nature mental health

Article

https://doi.org/10.1038/s44220-023-00103-2

Plasma proteomics discovery of mental health risk biomarkers in adolescents

Received: 22 December 2022

Accepted: 29 June 2023

Published online: 31 July 2023

Check for updates

Izaque de Sousa Maciel[1,8], Aino-Kaisa Piironen[1,8], Alexey M. Afonin[1,8], Mariia Ivanova[1], Arto Alatalo[1], Kaustubh Kishor Jadhav[1], Jordi Julvez[2,3,4], Maria Foraster[3,4,5,6], Irene van Kamp[7] & Katja M. Kanninen[1]

An estimated 10–20% of adolescents experience mental health conditions, and most of them remain underdiagnosed and undertreated. Discovering new susceptibility biomarkers is therefore important for identifying individuals at high risk of developing mental health problems, and for improving early prevention. Here we aimed to discover plasma protein-based susceptibility biomarkers in children/adolescents aged 11–16 years at risk of developing mental health issues. Risk was evaluated on the basis of self-reported Strengths and Difficulties Questionnaire (SDQ) scores, and plasma proteomic data were obtained for individuals participating in the Spanish WALNUTs cohort study by liquid chromatography–tandem mass spectrometry. Bioinformatic analyses were performed to identify the biological processes and pathways in which the identified biomarker candidates are involved; 58 proteins were significantly associated with the SDQ score. The most prominent enriched pathways related to these proteins included immune responses, blood coagulation, neurogenesis and neuronal degeneration. This exploratory study revealed several alterations of plasma proteins associated with the SDQ score in adolescents, which opens a new avenue to develop novel susceptibility biomarkers to improve early identification of individuals at risk of mental health problems.

Adolescence is a period of life of profound changes in the biological, psychosocial, cognitive and emotional domains[1–3]. The dynamic brain development during youth opens a critical window for cognitive improvement, but also the onset and development of mental disorders[4,5]. Several mental disorders, such as attention deficit hyperactivity disorder, phobias, obsessive compulsive disorder, eating disorder, substance use, mood and social anxiety disorder, begin before the individual reaches adulthood, with a peak age of onset of 14 years[5–7]. Mental disorders negatively impact adolescent development, leading to morbidity, mortality and dysfunction in later life[5,8]. Therefore, identifying adolescents with a high risk of developing mental health issues and improving early diagnostics could improve the clinical outcomes and decrease the socio-economic impact.

Globally the prevalence of mental health conditions in adolescents is estimated to be between 10–20%, and most cases remain underdiagnosed and undertreated[9,10]. The social stigma of mental disorders, the adolescent and parent perception of mental health care needs, and the lack of mental health resources are some factors that contribute to the number of adolescents without proper diagnosis and treatment[11,12]. Furthermore, misdiagnosis or overdiagnosis could expose the adolescent to unnecessary treatment[13]. Diagnostics of mental disorders are based on the International Classification of Diseases (ICD) and the

[1]A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland. [2]Clinical and Epidemiological Neuroscience (NeuroÈpia), Institut d'Investigació Sanitària Pere Virgili (IISPV), Reus, Spain. [3]ISGlobal, Barcelona, Spain. [4]CIBER Epidemiología y Salud Pública (CIBERESp), Madrid, Spain. [5]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [6]PHAGEX Research Group, Blanquerna School of Health Science, Universitat Ramon Llull (URL), Barcelona, Spain. [7]Centre for Sustainability, Environment and Health, National Institute for Public Health and the Environment, Bilthoven, the Netherlands. [8]These authors contributed equally: Izaque de Sousa Maciel, Aino-Kaisa Piironen, Alexey M. Afonin. e-mail: katja.kanninen@uef.fi

Diagnostic and Statistical Manual (DSM) classifications provided by the World Health Organization (WHO) and the American Psychiatric Association, respectively[14]. Clinical interviews and validated questionnaires used for symptom assessment (for example, the Beck Depression Inventory[15]) have a major role in mental health diagnostics. The complexity of adolescent behavior and the overlap of symptomatology of several mental disorders complicate the precise and objective diagnosis of diseases in youth. Furthermore, the difficulty in defining normative or atypical expected behavioral development in adolescence[16], and lack of access to professional expertise, contribute to inaccurate judgment and precise definition of mental health conditions in adolescents[17,18].

The Strengths and Difficulties Questionnaire (SDQ) is a screening questionnaire for emotional and behavioral problems in children and young people that assesses the impact of difficulties on the child's life, including (1) emotional symptoms, (2) conduct problems, (3) hyperactivity/inattention, (4) peer relationship problems and (5) prosocial behavior[19–21]. Past validation studies have shown that the total SDQ score can be considered as a predictive factor for mental health disorders, as children with high SDQ scores have an increased probability for clinical disorders[22,23]. Differences in the total SDQ score seem to reflect the differences in prevalence of mental health disorders, although cross-national differences exist[22,24]. Thus, developing additional tools such as biological measurements for assessing mental health issues could improve identification of adolescents at high risk of mental health dysfunction, and enhance more precise diagnostics.

Although studying human brain tissue may be the most revealing method for measuring alterations related to mental disorders, it poses several severe limitations, including tissue access[25] and high cost in the case of neuroimaging[25]. By contrast, biological fluids such as blood or urine are easier to access and are routinely used for clinical diagnostics. Alterations in the gene expression levels, proteins abundance and biological activity can serve as internal indicators present in biological fluids (biomarkers), of pathogenic processes or responses to an external exposome[25,26]. The blood connects the brain and periphery, and changes in plasma components such as proteins can reflect alterations in the brain associated with mental disorders due to the two-way communication between the central nervous system (CNS) and peripheral circulation[27,28]. Past studies have shown the plasma proteomic changes associated with mental disorders[29–32]. For example, significant reductions in glia maturation factor beta and brain-derived neurotrophic factor were observed in patients with schizophrenia (SCZ) when compared with healthy volunteers[29]. Thus, peripheral blood plasma is a suitable biological fluid for investigating molecular alterations that reflect those associated with mental health issues, and for providing new understanding on the bidirectional communication between the brain and body[27–30].

Limited knowledge exists on whether alterations in plasma proteins could serve as early susceptibility biomarkers to predict the risk of mental health issues, leading to proper clinical interventions before disease onset, even though alterations in pathways and molecules related to hormone signaling, energy metabolism, growth factors, inflammation, oxidation/reduction and protein synthesis have been commonly associated with psychiatric disorders[28]. However, studies by Mongan et al.[33] and English et al.[34] suggest that adolescents at high risk of psychosis could be identified on the basis of the changes in the blood proteome several years before the psychotic experiences manifest. The number of studies focusing on discovering new susceptibility or predictive plasma biomarkers for mental health diseases in adolescents is so far limited.

This explorative study aims to identify and characterize alterations of plasma proteins in adolescents at high risk of developing mental health issues. We identified 67 plasma proteins with abundances significantly associated with the SDQ score, offering new insight into using proteins as susceptibility biomarkers for early identification of adolescents at risk of mental health problems.

Table 1 | Sample characteristics

Group	Low		Raised	
	SDQ=0–14		SDQ=15–25	
Sex	**Male**	**Female**	**Male**	**Female**
SDQ scores (mean±s.d.)	3.50±1.41	3.50±1.47	16.92±2.01	18.00±2.42
Number (*n* (%))	24 (57.1)	18 (42.9)	27 (55.1)	22 (44.9)
Age (mean±s.d.)	13.54±1.07	13.53±0.73	13.93±1.03	14.37±1.13

n, number of samples (% of each group); s.d., standard deviation.

Plasma samples and behavioral outcome

The peripheral blood plasma samples were obtained and analyzed from a subsample of 91 adolescents, aged 11–16 years, of the WALNUTs regional Spanish study (Table 1). The samples were collected in 2016 at approximately the same time that the participants filled out the SDQ. This baseline subsample without any dietary intervention was selected on the basis of the availability of blood samples and filled SDQ questionnaires, as well as the total scores of the SDQ questionnaire. Based on the self-reported SDQ score, the plasma samples were categorized into lower (SDQ = 0–14) and raised (SDQ = 15–25) groups[35]. The plasma samples were stored undisturbed at −80 °C until they were thawed in 2021 for protein depletion and subsequent proteomic analysis. The studies were reviewed and approved by the CEIC Parc Salut Mar Clinical Research Ethics Committee (approval nos. 2015/6026, WALNUTs; 2020/9688, Equal-Life). Written informed consent to participate in the original WALNUTs study was provided by the participants' legal guardian/next of kin.

Plasma samples were pre-processed and analyzed using liquid chromatography electrospray ionization tandem mass spectrometry, which was performed at the Turku Proteomics Facility and supported by Biocenter Finland. The linear associations between the SDQ score and the protein abundances were investigated using linear modeling with DeqMS[36]. To characterize the biological processes and pathways related to the identified proteins, significantly differently abundant proteins (adjusted P-value ≤ 0.05) associated with the SDQ score were used in further bioinformatic data analyses. See the Methods for more details.

Results

Protein identification

Using mass spectrometry-based proteomics we successfully identified 1,485 proteins in the WALNUTs plasma samples (N = 91; mean of 1,228 proteins per sample; standard error = 117). The full list of the proteins is presented in the Supplementary Information. Out of these, 77 were identified as contaminants, and were removed. After that, 983 proteins were detected in at least 80% of the samples, and therefore these proteins were used for subsequent analysis.

The sex and age variables were added to the linear model to correct for the possible effects. In the analysis, 67 proteins a had linear relationship with the SDQ score, out which 48 were positively correlated with the SDQ score, and 19 were negatively correlated (Fig. 1b). The proteins associated with the SDQ score are presented in Table 2.

All the significantly altered proteins were used to create a heatmap (Fig. 1b) that shows the protein abundances (z-scores) in relation to the SDQ score.

Enriched pathways and biological processes

Of the highly abundant proteins that were depleted from the plasma samples before the mass spectrometry analysis, nine were found in the data; these were considered as a possible source of bias and were

https://doi.org/10.1038/s44220-023-00103-2

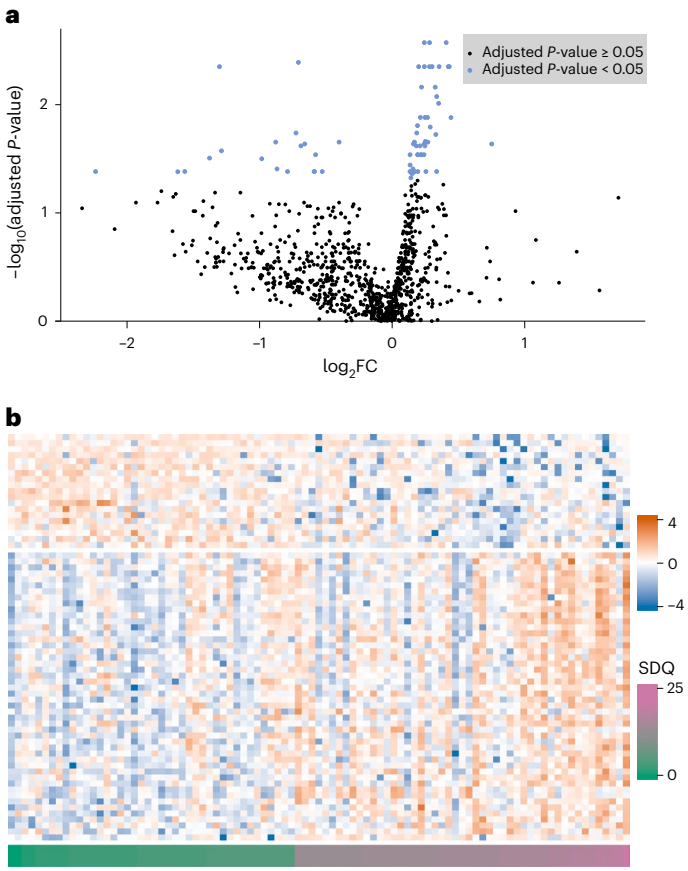

Fig. 1 | Significantly altered proteins identified in mass spectrometry-based proteomic analysis. a, Volcano plot of proteins associated with the SDQ score; 58 proteins were significantly changed, indicated in blue (adjusted *P*-value < 0.05). The *P*-values were calculated using DeqMS with SDQ as a continuous variable, and adjusted using the Benjamini–Hochberg method. **b**, A heatmap of protein abundances (*z*-scores) of proteins significantly associated with the SDQ score. The upper group represents negatively correlated proteins ($n = 19$), and the lower group positively correlated proteins ($n = 48$). SDQ scores are shown as the gradient at the bottom of the heatmap, green (pink) indicates individuals with a low (raised) SDQ score.

excluded from these analyses, leaving 58 proteins significantly associated with the SDQ score.

Those proteins were used for a clustering analysis of the differentially abundant proteins identified in our study using the STRING database[37]. The clustering analysis yielded three groups of proteins, as shown in Fig. 2a. Cluster 1 contained up- and down-regulated proteins involved in neuron growth, synaptic function, glial cell migration and cholesterol transport. The second cluster contained only up-regulated proteins mostly involved in the complement and coagulation cascades. Cluster 3 contained three down-regulated proteins involved in the olfactory system and three up-regulated proteins involved in protein degradation.

We also performed an analysis of enriched pathways with Reactome[38], using all of the identified proteins as the gene background. Enriched pathways were related to immune system, coagulation, complement cascade, and post-translational protein modification (Fig. 2b). In total, thirteen pathways were significantly (false discovery rate (FDR)-adjusted *P*-value < 0.05) enriched in the pathway analysis (Supplementary Table 1). The signaling pathways analysis in ingenuity pathway analysis (IPA) revealed that canonical pathways were associated with immune responses, coagulation, complement cascade and signaling, as in the Reactome analysis (Supplementary Table 2).

Predictive models generation

The relatively large number of samples made it possible to employ modern strategies to determine potentially predictive biomarkers for the low versus raised SDQ score groups. A novel QLattice algorithm[39] was used to create models containing predictive biomarkers that best separate the two groups with low and raised SDQ scores. The Bayesian information criterion (BIC) was used to ensure that the resulting models generalize well from the training to test set. We performed fivefold cross-validation of running logistical regression model with QLattice on different partitions of the data keeping the lowest BIC-scoring model from each partition. The receiver operator characteristic (ROC) curves and area under the ROC curve (AUC) for each of the models are presented in Supplementary Fig. 2.

Five diverse models were created using a fivefold cross-validation scheme. These models bring similar—albeit complementary—insights, as the whole dataset was split into training and validation sets five times, and each round contained different sub-samples of the data. The five unique models (Table 3) contained eleven proteins in total (Supplementary Table 3). Four of the five models contained proteins with a previously shown connection to the CNS. The first model contained three such proteins: amyloid beta precursor-like protein 1 (APLP1) (P51693), calcium/calmodulin dependent protein kinase II beta (CAMK2B) (Q13554/Q13555) and Reticulon 4 (RTN4; Q9NQC3), the ROC parameters for the models are shown in the Supplementary Fig. 2. Only the fifth model contained no proteins, previously connected to brain development. The proteins present in the models can be investigated further as potential biomarkers.

Discussion

Plasma proteomic biomarker studies in mental health diseases are a novel field. Increasing evidence shows alterations in plasma proteins associated with different mental disorders such as depression (MDD), SCZ, psychotic disorders and bipolar disorders[24,25,32]. Most altered pathways in mental disorders (such as complement cascade and signaling by interleukins) seem to be common to the above-mentioned major psychiatric disorders[32]. Here we report plasma protein alterations related to immune responses, blood coagulation, complement cascade, neuronal degeneration and neurogenesis in adolescents at high risk of mental dysfunction, which was evaluated based on the self-reported SDQ score. It should be kept in mind that assessing the risk of mental health problems in adolescents is associated with ethical issues, which should be appropriately considered.

In this study we used the total SDQ score as an indicator of mental health dysfunction and predisposition to mental health issues in adolescents. Becker and co-workers have shown the predictive value of the self-reported SDQ in clinical diagnostics, especially combined with parent and/or teacher versions[40]. Furthermore, the self-reported SDQ was shown to be a reliable and valid method for the assessment of behavioral problems in children and adolescents[40]. Goodman et al. have shown that multi-informant (parents, teachers, older children) SDQs in community samples can identify children and adolescents with a psychiatric diagnosis with a specificity of 94.6% and a sensitivity of 63.3%; SDQ scores successfully identified over 70% of individuals with conduct, hyperactivity, depressive and some anxiety disorders[19]. The SDQ performs well as a screening tool, but it is not intended to be used as a psychiatric diagnostic instrument as such[41]. It is therefore considered a useful and valid tool for screening children and adolescents at a high risk of mental disorders[19,41,42].

This study revealed 58 plasma protein alterations associated with the SDQ score in adolescents. The abundances of 39 proteins were enhanced in the raised SDQ score group, whereas 19 were reduced. We identified altered proteins such as clusterin, vitronectin, complement C2 and coagulation factor XI that have also been reported to be altered in past blood proteomics studies[33,43,44].

Table 2 | Plasma proteins with abundance changes associated with the SDQ score

Protein ID	Gene names	Protein names	Effect size	log2FC	Adjusted *P*-value
P13796	LCP1	Lymphocyte cytosolic protein 1	0.024393435	0.28230175	0.00266891423429772
Q5XPI4	RNF123	Ring finger protein 123	0.030214587	0.407569734	0.00266891423429772
P02763	ORM1	Orosomucoid 1*	0.041472001	0.582318552	0.00266891423429772
P00450	CP	Ceruloplasmin	0.021579105	0.242648946	0.00266891423429772
P01833	PIGR	Polymeric immunoglobulin receptor	−0.051686633	−0.707557707	0.00406569633896016
P35542	SAA4	Serum amyloid A4, constitutive	0.031761716	0.354177466	0.00444900848111389
P08697	SERPINF2	Serpin family F member 2	0.022213284	0.281759883	0.00444900848111389
P05155	SERPING1	Serpin family G member 1	0.019806854	0.24302875	0.00444900848111389
P02675	FGB	Fibrinogen β chain*	0.026409456	0.431154772	0.00444900848111389
P02679	FGG	Fibrinogen gamma chain*	0.025721779	0.423203306	0.00444900848111389
P01008	SERPINC1	Serpin family C member 1	0.018137241	0.199816818	0.00444900848111389
P25786	PSMA1	Proteasome 20S subunit α1	−0.095693078	−1.302777999	0.00444900848111389
Q9UK55	SERPINA10	Serpin family A member 10	0.023632886	0.301366666	0.00444900848111389
P05154	SERPINA5	Serpin family A member 5	0.027008252	0.324609321	0.00688379748291161
P04180	LCAT	Lecithin-cholesterol acyltransferase	0.020279571	0.221513014	0.00688379748291161
P02787	TF	Transferrin*	0.024014254	0.335644745	0.00840623855309166
P02671	FGA	Fibrinogen α chain*	0.021692839	0.350696404	0.00972440642891591
O95445	APOM	Apolipoprotein M	0.022035032	0.246726784	0.01311860384508440
P08571	CD14	CD14 molecule	0.018772876	0.212669508	0.01311860384508440
P43251	BTD	Biotinidase	0.022100724	0.267028414	0.01311860384508440
Q15485	FCN2	Ficolin 2	0.037120727	0.44342002	0.01314547997572250
P08603	CFH	Complement factor H	0.018284096	0.191087977	0.01561410991073270
P03951	F11	Coagulation factor XI	0.024018263	0.287092562	0.01608758697162620
P06681	C2	Complement C2	0.01678681	0.18562205	0.01826102103619690
P51693	APLP1	Amyloid β precursor-like protein 1	−0.048592355	−0.725496494	0.01826102103619690
P19652	ORM2	Orosomucoid 2*	0.027306639	0.329227308	0.01885805215970600
P01019	AGT	Angiotensinogen	0.018588708	0.250511329	0.02185375824202300
P00734	F2	Coagulation factor II, thrombin	0.0178514	0.16799325	0.02216801017152710
P26992	CNTFR	Ciliary neurotrophic factor receptor	−0.030575485	−0.400276076	0.02216801017152710
A6NE52	KIAA1875	WD repeat domain 97	−0.069097379	−0.878591149	0.02216801017152710
P02768	ALB	Albumin*	0.018244928	0.269527276	0.02216801017152710
P11171	EPB41	Erythrocyte membrane protein band 4.1	−0.048815886	−0.659499259	0.02306083695884880
P01042	KNG1	Kininogen 1	0.016314731	0.160500404	0.02306083695884880
P29622	SERPINA4	Serpin family A member 4	0.018413978	0.240780219	0.02306083695884880
Q13554;Q13555	CAMK2B, CAMK2G	Calcium/calmodulin-dependent protein kinase type II subunit β	0.051385424	0.751103187	0.02306083695884880
P06276	BCHE	Butyrylcholinesterase	0.020235784	0.247353801	0.02402185464772990
P62873	P62873	G protein subunit β 1	−0.054612995	−0.688126035	0.02402185464772990
P00751	CFB	Complement factor B	0.019090843	0.210533723	0.02402185464772990
P00742	F10	Coagulation factor X	0.017052085	0.1825767	0.02402185464772990
Q58FF3	HSP90B2P	Putative endoplasmin-like protein	−0.09637649	−1.287683106	0.02668279775737110
P04217	A1BG	α1-B glycoprotein*	0.019130571	0.194414326	0.02898562281342810
P05156	CFI	Complement factor I	0.018729496	0.217624798	0.02898562281342810
Q9UGM5	FETUB	Fetuin B	0.023263795	0.240217283	0.02898562281342810
P02652	APOA2	Apolipoprotein A2	0.020753485	0.191480134	0.02898562281342810
P11166	SLC2A1	Solute carrier family 2 member 1	−0.046363787	−0.577758688	0.02898562281342810
P00736	C1R	Complement C1r subcomponent	0.015778503	0.134858762	0.02898562281342810
P08185	SERPINA6	Serpin family A member 6	0.017945936	0.219458651	0.02898562281342810

https://doi.org/10.1038/s44220-023-00103-2

Table 2 (continued) | Plasma proteins with abundance changes associated with the SDQ score

Protein ID	Gene names	Protein names	Effect size	log2FC	Adjusted *P*-value
P07477;P07478;Q8NHM4	PRSS1, PRSS3P2, PRSS3P2	Serine protease 1	−0.100752815	−1.377624809	0.03115142391195680
Q6IF82	OR4A47	Olfactory receptor family 4 subfamily A member 47	−0.093345139	−0.983920911	0.03159089421977260
P04004	VTN	Vitronectin	0.015083377	0.135248758	0.03608129999454210
Q8N3T6	TMEM132C	Transmembrane protein 132C	−0.068568409	−0.868793155	0.03921664831857450
P02647	APOA1	Apolipoprotein A1*	0.017379495	0.156907997	0.03921664831857450
P22792	CPN2	Carboxypeptidase N subunit 2	0.022272689	0.254526745	0.04145715624086360
O95998	IL18BP	Interleukin 18 binding protein	−0.044789476	−0.588596594	0.04145715624086360
P13611	VCAN	Versican	−0.039810664	−0.586620961	0.04145715624086360
O95274	LYPD3	LY6/PLAUR domain-containing 3	−0.107072994	−1.564572013	0.04145715624086360
Q9NZP8	C1RL	Complement C1r subcomponent-like	0.016939083	0.17007371	0.04145715624086360
P12955	PEPD	Peptidase D	0.020336146	0.335635822	0.04145715624086360
P01023	A2M	α2-macroglobulin	0.01942573	0.193716524	0.04145715624086360
P55287	CDH11	Cadherin 11	−0.149776496	−2.237629717	0.04145715624086360
Q96RD9	FCRL5	Fc receptor-like 5	−0.034069414	−0.527369507	0.04145715624086360
P21926	CD9	CD9 molecule	−0.056132135	−0.78939602	0.04145715624086360
P02746	C1QB	Complement C1q subcomponent subunit B	0.013951103	0.134418819	0.04145715624086360
Q9NQC3	RTN4	Reticulon 4	−0.122514211	−1.617124146	0.04162111693323930
P10909	CLU	Clusterin	0.014330575	0.146386665	0.04293539920884360
P11279	LAMP1	Lysosome-associated membrane glycoprotein 1	0.013986417	0.153178004	0.04321300399755200
Q96PD5	PGLYRP2	*N*-acetylmuramoyl-L-alanine amidase	0.014620832	0.141416175	0.04742114492279560

Proteins are presented with their UniProt accession number and corresponding protein name. The asterisks indicate the highly abundant proteins that were depleted in the pre-processing stage. The effect size indicates the log2-fold-change in expression that results from a unit change in SDQ. The *P*-values were calculated using DeqMS SDQ score as continuous variable and adjusted using the Benjamini–Hochberg method. log2FC, log2-fold-change (ratio of means).

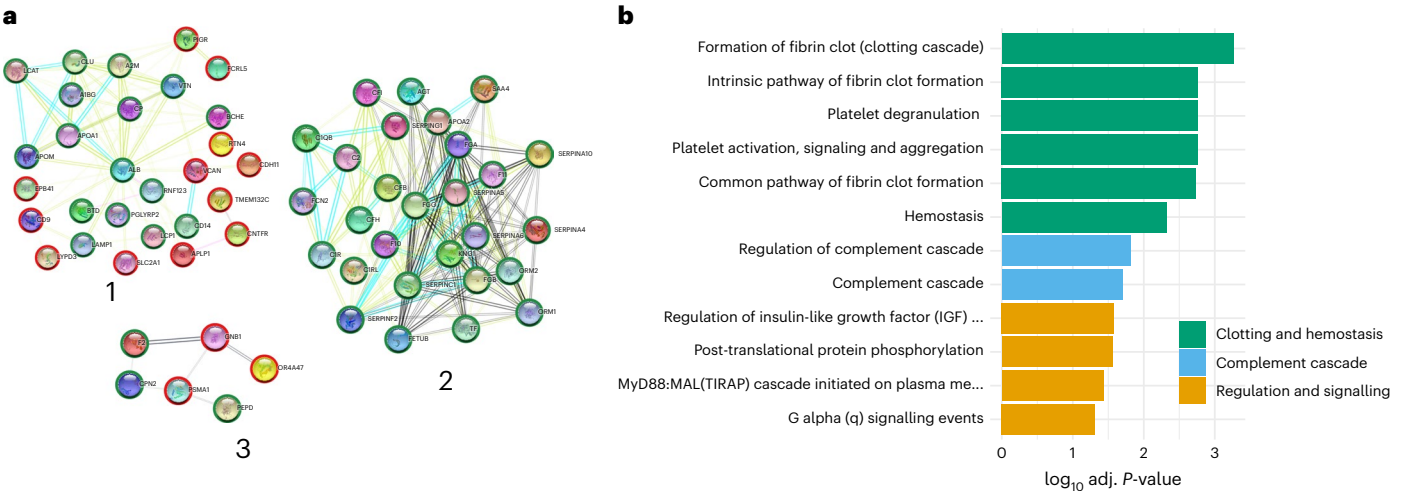

Fig. 2 | Enriched biological processes and pathways. a, The results of the STRINGdb clustering analysis; the proteins positively correlated with the SDQ score are highlighted green, whereas negatively correlated proteins are highlighted red. The number and colors of the lines represent the evidence for the protein connection according to the default STRING database scheme. **b**, Significantly enriched Reactome pathways. The colors indicate functional grouping. The enrichment was performed using the enrichPathway function of the ReactomePA package, which uses the hypergeometric model. The *P*-value was adjusted using the Benjamini–Hochberg method.

Blood coagulation and immune responses, including the complement cascade, were the most enriched pathways altered among proteins significantly associated with the SDQ score. Our clustering analysis revealed up-regulation of complement and blood coagulation cascades. Past studies have also shown associations between early changes in complement and coagulation cascades and increased risk of psychotic disorders in adolescents[33,34,45]. Changes in immune responses and blood coagulation have also been reported in SCZ, MDD and

Table 3 | The models returned by the QLattice with lowest BIC score

Model number	Model functional form genes	Model functional form accessions	BIC	Training set AUC
1	APLP1+CAMK2B x RTN4	P51693+Q13554 x Q9NQC3	49.33	0.95
2	PIGR+SERPINA4+CDH11	P01833+P29622+P55287	54.25	0.95
3	LYPD3+SERPING1+CAMK2B	O95274+P05155+Q13554;Q13555	49.88	0.95
4	LYPD3+BTD+APLP1	O95274+P43251+P51693	53.85	0.93
5	LYPD3+LCP1+CD9	O95274+P13796+P21926	48.46	0.94

Proteins in the models are potentially predictive biomarkers for the raised SDQ score in adolescents. The model for each protein contains gene names and protein accession numbers. Training set AUC performances and the BIC scores for each model are comparable. The first model contained non-linear interacting proteins CAMK2B and RTN4, whereas all of the other interactions were linear. The protein names and their relation to the accession codes are presented in Table 1.

bipolar disorder patients in several blood proteomic studies[43,44,46]. We found positive correlations with the SDQ score in coagulation factor XI, coagulation factor X and coagulation factor II (thrombin)—all of which are involved in blood coagulation. Increased levels of prothrombin and several coagulation factors (F5, F9, F12, F13A1) were also found by English et al. in adolescents who later developed a psychotic disorder[34].

Several complement components and factors such as C6, C1S, and CFI were altered (mostly increased) in high-risk psychotic disorder adolescents[33,34] and first-episode SCZ-patients[47]. In our study, complement proteins such as complement C1q and C1r subcomponents, complement factor I, complement factor H and complement C2 were significantly and mainly positively associated with the SDQ score. Jiang et al.[48] suggested that complement activation together with metabolic up-regulation can increase oxidative stress, which can induce protein damage and cell apoptosis, and thus contribute to the development of SCZ. Altogether, our results are in line with the current understanding of the role of altered immune responses and blood coagulation in pathophysiology of mental disorders. Furthermore, our findings support the increasing evidence on early changes in coagulation and complement cascades in predisposition to—and development of—mental health issues in adolescents.

A symbolic-regression-based algorithm, QLattice, was used to gain an insight into the potential of proteins to predict the SDQ status. The formed models comprised eleven proteins, four of which have a previously reported connection to the CNS, neurogenesis or mental health. Of the eleven proteins, eight were reported to belong to the first cluster according to STRING database analysis. Of the proteins not previously connected to the CNS, the LYPD3 protein was reported to be an amyloid precursor protein interactor[49], which can explain its coincidence with APLP1, CAMK2B and CD9 in our predicted models.

APLP1 is a protein residing predominantly in brain tissue, and it has been shown to be involved in brain development[50] and synaptogenesis during post-natal development in mice[51]. We identified a significant negative association between APLP1 and the SDQ score. Pandolfo and colleagues have suggested that amyloid could be a marker of cognitive impairment and altered neurodevelopment in mental diseases. Decreased β-amyloid proteins in CSF have been reported in patients with SCZ and MDD, and altered amyloid precursor protein metabolism in patients with bipolar disorders[52]. However, although the role of APLP1 in the pathogenesis of mental disorders is still unknown, this protein—on the basis of our data—warrants further investigation in the context of adolescent mental health.

Two other proteins—RTN4[53,54] and CAMK2B—have been reported to be connected to neuronal development and neuroplasticity[55,56]. CAMK2B was positively associated with the SDQ score. It is a protein connected to dendritic spine and synapse formation, neuronal plasticity and regulation of sarcoplasmic reticulum Ca^{2+} transport in skeletal muscle[57]. The beta subunit was reported to be brain specific[55], yet little is known of its involvement in mental disorders. Another protein present in the top model was RTN4, which was negatively associated with the SDQ score, and has been previously shown to be associated with SCZ[53,58]. RTN4 is a membrane shaping protein in the endoplasmic reticulum involved in the maintenance of the endoplasmic reticulum membrane tubular integrity. Impairment in the RTN4 process has been connected to neurodegeneration. The RTN4A-subtype, also known as Nogo-A, is localized in the CNS and has a role in neuronal growth and maturation during nervous system development[59]. Furthermore, RTN4 was shown to be connected with social behavior and spatial cognition in a study using mice with a missense mutation in the RTN4 receptor[59,60]. The involvement of RTN4 in adolescent mental health remains a topic worthy of detailed investigation. The fourth protein with reported connection to CNS was Cadherin 11 (P55287), which has been shown to be enriched in several brain areas during dendrite formation and synaptogenesis[61–63]. Given that the proteins reported as probable biomarkers belonged to the same cluster according to STRING database, and that four of the proteins were shown to be connected to the CNS and neuronal development, it is conceivable that all of the proteins from the cluster are connected to one process. Further investigation of this network might shed new light on the nuances of brain development and mental health in adolescents.

Girls in the raised SDQ score (>15) group reported slightly earlier puberty changes compared with the lower SDQ score (<14) group, whereas in boys, the self-reported puberty changes between the lower and the raised SDQ score groups seemed to be the opposite (Supplementary Table 4); however, differences in puberty changes were minor and sex was added as confounding factor in our linear models. We also examined other possible confounding factors available, including the education level for each parent, the levels of media consumption, levels of social media engagement, drug and alcohol use, and physical activity. The results showed no significant differences among the groups (P-value > 0.05). Furthermore, we used the information on the school of the subjects attended as a random variable in linear modeling. No differences in the number of significant proteins were found. We thus concluded that these factors were not likely to confound the investigation of the connection of SDQ to the protein abundance levels.

As the main limitation in this study, the sample number is low in relation to the identified proteins. Linear modeling was performed along with some group-based comparisons to strengthen the statistical power of the analysis, and we managed to detect statistically significant alterations with these sample numbers. Similar N-numbers have also been used in past studies on serum and plasma protein biomarkers in SCZ, MDD and bipolar disorder patients[45]. Furthermore, overnight fasting samples are preferred for proteomics analysis as food intake can influence the protein composition and concentrations in blood[64]. The plasma samples used in the current study were non-fasting samples due to the practical and ethical issues related to the implementation of the WALNUTs study as blood samples were drawn from the adolescents at school in the afternoon.

Conclusion

In this explorative study, we identified protein-based susceptibility biomarker candidates associated with the self-reported SDQ score in

adolescents reflecting a risk of developing mental health dysfunction. Significant alterations were found in proteins involved in the immune response, blood coagulation and hemostasis, neuronal degeneration and neurogenesis. Further studies are needed to confirm and validate these biomarker candidates in larger cohorts, as well as follow-up data and studies to evaluate whether these biomarkers are associated with the risk of transition to the clinical state and mental disorders.

Methods

Participant recruitment and sample collection

The studies were reviewed and approved by CEIC Parc Salut Mar Clinical Research Ethics Committee (approval nos. 2015/6026 WALNUTs and 2020/9688–Equal-life). Written informed consent to participate in the original WALNUTs study was provided by the participants' legal guardian/next of kin. No additional consent was needed for this study, all of the participants were offered free tickets to the science museum of Barcelona. The specifics of the WALNUTs cohort formation were described in previous publications[21,65]. The current manuscript used a subset of 372 baseline blood samples before any dietary intervention originally described in a previous work[65]. For this study, a sub-group of 91 samples was used to perform the proteomics analysis. These samples were selected on the basis of the SDQ scores: 42 with the lowest SDQ score (SDQ = 0–14) and 49 with the highest SDQ score (SDQ = 15–25). Samples were drawn by a nurse using K2EDTA plus tubes, rested for 1 h and then centrifuged at 2,500 × *g* for 20 min at 20 °C, refrigerated at 4 °C, and frozen to −80 °C within 4 h after extraction[65], stored at −80 °C, and were not thawed until the protein depletion was performed before the proteomics analysis.

High-abundance protein depletion

Albumin and IgG represent more than 70% of total protein levels in human plasma samples. The depletion of high-abundant proteins is therefore essential to the identification and analysis of low-abundant proteins. A commercial kit (High Select Top14 Abundant Protein Depletion Mini Spin Columns, catalogue no. A36370, ThermoScientific) was used to deplete the 14 most abundant proteins from plasma before the proteomic analyses. The depleted proteins were human serum albumin, albumin, IgG, IgA, IgM, IgD, IgE, kappa and lambda light chains, α1-acidglycoprotein, α1-antitrypsin, α2-macroglobulin, 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 × *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 supported by Biocenter Finland.

Protein precipitation and digestion

Samples were acetone precipitated and subjected to in-solution digestion. Shortly, four volumes of ice-cold acetone were used to precipitate proteins. Precipitated proteins were resuspended to 8 M Urea, 50 mM Tris-HCl for protein denaturation, reduced with 5 mM dithiothreitol and alkylated with 13 mM iodoacetamide. Proteins were digested to peptides with trypsin (Promega) (enzyme:protein ratio 1:30) at 37 °C overnight. After digestion the peptides were desalted with a Sep-Pak C18 96-well plate (Waters), evaporated and stored at −20 °C.

Mass spectrometry analysis

Digested peptide samples were dissolved in 0.1% formic acid and peptide concentrations were determined with a NanoDrop device. Samples were spiked with iRT peptides (Biognosys) for retention time calibration. Equal amounts of samples were analyzed on a nanoflow HPLC system (Easy-nLC1200, Thermo Fisher Scientific) coupled to the Q Exactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a nano-electrospray ionization source. Peptides were first loaded onto a trapping column and subsequently separated in-line on a 15 cm C18 column (75 µm × 15 cm, ReproSil-Pur 3 µm 120 Å C18-AQ, Dr. Maisch HPLC GmbH). 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 100 min gradient was used to elute peptides (50 min from 5% to 21% solvent B followed by 40 min from 21% to 36 min solvent B). Mass spectrometry data were acquired automatically by using Thermo Xcalibur v.4.1 software (catalog no. OPTON-30965; Thermo Scientific). In a data-independent acquisition (DIA) method, a duty cycle contained one full scan (400–1,000 *m*/*z*) and 40 DIA MS/MS scans covering the mass range 400–1,000 with variable width isolation windows.

Protein identification and quantification analysis

Data analysis consisted of protein identifications and label-free quantifications of protein abundances. The data were analyzed using the Spectronaut software (Biognosys; v.17.1.221229). The direct DIA approach was used to identify proteins and label-free quantifications were performed with the MaxLFQ algorithm in Spectronaut. The main data analysis parameters in Spectronaut were: (1) enzyme (Trypsin/P); (2) up to two missed cleavages; (3) fixed modification (carbamidomethyl (cysteine)); (4) variable modifications (acetyl (protein *N*-terminus) and oxidation (methionine)); (5) the precursor FDR cutoff (0.01); (6) the protein FDR cutoff (0.01); (7) the quantification MS level (MS2); (8) the quantification type (area under the curve within integration boundaries for each targeted ion); (9) the protein database (Swiss-Prot 2022_05 Homo Sapiens[66] and Universal Protein Contaminant database[67]); and (10) normalization (global median normalization). All of the peptides were used for quantification.

Statistical analysis

Data pre-processing and statistical analyses were performed using R (v.4.2.1). Principal component analysis was performed to assess the general quality of the dataset (Supplementary Fig. 1). Identified proteins with more than 20% missing values were excluded from the analysis. Sample normalization was performed using the medianCentering method from the proBatch method[68]. Missing values remaining in the dataset were input using the sample minimum method[69]. We have compared the low SDQ and raised SDQ groups for the following variables: the education level for each parent, the levels of media consumption, levels of social media engagement, drug and alcohol use, and physical activity. We tested whether there are differences of the confounding factors between the groups using the one-way ANOVA test. After correcting for multiple comparisons, no socio-economic, sociodemographic or other factors showed any significant difference between the groups, meaning that the analyzed groups do not have significant differences in between the mean values of those factors.

Bioinformatic data analysis

The DeqMS (v.1.16.0) package was used for the differential abundance analysis[36], with SDQ score used as a continuous variable. The sex and age of the adolescents were included into the linear model to ensure that the proteins reported are associated with SDQ and were not influenced by confounding factors. The differences in protein abundances were expressed as log2-fold-change (the ratio of the means of raised (numerator) and low (denominator) SDQ groups). The *P*-values were adjusted using Benjamini–Hochberg procedure.

Plasma proteomic datasets from adolescents represent a very low number of all the proteomic datasets[70], so to better investigate the functional enrichment the full list of all proteins found in this study was used as the background gene list in the enrichment analyses. To characterize the enriched pathways related to the identified proteins, significantly differently abundant proteins ($P \leq 0.05$) associated with the SDQ score were used in further bioinformatic data analyses. Proteins depleted before mass spectrometry analysis that showed significant differences between groups were considered a possible source of bias and

thus were excluded. The Reactome pathways were investigated using the ReactomePA (v.1.9.4) R package[38]. The enrichment was performed using the enrichPathway function of the ReactomePA package, which uses the hypergeometric model. The *P*-value was adjusted using the Benjamini–Hochberg method. We used IPA (Ingenuity Systems) for the further enrichment analyses In IPA core analysis, default software parameters were used (reference set: ingenuity knowledge base—genes only). The *z*-score values were used to identify canonical pathways that were expected to be changed by their activity. The STRINGdb package was used to get the protein–protein interaction information for the significantly differentially abundant proteins from the STRING database (v.11.5)[37]. The fastgreedy clustering function was used to extract gene clusters with strong associations. A novel symbolic-regression-based algorithm, QLattice, which is part of the Feyn package (v.3.0.3), was used to generate models combining proteins with the best predictive power for the SDQ score based on protein biomarkers[39]. The algorithm was used to find the models combining proteins with the best predictive power. Possible biomarkers were searched among the 58 proteins significantly associated with the SDQ score, with five rounds of cross-validation. The resulting models were identified as the top models in each of the cross-validation. Result visualizations were performed using ggplot2 (v.3.4.0)[71] and ComplexHeatmap[72] (v.2.14.0) packages.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data analyzed in this study are subject to the following licenses/restrictions: the WALNUTs data art not publicly available due to the restrictions of informed consent. The data contain personal information on children and, according to the ethical approval, they should be kept confidential. Data are available from the corresponding author on reasonable request for researchers who meet the criteria for access to confidential data. A data-use/transfer agreement is needed to ensure the protection of privacy and compliance with national data protection legislation, the content and specific clauses of which will depend on the nature of the requested data.

References

1. Juraska, J. M. & Willing, J. Pubertal onset as a critical transition for neural development and cognition. *Brain Res.* **1654**, 87–94 (2017).
2. Crone, E. A. & Dahl, R. E. Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nat. Rev. Neurosci.* **13**, 636–650 (2012).
3. Blakemore, S.-J. & Mills, K. L. Is adolescence a sensitive period for sociocultural processing? *Annu. Rev. Psychol* **65**, 187–207 (2014).
4. Blakemore, S. J. Adolescence and mental health. *Lancet* **393**, 2030–2031 (2019).
5. Solmi, M. et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol. Psychiatry* **27**, 281–295 (2022).
6. Preti, A. et al. The epidemiology of eating disorders in six European countries: results of the ESEMeD–WMH project. *J. Psychiatr. Res.* **43**, 1125–1132 (2009).
7. Kessler, R. C. et al. Age of onset of mental disorders: a review of recent literature. *Curr. Opin. Psychiatry* **20**, 359–364 (2007).
8. Mental health matters. *Lancet Glob. Health* **8**, e1352 (2020).
9. Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A. & Rohde, L. A. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *J. Child Psychol. Psychiatry.* **56**, 345–365 (2015).
10. *Adolescent Mental Health* (World Health Organization, 2021); https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health
11. Schnyder, N. et al. Perceived need and barriers to adolescent mental health care: agreement between adolescents and their parents. *Epidemiol. Psychiatr. Sci.* **29**, e60 (2019).
12. Islam, M. I. et al. The gap between perceived mental health needs and actual service utilization in Australian adolescents. *Sci Rep.* **12**, 5430 (2022).
13. Merten, E. C., Cwik, J. C., Margraf, J. & Schneider, S. Overdiagnosis of mental disorders in children and adolescents (in developed countries). *Child Adolesc. Psychiatry Ment. Health* **11**, 5 (2017).
14. First, M. B. et al. An organization- and category-level comparison of diagnostic requirements for mental disorders in ICD-11 and DSM-5. *World Psychiatry* **20**, 34–51 (2021).
15. Wang, Y. P. & Gorenstein, C. Assessment of depression in medical patients: a systematic review of the utility of the Beck Depression Inventory-II. *Clinics* **68**, 1274–1287 (2013).
16. Drabick, D. A. G. & Kendall, P. C. Developmental psychopathology and the diagnosis of mental health problems among youth. *Clin. Psychol.* **17**, 272–280 (2010).
17. O'Connor, C., Downs, J., Shetty, H. & McNicholas, F. Diagnostic trajectories in child and adolescent mental health services: exploring the prevalence and patterns of diagnostic adjustments in an electronic mental health case register. *Eur. Child Adolesc. Psychiatry* **29**, 1111–1123 (2020).
18. O'Connor, C., Kadianaki, I., Maunder, K. & McNicholas, F. How does psychiatric diagnosis affect young people's self-concept and social identity? A systematic review and synthesis of the qualitative literature. *Soc. Sci. Med.* **212**, 94–119 (2018).
19. Goodman, R., Meltzer, H. & Bailey, V. The strengths and difficulties questionnaire: a pilot study on the validity of the self-report version. *Int. Rev. Psychiatry* **15**, 173–177 (2003).
20. Ortuño-Sierra, J. et al. Screening mental health problems during adolescence: psychometric properties of the Spanish version of the Strengths and Difficulties Questionnaire. *J. Adolesc.* **38**, 49–56 (2015).
21. Julvez, J. et al. Walnuts, long-chain polyunsaturated fatty acids, and adolescent brain development: protocol for the walnuts smart snack dietary intervention trial. *Front. Pediatr.* **9**, 593847 (2021).
22. Goodman, A. & Goodman, R. Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. *J. Child Psychol. Psychiatry* **52**, 100–108 (2011).
23. Goodman, A. & Goodman, R. Strengths and difficulties questionnaire as a dimensional measure of child mental health. *J. Am. Acad. Child Adolesc. Psychiatry* **48**, 400–403 (2009).
24. Rodríguez Cerdeira, C., Sánchez-Blanco, E., Sánchez-Blanco, B. & González-Cespón, J. L. Protein biomarkers of mood disorders. *Int. J. Immunopathol. Pharmacol.* **30**, 7–12 (2017).
25. García-Gutiérrez, M. S. et al. Biomarkers in psychiatry: concept, definition, types and relevance to the clinical reality. *Front. psychiatry* **11**, 432 (2020).
26. Cagney, D. N. et al. The FDA NIH biomarkers, endpoints, and other tools (BEST) resource in neuro-oncology. *Neuro. Oncol.* **20**, 1162–1172 (2018).
27. Turck, C. W. et al. Proteomic differences in blood plasma associated with antidepressant treatment response. *Front. Mol. Neurosci.* **10**, 272 (2017).
28. Guest, P. C., Guest, F. L. & Martins-de Souza, D. Making sense of blood-based proteomics and metabolomics in psychiatric research. *Int. J. Neuropsychopharmacol.* **19**, pyv138 (2016).
29. Rodrigues-Amorim, D. et al. Proteomics in schizophrenia: a gateway to discover potential biomarkers of psychoneuroimmune pathways. *Front. psychiatry* **10**, 885 (2019).

30. Hartwig, F. P., Borges, M. C., Horta, B. L., Bowden, J. & Davey Smith, G. Inflammatory biomarkers and risk of schizophrenia: a 2-sample mendelian randomization study. *JAMA Psychiatry* **74**, 1226–1233 (2017).
31. Nilsson, I. A. K. et al. Plasma neurofilament light chain concentration is increased in anorexia nervosa. *Transl. Psychiatry* **9**, 180 (2019).
32. Fernandes, B. S., Dai, Y., Jia, P. & Zhao, Z. Charting the proteome landscape in major psychiatric disorders: from biomarkers to biological pathways towards drug discovery. *Eur. Neuropsychopharmacol.* **61**, 43–59 (2022).
33. Mongan, D. et al. Development of proteomic prediction models for transition to psychotic disorder in the clinical high-risk state and psychotic experiences in adolescence. *JAMA Psychiatry* **78**, 77–90 (2021).
34. English, J. A. et al. Blood-based protein changes in childhood are associated with increased risk for later psychotic disorder: evidence from a nested case–control study of the ALSPAC longitudinal birth cohort. *Schizophr. Bull.* **44**, 297 (2018).
35. Goodman, A. *Scoring the Strengths & Difficulties Questionnaire for age 4–17 or 18+* (2016);https://terapia.co.uk/wp-content/uploads/2020/05/SDQ-scoring_Instructions_4-18-years.pdf
36. Zhu, Y. et al. DEqMS: A method for accurate variance estimation in differential protein expression analysis. *Mol. Cell. Proteomics* **19**, 1047–1057 (2020).
37. Szklarczyk, D. et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).
38. Yu, G. & He, Q.-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479 (2016).
39. Christensen, N. J. et al. Identifying interactions in omics data for clinical biomarker discovery using symbolic regression. *Bioinformatics* **38**, 3749–3758 (2022).
40. Becker, A., Hagenberg, N., Roessner, V., Woerner, W. & Rothenberger, A. Evaluation of the self-reported SDQ in a clinical setting: do self-reports tell us more than ratings by adult informants?. *Eur. Child Adolesc. Psychiatry* **13**, 17–24 (2004).
41. Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A. & Janssens, J. M. A. M. Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. *Clin. Child Fam. Psychol. Rev.* **13**, 254–274 (2010).
42. Nielsen, L. G. et al. The predictive validity of the Strengths and Difficulties Questionnaire in preschool age to identify mental disorders in preadolescence. *PLoS ONE* **14**, e0217707 (2019).
43. Madrid-Gambin, F. et al. Integrated lipidomics and proteomics point to early blood-based changes in childhood preceding later development of psychotic experiences: evidence from the avon longitudinal study of parents and children. *Biol. Psychiatry* **86**, 25 (2019).
44. Föcking, M. et al. Complement pathway changes at age 12 are associated with psychotic experiences at age 18 in a longitudinal population-based study: evidence for a role of stress. *Mol. Psychiatry* **26**, 524–533 (2021).
45. Comes, A. L. et al. Proteomics for blood biomarker exploration of severe mental illness: pitfalls of the past and potential for the future. *Transl. Psychiatry* **8**, 160 (2018).
46. Chan, M. K. et al. Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. *Transl. Psychiatry* **5**, e601 (2015).
47. Ziani, P. R. et al. Potential candidates for biomarkers in bipolar disorder: a proteomic approach through systems biology. *Clin. Psychopharmacol. Neurosci.* **20**, 211–227 (2022).
48. Jiang, J. et al. Leukocyte proteomic profiling in first-episode schizophrenia patients: does oxidative stress play central roles in the pathophysiology network of schizophrenia? *Antioxidants Redox Signal.* **31**, 579–588 (2019).
49. Silva, J. V. et al. Amyloid precursor protein interaction network in human testis: sentinel proteins for male reproduction. *BMC Bioinf.* **16**, 12 (2015).
50. Erdinger, S. et al. Lack of APLP1 leads to subtle alterations in neuronal morphology but does not affect learning and memory. *Front. Mol. Neurosci.* **15**, 1028836 (2022).
51. Schilling, S. et al. APLP1 is a synaptic cell adhesion molecule, supporting maintenance of dendritic spines and basal synaptic transmission. *J. Neurosci.* **37**, 5345–5365 (2017).
52. Pandolfo, G. et al. Mental illness and amyloid: a scoping review of scientific evidence over the last 10 years (2011 to 2021). *Brain Sci.* **11**, 1352 (2021).
53. Novak, G., Kim, D., Seeman, P. & Tallerico, T. Schizophrenia and Nogo: elevated mRNA in cortex, and high prevalence of a homozygous CAA insert. *Mol. Brain. Res.* **107**, 183–189 (2002).
54. Wang, J. et al. RTN4/NoGo-receptor binding to BAI adhesion-GPCRs regulates neuronal development. *Cell* **184**, 5869–5885. e25 (2021).
55. Yasuda, R., Hayashi, Y. & Hell, J. W. CaMKII: a central molecular organizer of synaptic plasticity, learning and memory. *Nat. Rev. Neurosci.* **23**, 666–682 (2022).
56. Colbran, R. J. & Brown, A. M. Calcium/calmodulin-dependent protein kinase II and synaptic plasticity. *Curr. Opin. Neurobiol.* **14**, 318–327 (2004).
57. Rose, A. J., Kiens, B. & Richter, E. A. Ca^{2+}-calmodulin-dependent protein kinase expression and signalling in skeletal muscle during exercise. *J. Physiol.* **574**, 889–903 (2006).
58. Willi, R. & Schwab, M. E. Nogo and Nogo receptor: relevance to schizophrenia? *Neurobiol. Dis.* **54**, 150–157 (2013).
59. Dave, B. P. et al. Unveiling the modulation of Nogo receptor in neuroregeneration and plasticity: novel aspects and future horizon in a new frontier. *Biochem. Pharmacol.* **210**, 115461 (2023).
60. Lazar, N. L. et al. Missense mutation of the Reticulon-4 receptor alters spatial memory and social interaction in mice. *Behav. Brain Res.* **224**, 73–79 (2011).
61. Frei, J. A. et al. Regulation of neural circuit development by Cadherin-11 provides implications for autism. *eNeuro* https://doi.org/10.1523/ENEURO.0066-21.2021 (2021).
62. Wu, N., Wang, Y., Jia, J.-Y., Pan, Y.-H. & Yuan, X.-B. Association of CDH11 with autism spectrum disorder revealed by matched-gene co-expression analysis and mouse behavioral studies. *Neurosci. Bull.* **38**, 29–46 (2022).
63. Butler, M., Rafi, S. & Manzardo, A. High-resolution chromosome ideogram representation of currently recognized genes for autism spectrum disorders. *Int. J. Mol. Sci.* **16**, 6464–6495 (2015).
64. Khan, S. R., Manialawy, Y., Wheeler, M. B. & Cox, B. J. Unbiased data analytic strategies to improve biomarker discovery in precision medicine. *Drug Discov. Today* **24**, 1735–1748 (2019).
65. Pinar-Martí, A. et al. Red blood cell omega-3 fatty acids and attention scores in healthy adolescents. *Eur. Child Adolesc. Psychiatry* https://doi.org/10.1007/s00787-022-02064-w (2022).
66. Bateman, A. et al. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
67. Frankenfield, A. M., Ni, J., Ahmed, M. & Hao, L. Protein contaminants matter: building universal protein contaminant libraries for DDA and DIA proteomics. *J. Proteome Res.* **21**, 2104–2113 (2022).
68. Čuklina, J. et al. Diagnostics and correction of batch effects in large-scale proteomic studies: a tutorial. *Mol. Syst. Biol.* **17**, e10240 (2021).

69. Liu, M. & Dongre, A. Proper imputation of missing values in proteomics datasets for differential expression analysis. *Brief. Bioinform.* **22**, bbaa112 (2021).
70. Ignjatovic, V. et al. Mass spectrometry-based plasma proteomics: considerations from sample collection to achieving translational data. *J. Proteome Res.* **18**, 4085–4097 (2019).
71. Wickham, H. *ggplot2*. *ggplot2* (Springer, 2009); https://doi.org/10.1007/978-0-387-98141-3
72. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

Acknowledgements

We gratefully acknowledge the contribution of the Early Environmental Quality and Life-course Mental Health Effects (Equal-Life) project team, R. Koole and R. Bogers (RIVM) in providing the selection protocol, the Turku Proteomics Facility team supported by Biocenter Finland for mass spectrometry, and the laboratory assistance of M. Tikkanen. This WALNUTs project was supported by Instituto de Salud Carlos III through the projects 'CP14/00108, PI16/00261, and PI21/00266' (co-founded by the European Regional Development Fund 'A way to make Europe'). J.J. holds a Miguel Servet-II contract (grant CPII19/00015) awarded by the Instituto de Salud Carlos III (confounded by the European Social Fund 'Investing in your future'). The California Walnut Commission (CWC) gave support by supplying the walnuts for free for the WALNUTs Smart Snack Dietary Intervention Trial. The funders had no role in the study design, collection, management, analysis, and interpretation of data, the writing of the report, or the decision to submit it for publication. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 874724. Equal-Life is part of the European Human Exposome Network.

Author contributions

K.M.K., I.d.S.M. and I.v.K. designed the study. I.v.K., M.F. and J.J. guided the selection process and provided the samples. I.d.S.M. and A.A. pre-processed the samples. I.d.S.M., A.-K.P. and A.M.A. performed statistical and bioinformatics analyses. M.I. was involved in pathway analysis. K.K.J. participated in planning and literature analysis. I.d.S.M., A.-K.P., A.M.A., M.I. and K.M.K. drafted the manuscript. All of the authors read and approved the final version of the manuscript.

FundingInformation

Open access funding provided by University of Eastern Finland (UEF) including Kuopio University Hospital.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44220-023-00103-2.

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

Peer review information *Nature Mental Health* thanks Joao Rodrigues, Patricia Silveira and the other, anonymous reviewers for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Katja M. Kanninen

Last updated by author(s): 2023/06/10

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was analyzed by Spectronaut software (Biognosys; version 17.0. 17.1.221229). The direct DIA approach was used to identify proteins and label-free quantifications were performed with the MaxLFQ algorithm in Spectronaut. Main data analysis parameters in Spectronaut were: (i) Enzyme: Trypsin/P; (ii) Fixed modifications: Carbamidomethyl (cysteine); (iii) Variable modifications: Acetyl (protein N-terminus) and oxidation (methionine); (iv) Protein database: Homo sapiens Swiss-Prot reference proteome (Uniprot release 2021_02); and (v) Normalization: Global median normalization. All the peptides were used for quantification.
Data analysis	Data pre-processing and statistical analyses were performed using R (version 4.2.1.). For the differential abundance analysis DeqMS (v. 1.16.0) package was used 67, with SDQ score used as a continuous variable. QLattice a part of the Feyn package (v. 3.0.3), a symbolic-regression-based ML algorithm was used to build models to find the proteins with the best predictive power. Result visualisations were performed using ggplot2 (v3.4.0) and ComplexHeatmap (v.2.14.0) packages.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The String-DB (v. 11.5) database (https://string-db.org/) was used for data annotation.
The Reactome (v. 83) database (https://reactome.org/) was used for data annotation.
Ingenuity Pathway analysis was performed in this study, using the IPA

The data analysed in this study is subject to the following licenses/restrictions: The Walnuts data is not publicly available due to the restrictions of informed consent. The data contain personal information of children and according to the ethical approval, they should be kept confidential. Data are available from the corresponding author upon reasonable request for researchers who meet the criteria for access to confidential data. To ensure the protection of privacy and compliance with national data protection legislation, a data use/transfer agreement is needed, the content and specific clauses of which will depend on the nature of the requested data.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	The biological sex information was obtained from school databases for the original WALNUTs manuscript.
Population characteristics	The peripheral blood plasma samples were obtained and analyzed from a subsample of 91 adolescents aged 11-16 years, of the WALNUTs regional Spanish study (Table 1). The samples were collected in 2016-2018 approximately at the same moment as the participants filled out the Strengths and Difficulties Questionnaire (SDQ). The SDQ is a screening questionnaire for emotional and behavioral problems in children and young people assessing the impact of difficulties on the child's life, such as (i) emotional symptoms, (ii) conduct problems, (iii) hyperactivity/inattention, (iv) peer relationship problems, and (v) prosocial behavior. Based on the self-reported SDQ score, the plasma samples were categorized into low (0-14) and raised (>15-17) groups.
Recruitment	The original WALNUTs study performed age-, gender-, and maternal education-stratified random computerized sampling within each school to assign adolescents to one of the two groups.
Ethics oversight	The studies were reviewed and approved by CEIC Parc Salut Mar Clinical Research Ethics Committee (approval numbers: 2015/6026 Walnuts and 2020/9688–Equal-life). Written informed consent to participate in the original WALNUTs study was provided by the participants' legal guardian/next of kin. No additional consent was needed for this study

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This is an exploratory study for the investigation of possible connections between the SDQ and abundancies of plasma proteins in adolescents. All the individuals from with raised SDQ score were analysed in this study, and an equal number of individuals with low SDQ score were also included as the control group.
Data exclusions	No data was excluded from the analysis.
Replication	Since the samples were taken from a cohort study, all the samples were analysed. There was no other comparable cohort available. There was no longitudinal data available.
Randomization	The original study was randomised. The samples were also randomised prior to protein sequencing.
Blinding	The SDQ score was calculated after the plasma samples were taken. The samples were anonymised for the protein depletion and the proteomics analysis step.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging