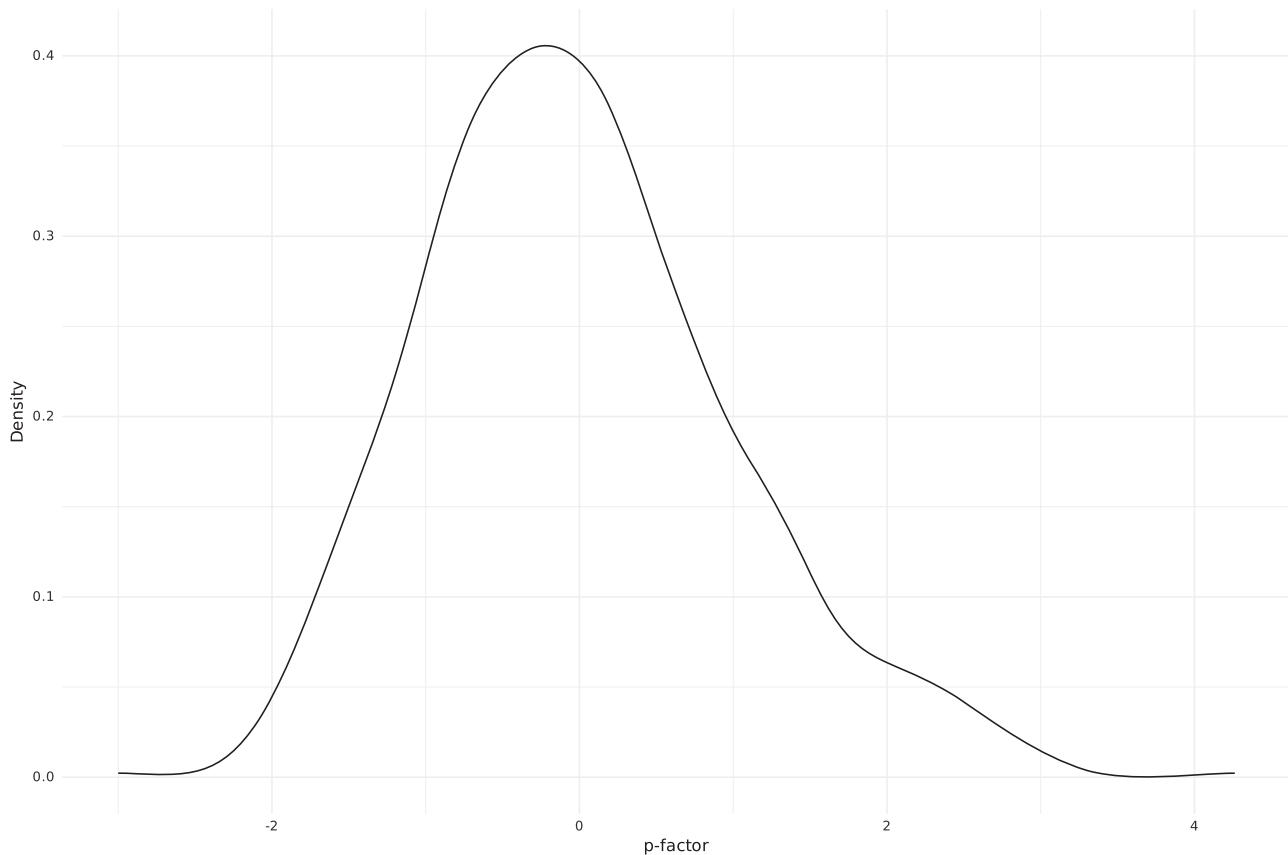
Fig. 1 Distribution of p-factor values in the FT12 cohort. The density distribution of the standardized p-factor values of the participants.

trapping column and subsequently separated inline on a 15 cm C18 column (75 μm x 15 cm, ReproSilPur 3 μm 120 Å C18-AQ, Dr. Maisch HPLC GmbH, Ammerbuch-Entringen, Germany). The mobile phase consisted of water with 0.1% formic acid (solvent A) or acetonitrile/water (80:20 (v/v)) with 0.1% formic acid (solvent B). A 50 min from 5% to 35% solvent B gradient was used to elute peptides. Samples were analyzed by a DIA LC-MS/MS method. MS data was acquired automatically by using Thermo Xcalibur 4.1 software (Thermo Fisher Scientific). In the DIA method, a duty cycle contained one full scan (400–1000 m/z) and 25 DIA MS/MS scans covering the mass range 400–1000 with variable width isolation windows.

Protein identification and quantification analysis
Data analysis consisted of protein identification and label-free quantifications of protein abundances. Data was analyzed using Spectronaut software (Biognosys; version 17.0.2). The direct DIA approach was used to identify proteins. The main data analysis parameters in Spectronaut were: (i) Enzyme: Trypsin/P; (ii) up to 2 missed cleavages; (iii) Fixed modifications: Carbamidomethyl (cysteine); (iv) Variable modifications: Acetyl (protein N-terminus) and oxidation (methionine); (v) Precursor FDR Cutoff: 0.01; (vi) Protein FDR Cutoff: 0.01; (vii) Quantification MS level: MS2; (viii) Quantification type: Area under the curve within integration boundaries for each targeted ion; (ix) Protein database: Homo sapiens Swiss-Prot reference proteome (Uniprot release 2022_01_07_HUMAN) [32], Universal Protein Contaminant database [33].

Raw data drift and batch correction
Protein abundances were analyzed by LC-MS/MS in three separate experimental runs or batches. Since the number of samples in each batch was relatively large, the data was normalized before further analysis, and batch effects were removed. For the ease of comparing the LC-MS/MS runs, 10 of the samples were analyzed in 2 out of the 3 runs. For the raw data analysis, we extracted the data from the .sne file using the iq export scheme [34]. The data used for normalization was the raw peak area of the peptide groups. These values were used in the further analyses.

The raw data investigation pre-processing and statistical analyses were performed in the R (version 4.2.1.) environment (R Core Team, 2022). The signal drift and the observed batch effect were corrected using the proBatch (v. 1.13.0) [35] package. The median abundance plots showed the samples forming four distinct groups, identical to the batches of instrument runs (Supplementary Fig. 1A). The figure also shows pronounced signal drift in the third and fourth batches. These effects were corrected for using the proBatch pipeline. After the batch effect correction, no significant drift or batch effect could be seen (Supplementary Fig. 1B).

Bioinformatic analysis
After drift and batch correction, the fastMaxLFQ method from the iq package (v. 1.9.10) [34] was used to transform the peptide abundancies into protein abundance values. The contaminant proteins and the proteins depleted in the sample pre-processing step were removed from the analysis. Only the identified proteins with quantified abundance levels in at least 80% of the samples were used in further analyses. Missing values remaining in the dataset were imputed using the Sample Minimum method [36]. As an additional sensitivity test, the same modeling was performed using the proteins present in 20, 40, 60, 80, and 98% of the samples, to ensure that the modeling was robust, and that the exclusion of rare proteins did not skew the analysis.

The connection between the p-factor and the protein abundances was analysed using the limma [37] package (v. 3.54.2). Sex and age were included into linear models as covariates to ensure reported associations were not due to sex or age effects. Limma modeling was used to investigate the association of protein abundance with the p-factor using linear and non-linear modeling. The possible non-linear relationship between the p-factor and the protein abundance was investigated by using splines in limma [38]. A basis matrix for representing the family of piecewise-cubic splines with 5 nodes were generated using the ns function from the p-factor variable (Splines package v3.6.2), and was used in limma modeling, also including sex and age as covariates.

Table 1. The plasma proteins significantly associated with the p-factor.

Protein ID	Gene name	Protein name	Linear effect size	Linear p value	Linear q value	Non-linear p value	Non-linear q value
Proteins with linear relationship with p-factor							
Q15828*	CST6	Cystatin E/M	−0.086	1.76E-04	0.024	8.43E-04	0.047
P98160	HSPG2	Heparan sulfate proteoglycan 2	−0.061	3.86E-05	0.006	1.27E-03	0.066
P23142	FBLN1	Fibulin-1	−0.065	2.49E-05	0.006	1.74E-03	0.075
P07911*	UMOD	Uromodulin	−0.145	1.19E-05	0.006	8.89E-05	0.013
Q9Y6R7*	FCGBP	Fc gamma binding protein	−0.082	8.11E-06	0.005	3.05E-05	0.007
Proteins with non-linear relationship with p-factor							
P04179	SOD2	Superoxide dismutase 2	–	3.90E-01	0.701	1.86E-06	0.001
Q9BY67	CADM1	Cell adhesion molecule 1	–	3.87E-02	0.329	4.24E-06	0.002
Q14126	DSG2	Desmoglein 2	–	8.53E-04	0.072	8.80E-05	0.010
Q8NBJ4	GOLM1	Golgi membrane protein 1	–	7.93E-01	0.866	2.68E-04	0.010
P07858	CTSB	Cathepsin B	–	1.22E-01	0.496	6.74E-05	0.010
Q86UN3	RTN4RL2	Reticulon-4 receptor-like 2	–	3.66E-01	0.675	8.59E-05	0.012
O75636	FCN3	Ficolin 3	–	9.16E-03	0.183	3.42E-04	0.026
P07942	LAMB1	Laminin subunit beta-1	–	9.11E-03	0.183	5.13E-04	0.035

*indicate proteins with both linear and non-linear relationship.

Moderate *F* test on the p-factor was carried out to assess the significance of non-linear associations of the protein abundance with the p-factor using the function lmFit and eBayes from the R limma package. *P* values for linear and non-linear modeling were corrected for multiple testing and the false discovery rate (FDR) was computed by using the Benjamini & Hochberg method [39], which were reported as q-values. The significance level considered in all analyses was 0.05. The linear effect size is reported as the log2-fold-change in expression that results from a unit (one standard deviation) change in p-factor.

The protein–protein interaction information for the significantly differentially abundant proteins was analysed using STRING database (v. 12.0) [40]. The enrichment analysis with Gene Ontology (Process, Function, and Component), KEGG and Reactome pathways, PubMed publications, UniProt Keywords, and PFAM/INTERPRO/SMART domain databases was performed using the STRINGdb package [41]. Result visualizations were performed using R and ggplot2 (v3.4.0) [42].

RESULTS

Cohort characteristics
The p-factor was calculated based on assessments by multiple raters at three different ages as described in the Materials and Methods. A combined p-factor value was available for 775 individuals (318 males and 457 females). The z-score-based p-factor distribution is presented in Fig. 1.

Protein identification
MS-based proteomics successfully identified 1494 proteins (DIA spectrometry intensity values) in the FT12 cohort ($N = 775$) the list of found proteins is presented in Supplementary File 1. After the removal of values of proteins depleted in the sample pre-processing step, 1415 proteins were left, with a mean number of identified proteins of 835 per sample (SD = 48). Proteins present in at least 80% of the samples were used, leaving 571 proteins.

Association of proteins with p-factor
The linear modeling showed 5 proteins inversely associated with the p-factor (Table 1). As the relationship between the altered proteins and p-factor is not known, the analysis was also performed using splines, which also made it possible to investigate non-linear relationships between the protein

abundance and the p-factor. These analyses showed 14 proteins associated with the p-factor (Table 1). The relationships between the p-factor and the protein abundance for the significantly associated proteins are presented in Supplementary Figure 2.

The sensitivity testing showed that 13 of 14 proteins were consistent across models, the S100 calcium binding protein A4 was significantly associated with the p-factor only when proteins missing in over 20% of samples were excluded. The results of additional analyses are presented in Supplementary Table 1.

Functional enrichment and annotation
The STRING protein–protein interaction networks functional enrichment analysis showed two connected clusters of proteins with: the first being cystatin-M (CST6) and cathepsin B (CTSB), the second containing laminin subunit beta-1 (LAMB1), basement membrane-specific heparan sulfate proteoglycan core protein (HSPG2), and fibulin-1 (FBLN1) (Fig. 2). Investigation of the first layer of the string network showed that nine of the significant proteins were linked specifically through the epidermal growth factor receptor (EGFR) and transthyretin (TTR) (Fig. 2). Both proteins were among the 636 proteins we investigated, though the q-values were above the significance threshold (q values for both EGFR and TTR were 0.066).

Enrichment analysis of function categories showed only the extracellular matrix structural constituent to be significantly enriched. Compartments, component, function, and tissue analyses showed significantly enriched terms, mostly connected to extracellular space and matrix, and cell–cell adhesion (Supplementary Table 2).

A connection to a disease of the CNS or other neurodegenerative diseases according to the Disease Ontology database was found for 6 of the 13 significant proteins [43], shown in Table 2.

DISCUSSION

The field of plasma proteomics is rapidly gaining traction in the realm of biomedical research, particularly in studies relating to mental health. There is increasing evidence that alterations in plasma protein profiles are associated with major psychiatric
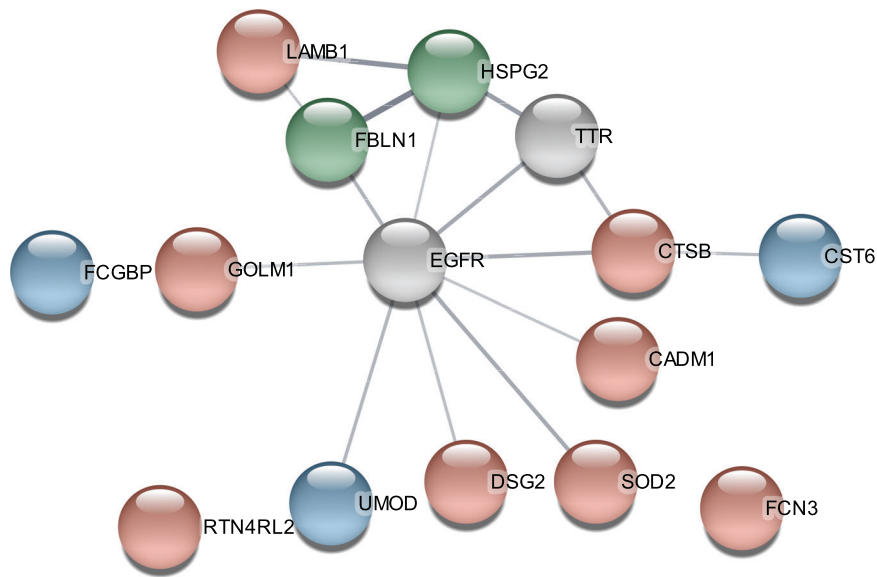
Fig. 2   The result of STRING analysis of the proteins significantly associated with the p-factor. Line thickness indicates the strength of data support. Green circles denote proteins with a linear relationship to the p-factor, red circles with non-linear relationships, and blue circles representing proteins with reported linear and non-linear relationships.

Table 2.   The disease ontology classification (STRINGdb).

Gene name	Disease ontology
LAMB1, FCN3, CTSB, HSPG2	Nervous system disease
FCN3, CTSB, HSPG2, SOD2	Neurodegenerative disease
FCN3, CTSB, HSPG2	Central nervous system disease

conditions, including major depressive disorder (MDD), schizophrenia, psychotic disorders (PSD), and bipolar disorders (BD) [44–46]. This study presents the first report of differences in plasma protein levels associated with the p-factor in a population sample of young adults.

We found 13 plasma proteins associated with p-factor scores in young adults. All but the FCGBP protein were present in the Human Plasma Proteome Database [47, 48], FCGBP, however, was previously reported in serum samples [49]. Of these proteins, ten belonged to a protein network connected to EGFR, eight being directly connected to EGFR. The EGF-related signaling pathways have been previously linked to neurodevelopment [50], synaptic plasticity [51, 52], chronic pain [53], fear [54], as well as mental health diseases [44, 52, 55–58]. For example, altered EGFR signaling has been reported in MDD and BD patients in blood proteomics studies [44].

In addition to EGFR signaling, we observed the p-factor to be negatively associated with heparan sulfate proteoglycan 2 (HSPG2). Heparan sulfate proteoglycans (HSPG) are membrane proteins and a major component of extracellular matrices involved in many cellular processes, as they function as co-receptors for growth factors [59]. HSPG2, combined with CEP350 and SMAD5, was recently presented as a potential diagnostic biomarker for MDD [60]. Furthermore, the *HSPG2* gene was previously connected to antipsychotic-induced adverse effects such as tardive dyskinesia [61, 62], specifically in SCZ patients [63], and the maintenance and repair of the blood-brain barrier in mice [64]. Moreover, downregulation of *HSPG2* and a depressive-like phenotype were revealed in mouse models of chronic mild stress and impaired glutamate function [65].

We also report a negative association with fibulin-1 (FBLN1) and the p-factor. The FBLN1 gene is connected to central nervous system development [66, 67] and modulation of neurotrophic activities of amyloid precursor protein in cultured rodent neural stem cells [68]. So far, little is known about the possible connection of FBLN1 to mental health. Shin [69]. reported decreased FBLN1 plasma protein levels in MDD patients compared to BD patients and healthy controls. The model proposed in that study also contained Fc gamma binding protein (FCGBP), which was reported to be significantly higher in BD patients compared to MDD patients but not in healthy controls. In our study, FCGBP was negatively associated with the p-factor. Additionally, increased plasma protein abundance of desmoglein 3 (DSG3) was reported in MDD patients and reduced abundance in BD patients compared to healthy controls [69]. DSG3 is a protein belonging to the same desmosomal cadherin family as the DSG2 reported in this study, which had a non-linear association with the p-factor with increased abundancies in the middle part of the p-factor scale. DSG2 was previously shown to have a similar function as DSG3 and was also shown to compensate for DSG3 in DSG3⁻ mouse models [70]. The Shin et al. paper investigated BD, which is classified as a Thought Disorder factor, and MDD, which is classified as an Internalizing factor [5], so the differences in the abundance changes observed compared to our work, where we used the combined p-factor, are to be expected. Laminin subunit beta-1 (LAMB1) was associated with the p-factor in this study. A polymorphism in *LAMB1* gene has been earlier associated with autism severity [71], neural development of embryonic stem cells [72] and pain sensitivity in mice [73]. *LAMB1* is expressed during the early development of nervous system [71] and in the hippocampus in the mature brain [74]. In rats, LAMB1 showed negative regulation of spatial learning through the inhibition of the ERK/MAPK-SGK1 signaling pathway in the hippocampus [74]. Furthermore, loss of LAMB1 in the anterior cingulate cortex was found to increase pain sensitivity and be associated with anxiety- and depressive-like behavior in mice [73].

Cathepsin B (CTSB) was identified here with a non-linear relationship to the p-factor. Moon *et al.* suggested CTSB as a mediator of exercise-induced effects on brain health by enhancing the expression of neurotrophins [75]. Exercise was found to increase plasma CTSB levels in monkeys and humans [75], but a 20-week exercise intervention in children did not find any significant connection between CTSB and brain health outcomes [76]. Additionally, CTSB has been connected to brain-related functions in several mice studies [77–79]. For example, a mouse

model for chronic social stress revealed increased activity of cathepsin B in the hypothalamus and nucleus caudatus with depressive-like behavior [79]. Contrarily, decreased cathepsin B activity was found after acute emotional stress in mice [77]. CTSB shows a potential mediator role in the brain induced by physical and mental stressors, which should be further investigated.

Other proteins significantly associated with the p-factor and directly connected to EGFR in our study were golgi membrane protein 1 (GOLM1), superoxide dismutase 2 (SOD2), and uromoduline UMOD. Increased *GOLM1* gene expression was found in soldiers with PTSD [80]. This effect occurs through the activation of ErbB4-BDNF signaling pathway [81, 82]. Particularly strong evidence supports the role of the Neuregulin-1 (NRG1)-ErbB4 signaling on synaptic plasticity [51, 52]. Neuregulins are a family of epidermal growth factor-related proteins acting on the ErbB tyrosine kinase receptors [51, 52]. SOD2 was found to play a role in neurodegenerative disease according to the Disease Ontology database, a polymorphism in the *sod2* gene was associated with differences in white matter microstructure and suboptimal brain aging [83]. For the uromodulin (UMOD) proteins and CST6 gene, no previous connection to mental health problems was reported.

We observed associations of plasma reticulon-4 receptor-like 2 (RTN4L2), and ficolin 3 (FCN3) with the p-factor. Reticulon-4 receptors (RTN4R), also known as NogoRs, are surface proteins expressed in neurons [84]. RTN4Rs are involved in synaptogenesis and inhibition of axonal and dendrite growth, and, thus neuronal plasticity [84, 85]. Human genetics studies have revealed the linkage between Nogo receptors and SCZ [85–87]. For example, a rare variant in *RTN4R*, affecting the formation of growth cones in vitro, was associated with SCZ [87]. The role of RTN4Rs in SCZ seems to be mediated by neurodevelopmental and myelin-related abnormalities [85]. However, further studies are needed to clarify the exact role of RTN4Rs in mental health. Interestingly, ficolin activation was negatively associated with severity of SCZ [88], and in our recent study, the plasma abundance of ficolin 2, a similar protein, was found to be positively correlated with the Strength and Difficulties Questionnaire (SDQ) score in adolescents.

Half of the proteins found to be significant in this study were connected to the extracellular matrix. HSPG2, FBLN1, and LAMB1 were also strongly connected to each other, according to the STRING database, being structural components of the basal membrane, specifically in the brain. Coupled with proteins related to neuronal plasticity, proteins identified in this study may be potentially connected to the previously noted inverse relationship between the p-factor and the microstructural integrity of white matter as observed through neuroimaging [89]. Further studies are needed to investigate the possible connections of the found proteins with the brain microstructure and functioning.

Large-scale proteomic studies with plasma samples can present multiple challenges that need to be addressed to generate robust and meaningful results. For instance, protein expression in plasma is dynamic, and both interindividual and sample variability can be notable. Furthermore, plasma proteomic studies differ in the pipelines and methods used due to a lack of standard protocols [19]. Additional challenges include ensuring consistent sample handling and processing [90], normalizing data, correcting signal drift and batch effects [35, 91], accounting for biological variability [92], improving reproducibility [93], and managing the resource-intensive nature of such studies [94]. Despite these limitations, proteomics remains a powerful tool that can contribute to better diagnostics of mental health [95, 96]. The major constraint in this study is that the proteomic data was only obtained once for each participant. This one-time snapshot of a dynamically evolving organism makes it challenging to conclusively link the identified biomarkers to the investigated p-factor. The true nature of these associations is also hard to determine based solely on these data. These correlations could be the outcome of underlying biological

processes or inherent biological traits of the participants, which might simultaneously influence both protein abundance and the p-factor (the observed behavior). Alternatively, the changes in protein abundance and the p-factor could be causally related, either as a cause or as an effect. Mental conditions may cause divergent effects on the abundance of plasma proteins, as demonstrated in the study by Shin and colleagues [69]. Further investigations will benefit from the inclusion of the disorder symptoms into the p-factor, which is missing from the score used in the present manuscript. These limitations suggest that a more detailed investigation into the various components of the p-factor may be needed to identify more specific biomarkers.

The strength of this study lies in its large cohort size, and the use of modern proteomics methods, which made it possible to obtain proteome profiles of hundreds of individuals, each comprising hundreds of plasma protein abundancies. This large scale allows us to identify common patterns in the proteomes of individuals with high and low p-factor values. While the changes in plasma abundancies of some of the proteins were previously reported, other proteins were linked to a vulnerability to the development of general psychopathology for the first time. Our research utilized the FT12 cohort, a large and thoroughly characterized population-based cohort with a broad range of measured characteristics, making the proteomic data gathered in this study an invaluable resource for future exploration and analysis.

CONCLUSIONS

The study suggests that examining plasma proteomic profiles makes it possible to elucidate the biological processes related to the p-factor, which may inform the future development of novel screening, diagnostic, or therapeutic strategies for mental disorders. The results revealed proteins with common cellular functions connected to the p-factor, reflecting the general psychopathology. However, further studies are needed to examine the identified proteins and their potential as biomarkers for mental health dysfunction. In the future, utilization of the p-factor may also have implications for the development of interventions targeting common underlying factors that contribute to multiple forms of mental disorders. By addressing these shared factors, interventions could potentially be more effective in improving mental health outcomes across a range of disorders.

REFERENCES

1. Organization WH. World mental health report: transforming mental health for all. (2022).
2. World Health Organization. Adolescent mental health. 2021. https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health. Accessed 1 September 2022.
3. Health at a Glance: Europe 2022. OECD; 2022.
4. Smith GT, Atkinson EA, Davis HA, Riley EN, Oltmanns JR. The general factor of psychopathology. Annu Rev Clin Psychol. 2020;16:75–98.
5. Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, et al. The p factor: one general psychopathology factor in the structure of psychiatric disorders? Clin Psychol Sci. 2014;2:119–37.
6. Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is there a general factor of prevalent psychopathology during adulthood? J Abnorm Psychol. 2012;121:971–7.

7. Ronald A. Editorial: the psychopathology p factor: will it revolutionise the science and practice of child and adolescent psychiatry? J Child Psychol Psychiatry. 2019;60:497–9.

8. Jensen AR. The g factor: the science of mental ability. Psycoloquy. 1999;10:36–2443.

9. Oltmanns JR, Smith GT, Oltmanns TF, Widiger TA. General factors of psychopathology, personality, and personality disorder: across domain comparisons. Clin Psychol Sci. 2018;6:581–9.

10. Laceulle OM, Vollebergh WAM, Ormel J. The structure of psychopathology in adolescence. Clin Psychol Sci. 2015;3:850–60.

11. Shanmugan S, Wolf DH, Calkins ME, Moore TM, Ruparel K, Hopson RD, et al. Common and dissociable mechanisms of executive system dysfunction across psychiatric disorders in youth. Am J Psychiatry. 2016;173:517–26.

12. Hartwig FP, Borges MC, Horta BL, Bowden J, Davey Smith G. Inflammatory biomarkers and risk of schizophrenia: a 2-sample mendelian randomization study. JAMA Psychiatry. 2017;74:1226–33.

13. Haywood D, Baughman FD, Mullan BA, Heslop KR. What accounts for the factors of psychopathology? An investigation of the neurocognitive correlates of internalising, externalising, and the p-factor. Brain Sci. 2022;12:421.

14. Pulkkinen L. Human development from middle childhood to middle adulthood. London: Routledge; 2017.

15. Walters RK, Polimanti R, Johnson EC, McClintick JN, Adams MJ, Adkins AE, et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Nat Neurosci. 2018;21:1656–69.

16. Munn-Chernoff MA, Johnson EC, Chou Y, Coleman JRI, Thornton LM, Walters RK, et al. Shared genetic risk between eating disorder- and substance-use-related phenotypes: Evidence from genome-wide association studies. Addict Biol. 2021;26:e12880.

17. Orlando E, Aebersold R. On the contribution of mass spectrometry-based platforms to the field of personalized oncology. TrAC Trends Anal Chem. 2019;110:129–42.

18. Zhou B, Zhou Z, Chen Y, Deng H, Cai Y, Rao X, et al. Plasma proteomics-based identification of novel biomarkers in early gastric cancer. Clin Biochem. 2020;76:5–10.

19. Ignjatovic V, Geyer PE, Palaniappan KK, Chaaban JE, Omenn GS, Baker MS, et al. Mass spectrometry-based plasma proteomics: considerations from sample collection to achieving translational data. J Proteome Res. 2019;18:4085–97.

20. 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;8:160.

21. Domenici E, Willé 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:e9166.

22. Ziani PR, Feiten JG, Goularte JF, Colombo R, Antqueviezc B, Géa LP, et al. Potential candidates for biomarkers in bipolar disorder: a proteomic approach through systems biology. Clin Psychopharmacol Neurosci. 2022;20:211–27.

23. Malik S, Singh R, Arora G, Dangol A, Goyal S. Biomarkers of major depressive disorder: knowing is half the battle. Clin Psychopharmacol Neurosci. 2021;19:12–25.

24. Campeau A, Mills RH, Stevens T, Rossitto L-A, Meehan M, Dorrestein P, et al. Multiomics of human plasma reveals molecular features of dysregulated inflammation and accelerated aging in schizophrenia. Mol Psychiatry. 2022;27:1217–25.

25. Kaprio J. Twin studies in Finland 2006. Twin Res Hum Genet. 2006;9:772–7.

26. Rose RJ, Salvatore JE, Aaltonen S, Barr PB, Bogl LH, Byers HA, et al. FinnTwin12 cohort: an updated review. Twin Res Hum Genet. 2019;22:302–11.

27. Whipp AM, Heinonen-Guzejev M, Pietiläinen KH, van Kamp I, Kaprio J. Branched-chain amino acids linked to depression in young adults. Front Neurosci. 2022;16:935858.

28. Pulkkinen L, Kaprio J, Rose RJ. Peers, teachers and parents as assessors of the behavioural and emotional problems of twins and their adjustment: the Multidimensional Peer Nomination Inventory. Twin Res. 1999;2:274–85.

29. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. Psychol Bull. 1987;101:213–32.

30. Whipp AM, Korhonen T, Raevuori A, Heikkilä K, Pulkkinen L, Rose RJ, et al. Early adolescent aggression predicts antisocial personality disorder in young adults: a population-based study. Eur Child Adolesc Psychiatry. 2019;28:341–50.

31. Whipp AM, Vuoksimaa E, Bolhuis K, de Zeeuw EL, Korhonen T, Mauri M, et al. Teacher-rated aggression and co-occurring behaviors and emotional problems among schoolchildren in four population-based European cohorts. PLoS One. 2021;16:e0238667.

32. Bateman A, Martin M-J, Orchard S, Magrane M, Ahmad S, Alpi E, et al. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2023;51:D523–D531.

33. Frankenfield AM, Ni J, Ahmed M, Hao L. Protein contaminants matter: building universal protein contaminant libraries for DDA and DIA proteomics. J Proteome Res. 2022;21:2104–13.

34. Pham TV, Henneman AA, Jimenez CR. iq: an R package to estimate relative protein abundances from ion quantification in DIA-MS-based proteomics. Bioinformatics. 2020;36:2611–3.

35. Čuklina J, Lee CH, Williams EG, Sajic T, Collins BC, Rodríguez Martínez M, et al. Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. Mol Syst Biol. 2021;17:e10240.

36. Liu M, Dongre A. Proper imputation of missing values in proteomics datasets for differential expression analysis. Brief Bioinform. 2021;22:bbaa112.

37. Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Ann Appl Stat. 2016;10:946–63.

38. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.

39. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57:289–300.

40. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021;49:D605–D612.

41. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, 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.

42. Wickham H ggplot2. New York, NY: Springer New York; 2009.

43. Schriml LM, Munro JB, Schor M, Olley D, McCracken C, Felix V, et al. The human disease ontology 2022 update. Nucleic Acids Res. 2022;50:D1255–D1261.

44. Fernandes BS, Dai Y, Jia P, Zhao Z. Charting the proteome landscape in major psychiatric disorders: from biomarkers to biological pathways towards drug discovery. Eur Neuropsychopharmacol. 2022;61:43–59.

45. García-Gutiérrez MS, Navarrete F, Sala F, Gasparyan A, Austrich-Olivares A, Manzanares J. Biomarkers in psychiatry: concept, definition, types and relevance to the clinical reality. Front Psychiatry. 2020;11:432.

46. Rodríguez Cerdeira C, Sánchez-Blanco E, Sánchez-Blanco B, González-Cespón JL. Protein biomarkers of mood disorders. Int J Immunopathol Pharm. 2017;30:7–12.

47. Human Protein Atlas. proteinatlas.org. 2022.

48. Uhlén M, Karlsson MJ, Hober A, Svensson A-S, Scheffel J, Kotol D, et al. The human secretome. Sci Signal. 2019;12:eaaz0274.

49. Ji E, Boerrigter D, Cai HQ, Lloyd D, Bruggemann J, O'Donnell M, et al. Peripheral complement is increased in schizophrenia and inversely related to cortical thickness. Brain Behav Immun. 2022;101:423–34.

50. Birchmeier C. ErbB receptors and the development of the nervous system. Exp Cell Res. 2009;315:611–8.

51. Ledonne A, Mercuri NB. On the modulatory roles of neuregulins/ErbB signaling on synaptic plasticity. Int J Mol Sci. 2019;21:275.

52. Mei L, Nave K-A. Neuregulin-ERBB signaling in the nervous system and neuropsychiatric diseases. Neuron. 2014;83:27–49.

53. Borges JP, Mekhail K, Fairn GD, Antonescu CN, Steinberg BE. Modulation of pathological pain by epidermal growth factor receptor. Front Pharm. 2021;12:642820.

54. Chen Y-H, Hu N-Y, Wu D-Y, Bi L-L, Luo Z-Y, Huang L, et al. PV network plasticity mediated by neuregulin1-ErbB4 signalling controls fear extinction. Mol Psychiatry. 2022;27:896–906.

55. Fiori LM, Kos A, Lin R, Théroux J-F, Lopez JP, Kühne C, et al. miR-323a regulates ERBB4 and is involved in depression. Mol Psychiatry. 2021;26:4191–204.

56. Mei L, Xiong W-C. Neuregulin 1 in neural development, synaptic plasticity and schizophrenia. Nat Rev Neurosci. 2008;9:437–52.

57. Nawwar DA, Zaki HF, Sayed RH. Role of the NRG1/ErbB4 and PI3K/AKT/mTOR signaling pathways in the anti-psychotic effects of aripiprazole and sertindole in ketamine-induced schizophrenia-like behaviors in rats. Inflammopharmacology. 2022;30:1891–907.

58. Shi L, Bergson CM. Neuregulin 1: an intriguing therapeutic target for neurodevelopmental disorders. Transl Psychiatry. 2020;10:190.

59. Bishop JR, Schuksz M, Esko JD. Heparan sulphate proteoglycans fine-tune mammalian physiology. Nature. 2007;446:1030–7.

60. Long Q, Wang R, Feng M, Zhao X, Liu Y, Ma X, et al. Construction and analysis of a diagnostic model based on differential expression genes in patients with major depressive disorder. Front Psychiatry. 2021;12:762683.

61. MacNeil RR, Müller DJ. Genetics of common antipsychotic-induced adverse effects. Complex Psychiatry. 2016;2:61–78.

62. Zai CC, Maes MS, Tiwari AK, Zai GC, Remington G, Kennedy JL. Genetics of tardive dyskinesia: promising leads and ways forward. J Neurol Sci. 2018;389:28–34.

63. Zai CC, Lee FH, Tiwari AK, Lu JY, Luca Vde, Maes MS, et al. Investigation of the HSPG2 gene in tardive dyskinesia – new data and meta-analysis. Front Pharm. 2018;9:974.

64. Nakamura K, Ikeuchi T, Nara K, Rhodes CS, Zhang P, Chiba Y, et al. Perlecan regulates pericyte dynamics in the maintenance and repair of the blood–brain barrier. J Cell Biol. 2019;218:3506–25.

A.M. Afonin et al.

65. Tordera RM, Garcia-García AL, Elizalde N, Segura V, Aso E, Venzala E, et al. Chronic stress and impaired glutamate function elicit a depressive-like phenotype and common changes in gene expression in the mouse frontal cortex. Eur Neuropsychopharmacol. 2011;21:23–32.

66. Bohlega S, Al-Ajlan H, Al-Saif A. Mutation of fibulin-1 causes a novel syndrome involving the central nervous system and connective tissues. Eur J Hum Genet. 2014;22:640–3.

67. Cooley MA, Kern CB, Fresco VM, Wessels A, Thompson RP, McQuinn TC, et al. Fibulin-1 is required for morphogenesis of neural crest-derived structures. Dev Biol. 2008;319:336–45.

68. Ohsawa I, Takamura C, Kohsaka S. Fibulin-1 binds the amino-terminal head of β-amyloid precursor protein and modulates its physiological function. J Neurochem. 2001;76:1411–20.

69. Shin D, Rhee SJ, Lee J, Yeo I, Do M, Joo E-J, et al. Quantitative proteomic approach for discriminating major depressive disorder and bipolar disorder by multiple reaction monitoring-mass spectrometry. J Proteome Res. 2021;20:3188–203.

70. Hartlieb E, Rötzer V, Radeva M, Spindler V, Waschke J. Desmoglein 2 compensates for desmoglein 3 but does not control cell adhesion via regulation of p38 mitogen-activated protein kinase in keratinocytes. J Biol Chem. 2014;289:17043–53.

71. Kim YJ, Park JK, Kang WS, Kim SK, Park HJ, Nam M, et al. LAMB1 polymorphism is associated with autism symptom severity in Korean autism spectrum disorder patients. Nord J Psychiatry. 2015;69:594–8.

72. Sun Y-M, Cooper M, Finch S, Lin H-H, Chen Z-F, Williams BP, et al. Rest-mediated regulation of extracellular matrix is crucial for neural development. PLoS One. 2008;3:e3656.

73. Li Z-Z, Han W-J, Sun Z-C, Chen Y, Sun J-Y, Cai G-H, et al. Extracellular matrix protein laminin β1 regulates pain sensitivity and anxiodepression-like behaviors in mice. J Clin Invest. 2021;131:e146323.

74. Yang YC, Ma YL, Liu WT, Lee EH. Laminin-β1 impairs spatial learning through inhibition of ERK/MAPK and SGK1 signaling. Neuropsychopharmacology. 2011;36:2571–86.

75. Moon HY, Becke A, Berron D, Becker B, Sah N, Benoni G, et al. Running-induced systemic cathepsin B secretion is associated with memory function. Cell Metab. 2016;24:332–40.

76. Rodriguez-Ayllon M, Plaza-Florido A, Mendez-Gutierrez A, Altmäe S, Solis-Urra P, Aguilera CM, et al. The effects of a 20-week exercise program on blood-circulating biomarkers related to brain health in overweight or obese children: the ActiveBrains project. J Sport Health Sci. 2023;12:175–85.

77. Sharanova NE, Kirbaeva NV, Toropygin IY, Khryapova EV, Koplik EV, Soto CK, et al. Effect of acute emotional stress on proteomic profile of selected brain areas and lysosomal proteolysis in rats with different behavioral activity. Bull Exp Biol Med. 2016;161:355–8.

78. Wang Z, Li P, Wu T, Zhu S, Deng L, Cui G. Axon guidance pathway genes are associated with schizophrenia risk. Exp Ther Med. 2018;16:4519–26.

79. Zhanaeva SY, Rogozhnikova AA, Alperina EL, Gevorgyan MM, Idov GV. Changes in activity of cysteine cathepsins b and l in brain structures of mice with aggressive and depressive-like behavior formed under conditions of social stress. Bull Exp Biol Med. 2018;164:425–9.

80. Boscarino C, Nalpathamkalam T, Pellecchia G, Li W, Thiruvahindrapuram B, Merico D. Using next-generation sequencing transcriptomics to determine markers of post-traumatic symptoms: preliminary findings from a post-deployment cohort of soldiers. G3. 2019;9:463–71.

81. Han Q-Q, Wu P-F, Li Y-H, Cao Y, Chen J-G, Wang F. SVCT2–mediated ascorbic acid uptake buffers stress responses via DNA hydroxymethylation reprogramming of S100 calcium-binding protein A4 gene. Redox Biol. 2022;58:102543.

82. Pankratova S, Klingelhofer J, Dmytriyeva O, Owczarek S, Renziehausen A, Syed N, et al. The S100A4 protein signals through the ErbB4 receptor to promote neuronal survival. Theranostics. 2018;8:3977–90.

83. Salminen LE, Schofield PR, Pierce KD, Bruce SE, Griffin MG, Tate DF, et al. Vulnerability of white matter tracts and cognition to the SOD2 polymorphism: a preliminary study of antioxidant defense genes in brain aging. Behav Brain Res. 2017;329:111–9.

84. Wang J, Miao Y, Wicklein R, Sun Z, Wang J, Jude KM, et al. RTN4/NoGo-receptor binding to BAI adhesion-GPCRs regulates neuronal development. Cell. 2021;184:5869–.e25.

85. Willi R, Schwab ME. Nogo and Nogo receptor: relevance to schizophrenia? Neurobiol Dis. 2013;54:150–7.

86. Jitoku D, Hattori E, Iwayama Y, Yamada K, Toyota T, Kikuchi M, et al. Association study of Nogo-related genes with schizophrenia in a Japanese case-control sample. Am J Med Genet Part B Neuropsychiatr Genet. 2011;156:581–92.

87. Kimura H, Fujita Y, Kawabata T, Ishizuka K, Wang C, Iwayama Y, et al. A novel rare variant R292H in RTN4R affects growth cone formation and possibly contributes to schizophrenia susceptibility. Transl Psychiatry. 2017;7:e1214.

88. Gracia DFK, Gomes EMP, Coelho TA, Carriello M, de Andrade FA, Bavia L, et al. Ficolin activation as a potential biomarker of the severity of schizophrenia. Psychiatry Res. 2021;304:114122.

89. Romer AL, Knodt AR, Sison ML, Ireland D, Houts R, Ramrakha S, et al. Replicability of structural brain alterations associated with general psychopathology: evidence from a population-representative birth cohort. Mol Psychiatry. 2021;26:3839–46.

90. Bell AW, Deutsch EW, Au CE, Kearney RE, Beavis R, Sechi S, et al. A HUPO test sample study reveals common problems in mass spectrometry–based proteomics. Nat Methods. 2009;6:423–30.

91. Lazar C, Gatto L, Ferro M, Bruley C, Burger T. Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. J Proteome Res. 2016;15:1116–25.

92. Paul D, Kumar A, Gajbhiye A, Santra MK, Srikanth R. Mass spectrometry-based proteomics in molecular diagnostics: discovery of cancer biomarkers using tissue culture. Biomed Res Int. 2013;2013:1–16.

93. Ioannidis JPA. Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses. JAMA. 2011;305:2200.

94. Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. Nat Methods. 2014;11:1114–25.

95. Davalieva K, Kostovska IM, Dwork AJ. Proteomics research in schizophrenia. Front Cell Neurosci. 2016;10:18.

96. Rodrigues-Amorim D, Rivera-Baltanás T, Vallejo-Curto MdelC, Rodriguez-Jamardo C, de las Heras E, Barreiro-Villar C, et al. Proteomics in schizophrenia: a gateway to discover potential biomarkers of psychoneuroimmune pathways. Front Psychiatry. 2019;10:885.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

KMK, ISM, JK, and IvK designed the study. JK, AW LP, and RJR guided the process and provided the samples. ISM, MI, and AA pre-processed the samples. AMA performed statistical and bioinformatics analyses. A-KP, AMA, ISM, and MI participated in planning and literature analyses. AMA, A-KP, AW, and KMK drafted the manuscript. AW, JK, LP, and RJR provided critical input to the draft manuscript. All authors read and approved the final version of the manuscript.

CONFLICT OF INTEREST

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-024-02751-z.

Correspondence and requests for materials should be addressed to Katja M. Kanninen.

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

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