plasma proteome, combined with easy accessibility of blood, makes plasma proteins promising candidates for disease diagnosis and risk prediction[9,10]. For example, a recent study demonstrated the ability to predict dementia by leveraging a combination of key plasma biomarkers and basic demographic indicators, achieving accurate predictions more than a decade before clinical diagnosis[11].

Despite these advantages, research specifically focused on plasma protein profiles in relation to SBs remains limited. For example, a recent study involving 288 patients with bipolar disorder analyzed 92 plasma proteins and identified 16 potential biomarkers associated with future suicide attempts using a multivariate classification approach[12]. Of these, three proteins (SCGB1A1, ANXA10 and CETN2) demonstrated significant associations in individual analyses, although the overall effect sizes observed were modest. Similarly, an investigation of 39 peripheral cells and proteins related to immune function, oxidative stress and plasticity in 266 patients with mood disorders reported that elevated plasma levels of TSP-2 (thrombospondin-2) and C-reactive protein were associated with past-month suicide attempts, while higher levels of TSP-1 and PDGF-AB (platelet-derived growth factor-AB) predicted future attempts[13]. These findings suggest potential temporal differences in peripheral biomarker profiles across different stages of the suicidal process. The role of inflammatory biomarkers in suicidality was further supported by a Mendelian randomization (MR) study, which demonstrated that genetic upregulation of IL-6 (interleukin-6) signaling was associated with suicidal ideation[14]. More recently, a large-scale proteomic analysis of 4,719 blood-derived proteomes from a cohort of 35,559 individuals identified one protein, platelet endothelial aggregation receptor 1 (PEAR1), as potentially causally associated with suicide attempts through MR analysis[15]. However, limited evidence of colocalization for PEAR1 reduces confidence in its relevance to suicide attempts. Taken together, these studies implicate inflammatory and other proteins in SBs but remain constrained by small sample sizes, modest effect sizes and incomplete mechanistic insight. These limitations highlight the need for larger-scale proteomic studies using well-characterized case–control datasets to more comprehensively define plasma proteome profiles associated with SBs and to develop reliable biomarkers for SB risk prediction.

To address these research gaps, we conducted a comprehensive proteomic analysis within a large cohort from the UK Biobank, including 2,920 plasma protein measurements from over 50,000 adults, with plasma proteins in classifying past SBs and predicting future occurrences (see Fig. 1 for an overview of the study). Collectively, these integrative approaches may provide novel insights into the biological mechanisms underlying SBs and facilitate the identification of biomarkers that could inform precision prevention and guide the development of targeted therapeutic strategies.

Results

A total of 53,026 individuals (mean age 56.81 years; standard deviation (s.d.) 8.21; 53.91% women; 93.69% white) with measurements of 2,920 unique plasma proteins were included in our analysis. Of these, 268 individuals had engaged in SBs before protein data collection (cross-sectional cases), while 202 individuals developed SB during a follow-up period of up to 13.46 years after protein data collection (median 4.58 years; longitudinal cases). Baseline sociodemographic and clinical characteristics of cases and controls in both the cross-sectional and longitudinal analyses are provided in Supplementary Table 1.

Identification of plasma proteins associated with SBs

Cross-sectional associations. To identify plasma proteins associated with SBs, we first examined the associations between 2,920 unique plasma protein levels and SBs using logistic regression analyses in cross-sectional data. The analyses were adjusted for sex, age, assessment centers, ethnicity, socioeconomic status, body mass index (BMI), protein batch, the time lag between blood sample collection and proteomic testing, smoking status and six categories of disease diagnoses at baseline. Results revealed that 421 (14.42%) proteins were significantly associated with SBs after false discovery rate (FDR) correction for multiple comparisons ($\beta = -1.30$ to 1.49, β are regression coefficients throughout, FDR-adjusted P value < 0.05, $P_{FDR} = 6.23 \times 10^{-18}$ to 0.049; Fig. 2a). Of these, 388 proteins were positively correlated with SBs, while 33 were negatively correlated (Fig. 2a).

Among the identified proteins, PLAUR demonstrated the strongest association with SBs ($\beta = 1.49$, $P_{FDR} = 6.23 \times 10^{-18}$), followed by LAMP3 ($\beta = 0.70$, $P_{FDR} = 4.68 \times 10^{-13}$), WFDC2 ($\beta = 0.83$, $P_{FDR} = 6.56 \times 10^{-12}$), PIGR ($\beta = 0.96$, $P_{FDR} = 7.77 \times 10^{-12}$) and GDF15 ($\beta = 0.59$, $P_{FDR} = 1.19 \times 10^{-10}$). A full list of protein identifiers with their full names is provided in Supplementary Data Table 1, and the complete results of the cross-sectional protein analysis are available in Supplementary Data Table 2.

nature mental health

Article

https://doi.org/10.1038/s44220-025-00582-5

Plasma proteomic profiles linked to suicidal behaviors

Received: 22 January 2025

Accepted: 4 December 2025

Published online: 23 February 2026

Check for updates

Bei Zhang[1,12], Jia You[2,3,12], Edmund T. Rolls[2,4], Peng Ren[2,3], Yuzhu Li[2,3], Wei Zhang[2,3], Barbara J. Sahakian[2,5,6], Fei Li[7], Jianfeng Feng[2,3,4,8,9] ✉ & Wei Cheng[2,3,9,10,11] ✉

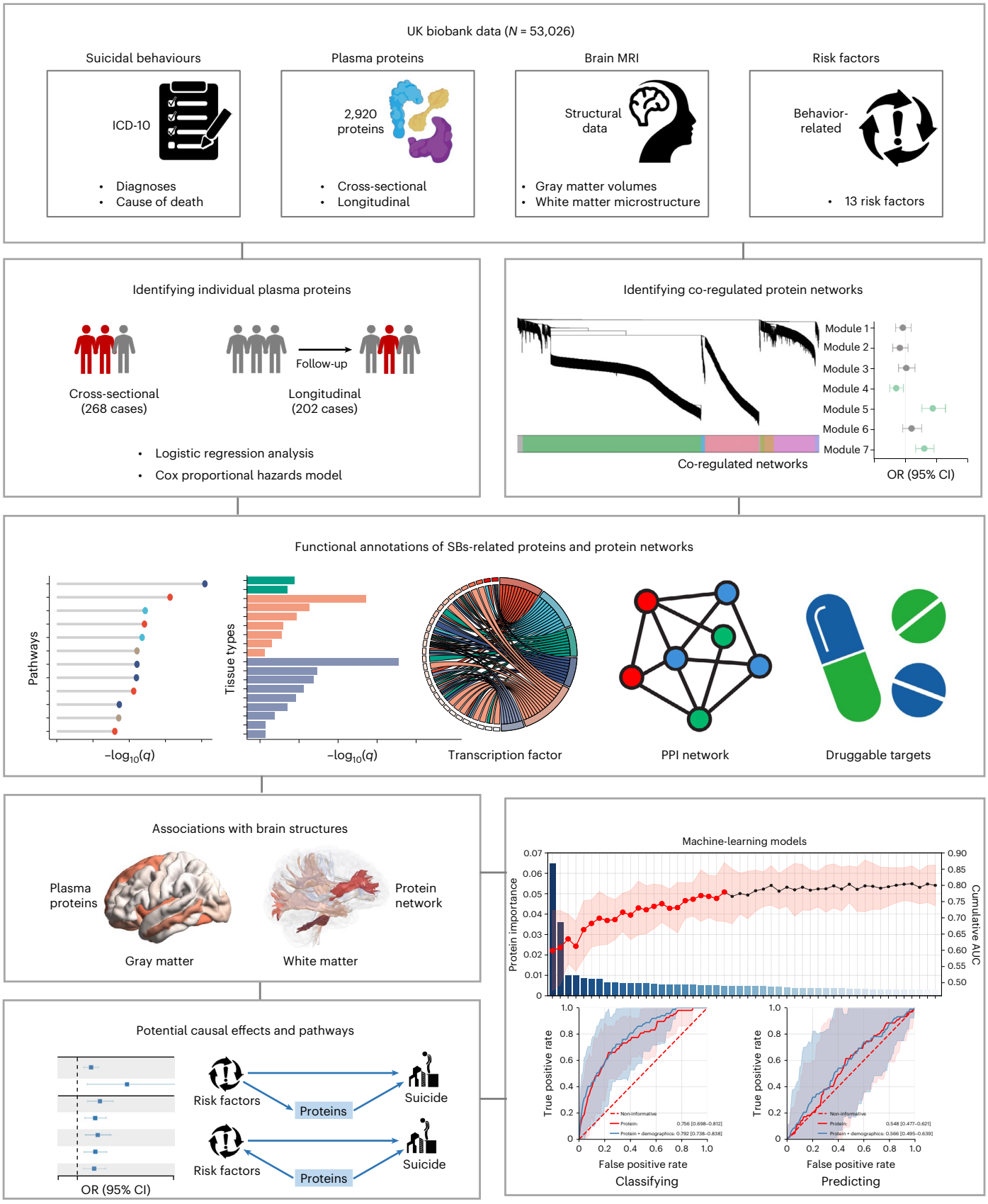
Suicidality is a major public health concern[1], e
ideation, suicide attempt and death by suicide[2].
fatal outcomes, suicidal behaviors (SBs, includi
death by suicide) have become a critical focus of
the biopsychosocial model, SBs are influenced
of psychological, social and biological factors
factors for SBs have been identified[5], their predi
ited, particularly in accurately determining whc
in suicidal behaviors[6]. In addition, traditional su
which rely heavily on psychological evaluation

A full list of affiliations appears at the end of the pa

https://doi.org/10.1038/s44220-025-00582-5

Functional annotation of SBs-related proteins

Enrichment analysis. To elucidate the biological relevance and regulatory pathways of SBs-associated proteins, we conducted three types of enrichment analyses on the gene-set encoding the 421 significant proteins identified in cross-sectional associations. First, pathway enrichment analysis using six biological pathway databases available on the Metascape platform revealed significant FDR-corrected enrichment in ten pathways ($P_{FDR} = 7.45 \times 10^{-6}$ to 0.027), including cytokine–cytokine receptor interaction, tumor necrosis factors (TNFs) binding to their physiological receptors, NABA



matrisome-associated, antimicrobial humoral response, and pancreatic cancer subtypes (Fig. 2f and Supplementary Data Table 5). Second, tissue-specific enrichment using the GTEx v8 database showed significant FDR-corrected enrichment among up-regulated differentially expressed genes (DEGs) from the lung, salivary gland, vagina, colon, adipose tissue, small intestine, breast, blood and stomach, as well as among down-regulated DEGs from testis and pituitary tissues ($P_{FDR} = 2.21 \times 10^{-12}$ to 0.045; Fig. 2d and Supplementary Data Table 6). Third, transcription factors (TFs) analysis using the TRRUST (transcriptional regulatory relationships unraveled by sentence-based text-mining) database identified 71 TFs significantly involved in regulating these proteins after FDR correction ($P_{FDR} = 5.87 \times 10^{-19}$ to 0.050). The top TFs were NFKB1 (the nuclear factor of kappa light polypeptide gene enhancer in B-cells 1), RELA (the v-rel reticuloendotheliosis viral oncogene homolog A), STAT3 (the signal transducer and activator of transcription 3), JUN (the jun proto-oncogene), SP1 (the Sp1 transcription factor), and STAT1 (the signal transducer and activator of transcription 1) (Fig. 2e, Supplementary Data Table 7).

PPI network analysis. Protein–protein interaction (PPI) was conducted to explore the interconnections among the 421 SBs-related proteins using the STRING database. By applying a stringent confidence threshold of 0.9, which corresponds to the highest level of interaction confidence in STRING and indicates that the retained interactions are supported by very strong evidence from multiple sources, we generated a robust network of 417 nodes and 184 edges, demonstrating the extensive interconnectivity of these proteins. To prioritize key proteins within the network, we ranked nodes by degree centrality, which reflects the number of direct connections each protein has and is widely used as an importance score in network analyses. IL-6, a key pro-inflammatory cytokine, was identified as the top-ranked hub protein (Fig. 2g).

Druggability assessment. To explore potential druggable targets, SBs-associated proteins were mapped to DrugBank and the Therapeutic Target Database. Targets were annotated according to either the International Classification of Diseases, Tenth Revision (ICD-10) or the Anatomical Therapeutic Chemical (ATC) system. Using ICD-10 categories, we identified significant enrichment across 21 disease groups, including inflammatory polyarthropathies and malignant neoplasms-related conditions. Notably, inflammatory polyarthropathies showed the strongest enrichment, driven by 23 genes such as *IL-6* and *TNFRSF11* (Supplementary Data Table 8). Using the ATC system, 18 genes, including *IL-6* and *TNFRSF11*, were mapped to existing drugs, primarily antineoplastic and immunomodulating agents. However, this enrichment reached only marginal statistical significance ($P = 0.09$; Supplementary Data Table 8).

Identification of co-regulated protein networks associated with SBs

In addition to examining individual plasma proteins, we grouped the full set of 2,920 plasma proteins into co-regulated networks according to their co-expression patterns using a data-driven approach (Netboost)[16,17]. This analysis identified seven non-overlapping protein networks, with sizes ranging from 35 to 1,726 proteins (Fig. 3a). We then quantified the association between the expression of each network and SBs in both cross-sectional and longitudinal datasets, adjusting for covariates, including sex, age, assessment center, ethnicity, socioeconomic status, and BMI, smoking status and six categories of baseline disease diagnoses. In the cross-sectional analyses, three protein networks were significantly associated with SBs after FDR correction (Fig. 3b and Supplementary Data Table 9): module 5 (odds ratio (OR) = 1.43, $P_{FDR} = 2.50 \times 10^{-7}$, hub proteins: IFNGR1, CD59, BSG, BTN2A1, CLMP, COLEC12), module 7 (OR = 1.30, $P_{FDR} = 1.06 \times 10^{-5}$, hub proteins: DSG3, SPINK5, TACSTD2, CST6, CA9, DSG4) and module 4 (OR = 0.86, $P_{FDR} = 0.032$, hub proteins: NOTCH1, ALCAM, NOTCH3, LTBP2, ADGRE5, ADGRE2). However, no significant associations with SBs were observed for any of the networks in the longitudinal datasets (Fig. 3c,d).

To elucidate the functional roles of these networks, we performed pathway enrichment analyses on the gene-set encoding proteins within the three identified networks/modules. The gene-set encoding proteins in module 5 were significantly enriched in 20 pathways, most of which were related to inflammatory processes ($P_{FDR} = 4.95 \times 10^{-8}$ to 0.011; Fig. 3e), including cytokine–cytokine receptor interaction, immunoregulatory interactions between lymphoid and non-lymphoid cells, and regulation of lymphocyte-mediated immunity. By contrast, proteins in module 4 were significantly enriched in pathways related to cell–cell adhesion ($P_{FDR} = 0.001$; Fig. 3f). No significant pathways were observed for the proteins in module 7.

Associations of SBs-related proteins and protein networks with brain structures

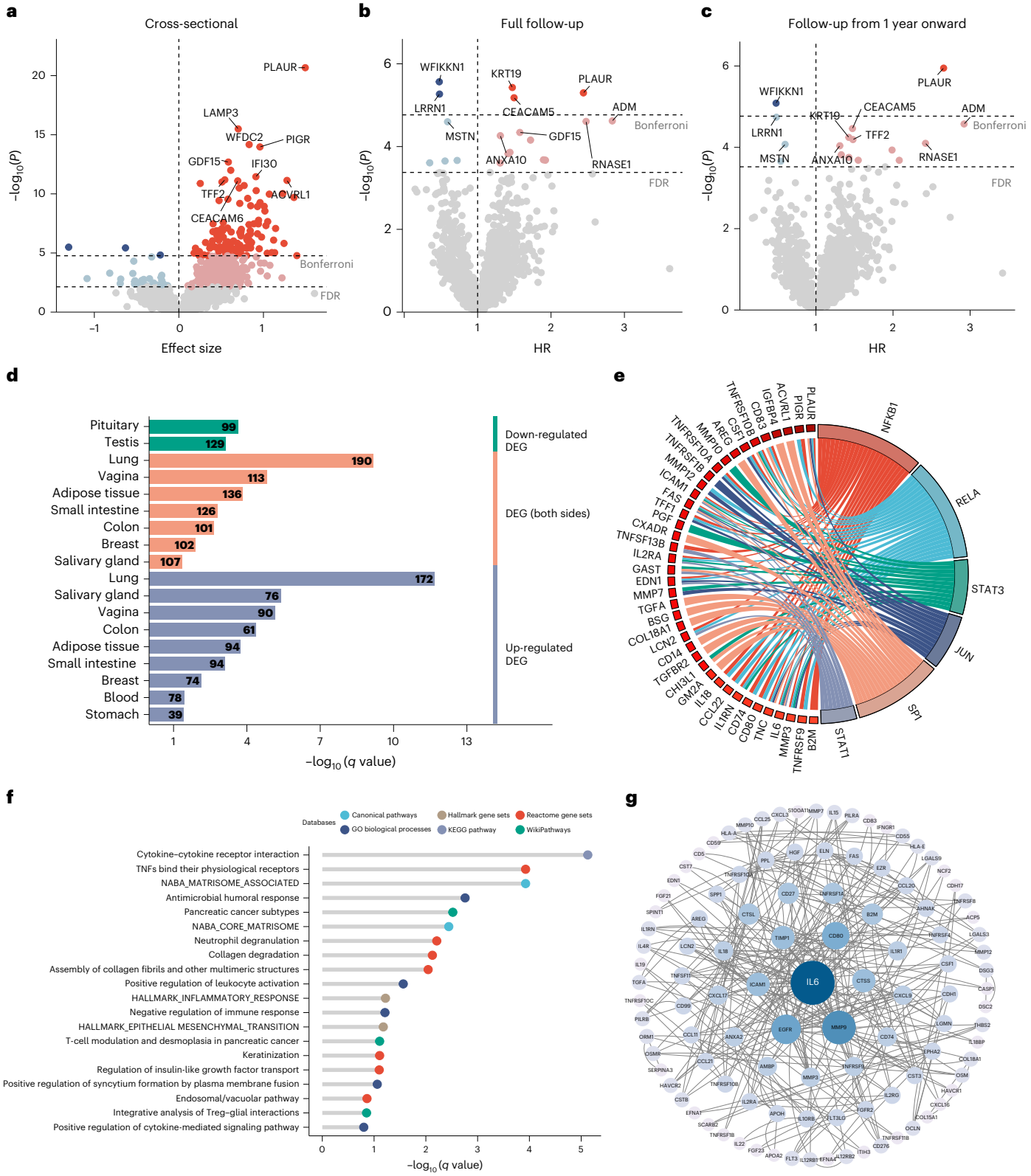
Protein associations. To assess the relationship between SBs-related proteins and brain structure measurements, we conducted linear regression analyses examining 421 proteins in relation to 82 gray matter volumes and 135 white matter microstructures. For gray matter volumes, 381 protein–gray matter pairs were found to be significant after FDR correction ($P_{FDR} = 3.13 \times 10^{-4}$ to 0.049), with 88 positive and 293 negative correlations. Notably, 62 out of the 82 gray matter volumes exhibited associations with at least one SBs-related protein. Cortical

Fig. 2 | Proteins associated with SBs and their functional highlights. a, Cross-sectional associations between plasma proteins and SBs. The *x* axis represents the effect size (β), and the *y* axis represents $-\log_{10}$ of uncorrected *P* value resulting from two-sided logistic regression tests between SBs and each plasma protein. Positive associations are shown in red and negative associations in blue. The dashed lines indicate significance thresholds after multiple testing correction using Bonferroni correction (top) and FDR correction (bottom), respectively. The top ten significant proteins are labeled (FDR-adjusted *P* value < 0.05). **b**,**c**, Longitudinal associations between plasma proteins and incident SBs during the entire follow-up period (**b**) and starting from 1 year onward (**c**). The *x* axis represents the HR, and the *y* axis represents $-\log_{10}$ of uncorrected *P* value resulting from two-sided Cox proportional hazard models between SBs and each plasma protein. Positive associations are shown in red and negative associations in blue. The dashed lines indicate significance thresholds after multiple testing correction using Bonferroni correction (top) and FDR correction (bottom), respectively. The top significant proteins are labeled (FDR-adjusted *P* value < 0.05). **d**, Tissue-specific enrichment analysis for the gene-set encoding SBs-related proteins using the GTEx v8 database. Green, blue and orange bars indicate down-regulated, up-regulated and dually regulated DEGs across various tissues, respectively. Numbers on the bars indicate the number of genes involved in this tissue-specific enrichment. The *x* axis represents the $-\log_{10}$ of the *q* value for each tissue, reflecting the level of statistical significance after FDR correction. **e**, Transcription factor enrichment analysis for genes encoding SBs-associated proteins using the TRRUST database. The top six transcription factors are annotated. Each enriched transcription factor is linked to its related gene. The red color bar indicates the significance of the cross-sectional association between plasma proteins and SBs, with darker colors denoting higher levels of significance. **f**, Pathway enrichment analysis of the gene-set encoding SBs-related proteins using six databases within the Metascape platform: Gene Ontology (GO) biological processes, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, Reactome pathways, WikiPathways, Hallmark pathways and Canonical pathways. The *x* axis represents the $-\log_{10}$ of the *q* value for each pathway, and the *y* axis lists different pathways, distinguished by color coding based on the source database. **g**, Protein–protein interaction network of SBs-associated proteins. Each node represents a protein, with darker shades indicating higher importance. The edges connecting nodes represent interaction strength, with thicker edges indicating stronger interactions according to the combined score value.

https://doi.org/10.1038/s44220-025-00582-5

regions such as the insula, superior frontal cortex, fusiform, middle temporal cortex, and medial and lateral orbitofrontal cortex, along with subcortical regions, including thalamus, pallidum and hippocampus, demonstrated the highest number of associations (Fig. 4a and Supplementary Data Table 10). Among proteins, ITGA11 and GDF15 exhibited the most associations, linking to 23 and 17 distinct gray matter regions, respectively.

For white matter microstructures, 182 protein–white matter pairs were significant after FDR correction ($P_{FDR} = 2.24 \times 10^{-6}$ to 0.050), with 120 positive and 62 negative correlations. These significant associations involved predominantly measures of mean diffusivity (MD, such as inferior and superior longitudinal fasciculus) and isotropic volume fraction (ISOVF, such as middle cerebellar peduncle and posterior thalamic radiation; Fig. 4b and Supplementary Data Table 10). Of these,



https://doi.org/10.1038/s44220-025-00582-5

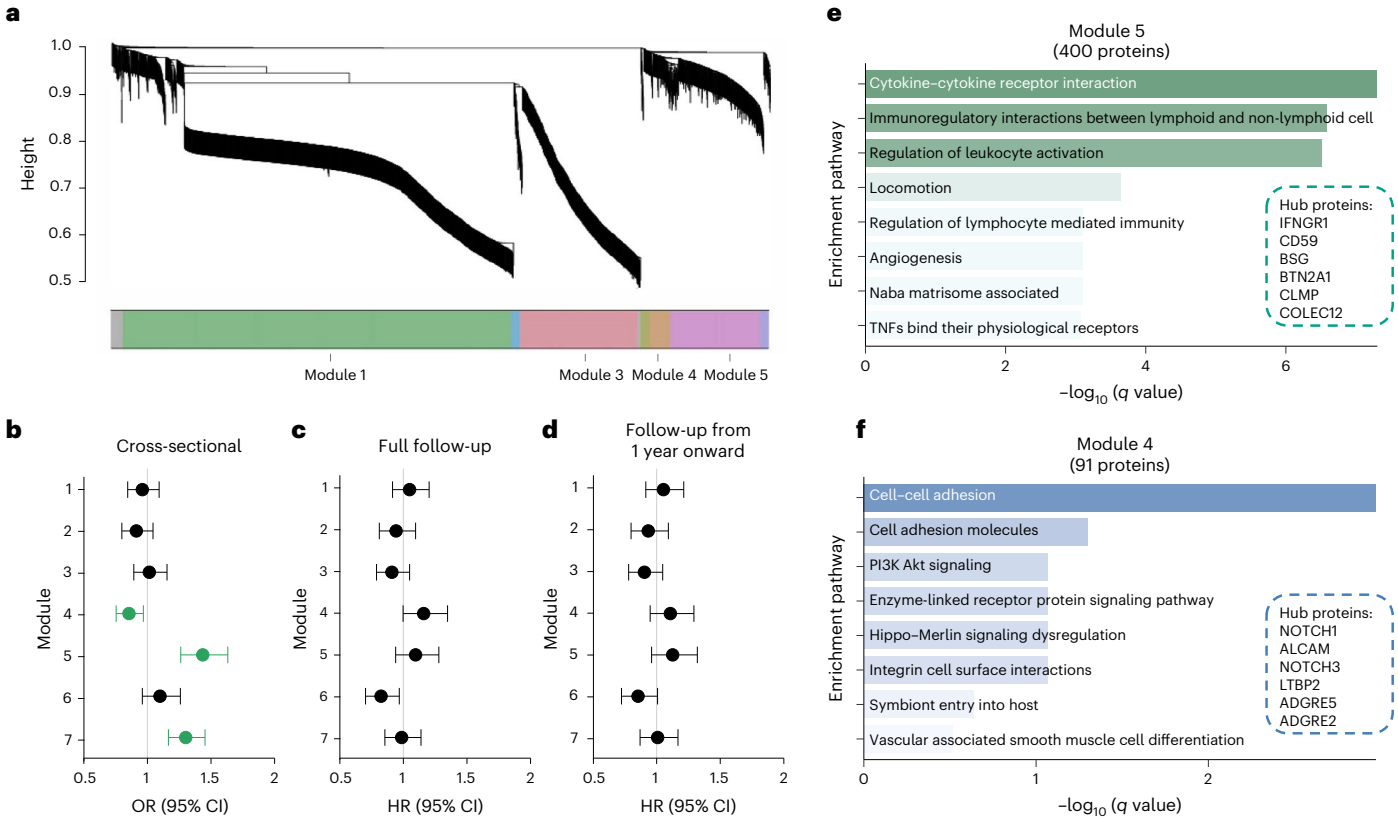


Fig. 3 | Co-regulated protein networks associated with SBs and their functional annotations. a, A hierarchical cluster tree of 2,920 plasma proteins, with the band indicating the division of proteins into seven distinct modules using Netboost clustering (53,026 participants). **b**, Associations between module expression and SBs in the cross-sectional datasets (N_{case} = 268, $N_{control}$ = 52,662). Data are presented as ORs with 95% confidence intervals (CIs). Protein networks significantly associated with SBs (FDR corrected) are highlighted in light green. **c**,**d**, Associations between module expression and SBs in the longitudinal datasets during the entire follow-up period (N_{case} = 202, $N_{control}$ = 52,303) (**c**) and starting from one year onward (N_{case} = 185, $N_{control}$ = 52,181) (**d**). Data are presented as HRs with 95% CIs. **e**,**f**, Pathway enrichment analysis of the gene-set encoding proteins in modules 5 (**e**) and 4 (**f**) based on six databases within the Metascape platform. The top eight enriched pathways are shown, and the *x* axis represents the $-\log_{10} q$ value for each term. The six hub proteins, which are most strongly correlated with overall network expression in each module, are also displayed.

growth/differentiation factor 8 (MSTN) showed the most associations, connected with 17 white matter microstructures.

Co-regulated protein network associations. We also examined the relationship between the three identified co-regulated protein networks and brain structure measurements. Protein module 5 was associated primarily with cortical regions, including the insula, medial orbitofrontal cortex, transverse temporal cortex and precuneus ($P_{uncorrected} < 0.021$; Fig. 4c). By contrast, protein module 7 was associated mainly with the hippocampus, amygdala, pars opercularis and middle temporal cortex ($P_{uncorrected} < 0.05$; Fig. 4e). Protein module 4 was associated with the middle temporal cortex, supramarginal gyrus, superior temporal cortex, precuneus, pallidum and thalamus ($P_{uncorrected} < 0.02$; Fig. 4d). However, none of the network-level associations survived FDR correction.

Causal effects between proteins, co-regulated protein networks and SBs

To investigate potential causal relationships between proteins, co-regulated protein networks and SBs, we conducted bidirectional two-sample MR analyses between 421 plasma proteins and SBs, as well as between the three protein networks and SBs. Forward MR results obtained using the inverse variance weighted (IVW) method indicated that, of the 424 proteins or protein networks examined for potential causal effects on SBs, 43 reached significance at $P_{IVW} < 0.05$, and one protein, GGH (gamma-glutamyl hydrolase), remained significant after Bonferroni correction (OR_{IVW} = 1.07, P_{BOF} = 0.035; Supplementary Data Table 11). This association did not show substantial horizontal pleiotropy (as indicated by $P_{Intercept}$ for MR–Egger intercept > 0.05) and single nucleotide polymorphisms (SNPs) heterogeneity ($P_{Q\text{-}test} > 0.05$). In the reverse MR analyses, we identified potential causal effects of SBs on 35 proteins at $P_{IVW} < 0.05$. However, none of these proteins survived after FDR correction.

Causal pathways linking behavior-related risk factors, proteins and SBs

On the basis of the preceding bidirectional two-sample MR analysis, which suggested GGH as a potential causal protein for SBs, we further explored potential pathways linking behavior-related risk factors, GGH and SBs. Thirteen behavior-related risk factors were selected: age at the first intercourse, smoking, BMI, neuroticism, loneliness, risk-taking, cumulative traumatic events and childhood adversity (and its five subtypes: physical neglect, emotional neglect, sexual abuse, physical abuse and emotional abuse).

First, univariable MR analyses were performed to assess the potential causal effects of behavior-related risk factors on SBs. After FDR correction, nine factors remained significantly associated with SBs ($P_{FDR} = 6.11 \times 10^{-29}$ to 0.049; Supplementary Data Table 11): smoking, age at the first intercourse, neuroticism, cumulative traumatic events, childhood adversity, risk-taking, BMI, emotional neglect and loneliness. We subsequently explored the potential causal relationship between these nine risk factors and GGH using univariable MR analysis. Three factors showed evidence consistent with a potential influence on GGH after FDR correction ($P_{FDR} = 2.43 \times 10^{-17}$ to 0.023; Supplementary

https://doi.org/10.1038/s44220-025-00582-5

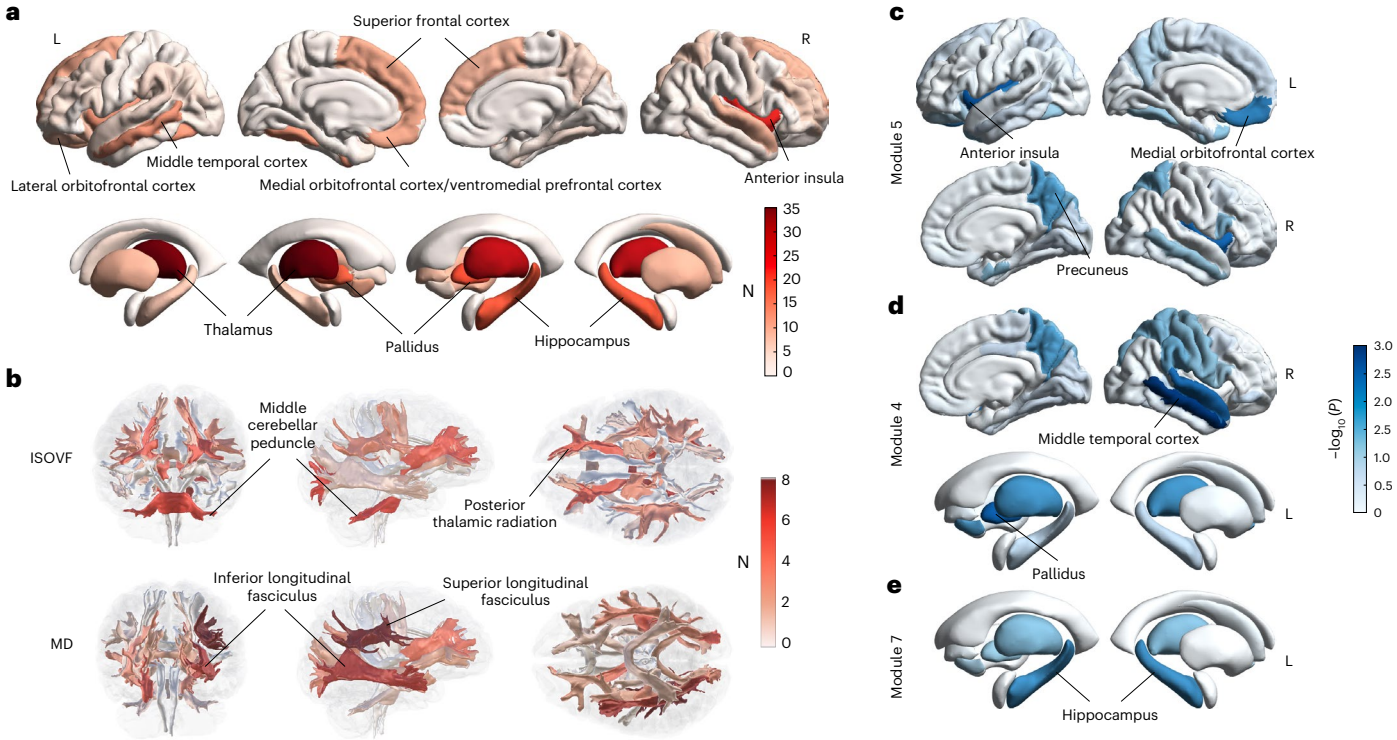


Fig. 4 | Associations of SBs-associated proteins and co-regulated protein networks with brain structures. a,**b**, The number of proteins significantly associated with gray matter volumes, including cortical and subcortical regions (**a**), and associated with white matter microstructures, such as ISOVF and MD (**b**). Darker shading represents a higher number of significant protein–brain structure associations after FDR correction. **c**–**e**, Significance plots showing associations between co-regulated protein network expression and gray matter volumes in module 5 (**c**), module 4 (**d**) and module 7 (**e**). The color bar indicates the $-\log_{10}$ of uncorrected *P* value resulting from a two-sided linear regression test between gray matter volumes and co-regulated network expression in each module.

Data Table 11): BMI (OR_{IVW} = 1.22, β = 0.20), age at first intercourse (OR_{IVW} = 0.88, β = -0.13) and neuroticism (OR_{IVW} = 1.16, β = 0.14).

Drawing on these findings, we proposed potential pathways interlinking behavior-related risk factors, GGH and SBs and applied multivariable MR analysis to assess potential mediation effects. This analysis supported one primary pathway: behavior-related risk factor → GGH → SBs, with GGH acting as a potential mediator. Specifically, GGH mediated 7.22% of the effect of BMI on SBs (Supplementary Data Table 11).

Proteomics-based machine-learning models for SBs

To evaluate the effectiveness of plasma proteins in identifying past SBs and predicting future occurrences, we developed machine-learning models using both cross-sectional and longitudinal protein measurements, with or without demographic information (age and sex). For the machine-learning model focused on identifying past SBs using cross-sectional protein data, a set of top 50 plasma proteins was exhibited and ranked by their importance to the classification task (Fig. 5a and Supplementary Data Table 12). Through a sequential forward selection strategy, the 23 most relevant proteins were selected as the final predictors for the model development. Among these, PLAUR emerged as the most significant protein, aligning with findings from logistic regression analyses. This model achieved an area under the receiver operating characteristic curve (AUC) of 0.756 in the cross-sectional classification (Fig. 5c). A notable improvement was observed when demographic information was included (AUC = 0.792; Fig. 5c). Importantly, we also developed a machine-learning model to predict future SBs using longitudinal protein data. The top 27 proteins were chosen as the final predictors for the machine-learning model development using a sequential forward selection strategy (Fig. 5b and Supplementary Data Table 12). This predictive model yielded an AUC of 0.548 (Fig. 5d). Including demographic variables further enhanced the model's performance, resulting in an AUC of 0.566 (Fig. 5d). Additional models incorporating BMI alongside proteins and demographic variables yielded little further improvement (cross-sectional: AUC = 0.791; longitudinal: AUC = 0.545; Supplementary Data Table 12).

Discussion

This large-scale study with 2,920 plasma proteins measured in more than 50,000 participants revealed plasma proteomic profiles associated with SBs in both cross-sectional and longitudinal analyses. We identified 421 unique plasma proteins associated with past SBs, 15 of which also showed significant associations with future SBs. These SBs-related proteins were enriched in pathways such as cytokine–cytokine receptor and TNF–receptor interactions, highlighting the role of the immune–inflammatory pathway in SBs. Co-regulated network analysis further identified three protein networks linked to SBs, involving inflammatory and cell–cell adhesion pathways. In addition, these SBs-related proteins and co-regulated protein networks were correlated with the volume of brain regions involved in emotion, including the medial and lateral orbitofrontal cortex, insula, superior frontal cortex and middle temporal cortex. Notably, our findings suggest one specific protein, GGH, as a potential causal factor for SBs, and GGH mediates the effect of behavior-related risk factors such as BMI on SBs. Finally, our machine-learning models demonstrated that plasma protein levels, combined with demographic data, could accurately identify past SBs.

Our analyses identified 421 plasma proteins associated with SBs in cross-sectional analyses, with 15 of these associations replicated in the longitudinal analysis. Among these, PLAUR, an inflammation-related protein, showed particularly strong associations with SBs, whereas IL-6, a key pro-inflammatory cytokine, emerged as a hub protein in PPI network analysis. Functional enrichment analysis further indicated that these SBs-related proteins were involved predominantly in inflammatory pathways, such as cytokine–cytokine receptor interactions and TNF–receptor interactions. Collectively, these results

https://doi.org/10.1038/s44220-025-00582-5

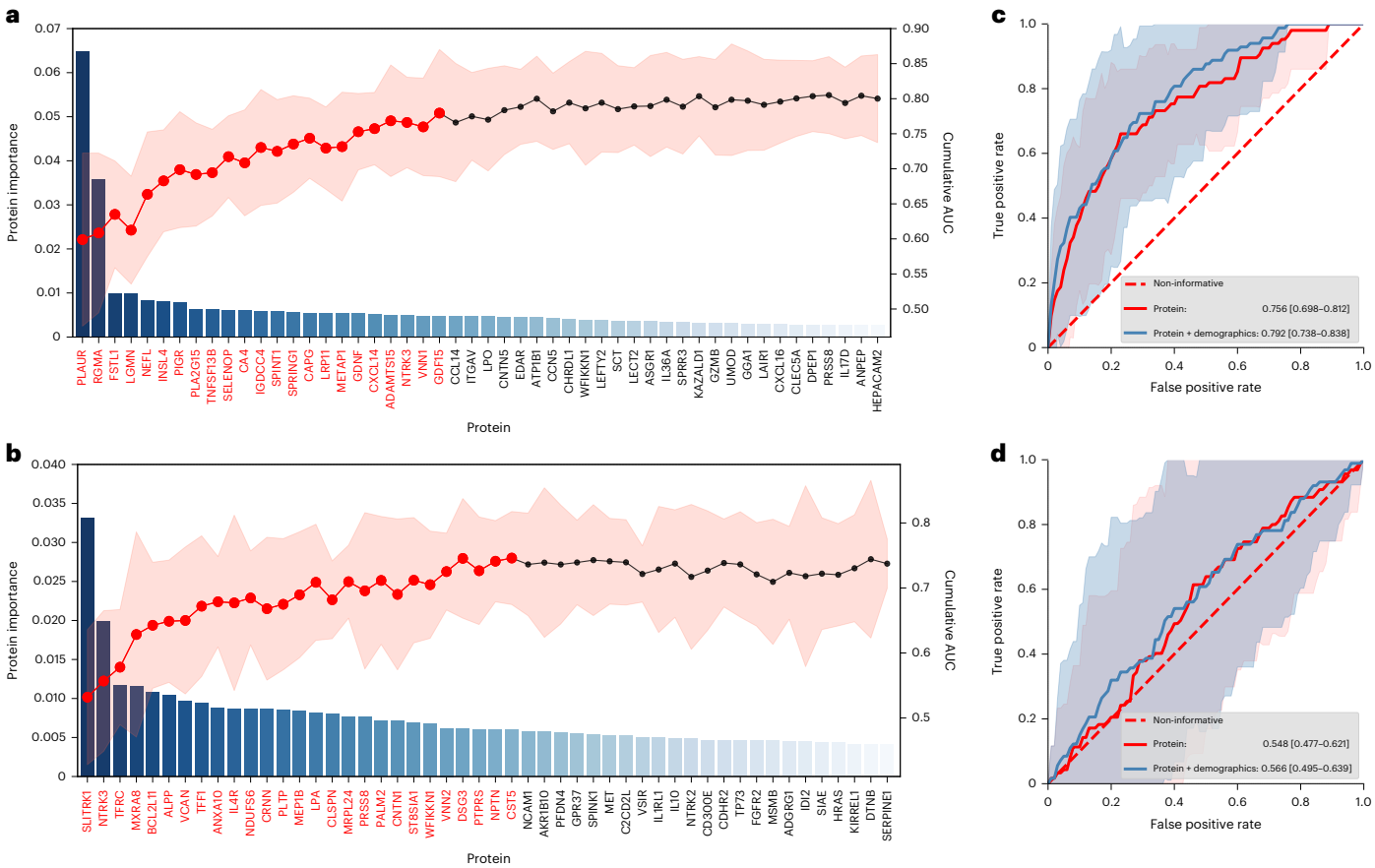


Fig. 5 | Predictor selection and performance of machine-learning models of SB. a,**b**, Sequential forward selection of plasma proteins in the derivational dataset (randomly partitioned 2/3 dataset) presents classification (**a**) and demonstrates prediction (**b**) tasks. The bar chart displays proteins ranked by their importance in the task for SBs. The line chart illustrates the cumulative AUCs (right axis) on the inclusion of proteins, one by each iteration, and shaded regions represent the corresponding 95% CIs derived from cross-validation within the derivational dataset. The top 23 proteins (**a**) and 27 proteins (**b**), highlighted in red, were selected as final predictors for model development. **c**,**d**, AUC plots display the performance in a holdout testing dataset (remaining 1/3 dataset) for SB classification (**c**) and prediction (**d**). Shaded regions represent 95% CIs derived from the holdout testing dataset.

suggest that inflammatory processes play a critical role in the pathophysiology of SBs. Our results align with previous studies demonstrating that an inflammatory profile is linked to increased SB risk. For example, abnormal levels of pro-inflammatory cytokines (such as IL-1β, IL-6 and TNF) in the plasma, cerebrospinal fluid and postmortem brain tissue have consistently been reported in individuals with SBs[18–22]. The co-regulated protein network analyses further support the involvement of inflammatory processes in SB risk. In particular, protein module 5 was enriched for multiple immune-related pathways, such as cytokine–cytokine receptor interaction and immunoregulatory interactions between lymphoid and non-lymphoid cells, and regulation of lymphocyte-mediated immunity. Consistent with these findings, our previous systems-level study also demonstrated that a biological cluster comprising elevated lymphocyte count, monocyte count, white blood cell count and C-reactive protein was significantly associated with increased SBs[5]. Taken together, our findings suggest that inflammatory factors may serve as biomarkers reflecting the severity of SB risk and offer potential targets for future therapeutic interventions.

Beyond inflammatory processes, our results highlighted additional biological processes implicated in SBs, notably those related to cell–cell adhesion, as revealed by enrichment of protein module 4. Cell adhesion molecules (CAMs) are essential for neural circuit development and maintenance by regulating synapse formation, stabilization and plasticity[23,24]. Dysregulation of CAM signaling may compromise synaptic integrity and plasticity, potentially disrupting the structural morphology and connectivity of neural circuits that underlie emotion regulation and impulse control[23,25]. Consistent with this notion, protein module 4 was correlated with structural measures of multiple brain regions, including the middle temporal cortex, supramarginal gyrus, superior temporal cortex, precuneus, pallidum and thalamus, which are involved in integrative processing of emotion regulation, social cognition, self-referential processing, reward and motivation. Alterations in these brain regions have also been reported in individuals with suicidality, as evidenced by neuroimaging and postmortem studies showing reduced synaptic density and aberrant connectivity[26–28]. Taken together, these findings suggest that disruptions in CAM-mediated processes could contribute to neural alterations underlying SBs.

Our findings indicated that various cortical brain regions, including the insula, superior frontal cortex, precuneus, middle temporal cortex, and medial and lateral orbitofrontal cortex, were highly correlated with SBs-associated proteins. Reduced gray matter volume in these brain areas was also significantly associated with increased polygenic risk scores for suicide attempts in our previous work[5], particularly in the right insula, ventromedial prefrontal cortex, and medial and lateral orbitofrontal cortex. These regions substantially overlap with those identified in a recent postmortem study of individuals who died by suicide, which reported significant gray matter volume reductions in the ventral anterior cingulate, orbitofrontal cortex, insula and precuneus[28]. Collectively, these findings highlight the potential critical involvement of the insula, precuneus and orbitofrontal cortex in the neurobiology of SBs. These regions may form a functional network

integrating interoception, self-reflection, emotional regulation and value-based decision-making[29,30]. Dysfunction within this network could contribute to maladaptive cognitive and behavioral patterns, thereby increasing the risk of SBs.

In our MR analyses, elevated levels of GGH were implicated as a potential causal factor for SBs. GGH is a key enzyme in folate metabolism, catalyzing the hydrolysis of polyglutamylated folates into monoglutamyl forms that are transportable and utilizable in one-carbon metabolism[31]. Clinically, reduced serum folate levels have been observed in individuals with SBs[32,33], and folic acid supplementation has been associated with reduced SBs[34]. One possible explanation for our finding is that upregulation of GGH may occur in response to reduced folate availability, reflecting a compensatory attempt to increase folate turnover. However, excessive GGH activity may paradoxically deplete intracellular folate as the conversion to monoglutamyl forms facilitates folate efflux and destabilizes cellular folate storage[35]. In addition, we further observed that GGH mediates the effects of behavior-related risk factors, such as BMI, on SBs. Elevated BMI is linked to chronic low-grade inflammation and oxidative stress[36], both of which can impair one-carbon metabolism and reduce folate availability. In this context, GGH may act as a mechanistic intermediary linking metabolic and inflammatory disturbances to folate dysregulation and increased SBs risk.

PLAUR (urokinase plasminogen activator surface receptor, uPAR), a stable inflammatory biomarker and central component of the plasminogen activation system[37], emerged as the protein most strongly associated with SBs in our study. This is consistent with previous findings of elevated plasma soluble uPAR (suPAR) levels in suicide attempters, where suPAR outperformed C-reactive protein in distinguishing attempters from controls[38]. These results suggest that PLAUR/suPAR may have clinical utility in identifying individuals at elevated SBs risk. Beyond serving as a robust inflammatory indicator, PLAUR may also represent a mechanistic link between immune activation and vascular biology in shaping SBs vulnerability. Immunologically, uPAR regulates immune cell adhesion and migration[39], with elevated suPAR reflecting chronic systemic immune activation. Vascularly, uPAR participates in fibrinolysis and angiogenesis, with potential implications for endothelial function and blood–brain barrier integrity[37]. Together, these processes suggest a putative 'immune–vascular–brain' pathway, through which peripheral inflammatory signals may impact central nervous system circuits relevant to SBs[40]. Thus, rather than serving solely as a passive biomarker, PLAUR/suPAR could be considered a candidate hub mediating the interplay between immune and vascular pathology in SBs risk, although this hypothesis warrants further mechanistic investigation.

Our study also has several conceptual implications that warrant further discussion. First, while our machine-learning model achieved moderate performance in identifying past SBs, its performance in predicting future SBs was poor. This highlights the limited capacity of proteomic and demographic data alone for prospective risk prediction and suggests that more integrative, multimodal approaches (such as combining proteomics with genomics, neuroimaging and psychosocial markers) may be necessary to enhance predictive performance. Second, our findings reinforce the role of inflammatory factors, including IL-6, cytokine–cytokine receptor interaction and TNF–receptor signaling, in SBs. These factors are broadly implicated across psychiatric and medical conditions. Their non-specificity, however, complicates causal interpretation in the context of SBs. Elevated inflammatory signaling may influence SB vulnerability through diverse mechanisms, including altered neurotransmitter metabolism, hypothalamic–pituitary–adrenal axis dysregulation and impaired neuroplasticity[4]. Therefore, inflammatory factors, including IL-6, cytokine–cytokine receptor interaction and TNF–receptor signaling, may be best conceptualized as components of a wider landscape of systemic dysregulation, rather than SBs-specific pathways. Finally, enrichment in the pancreatic cancer subtypes pathway does not imply a direct association between SBs and pancreatic cancer. Instead, it may reflect shared molecular processes, such as immune regulation and metabolism, which are relevant to both cancer biology and vulnerability to SBs.

While our study has many strengths, such as benefitting from a large population-based sample and utilization of various data-driven approaches, several limitations should be considered when interpreting these results. First, the measurement data for the approximately 3,000 plasma proteins were obtained using the Olink platform. This platform does not provide comprehensive coverage of the plasma proteome and may introduce bias toward certain proteins. Future research expanding proteomic coverage to include cell-specific and organ-specific splice isoforms and post-translational modifications could offer deeper biological insights into SBs. Second, given that plasma protein levels are subject to biological variations, repeated measurements over time would be necessary to improve the reliability of the findings. Third, it is essential to acknowledge the inherent limitations of the UK Biobank dataset, such as the 'healthy volunteer' bias[41] and the fact that participants are primarily middle- to late-aged individuals of European ancestry. These factors may limit the generalizability of our findings to more diverse populations. Fourth, although we applied stringent multiple testing corrections and sensitivity analyses, the combination of modest case numbers with high-dimensional proteomic data may still constrain statistical power and stability of results. Therefore, our findings should be regarded as exploratory, and future studies with larger and more diverse cohorts will be essential to validate and extend these results. Fifth, inclusion of baseline medication use as additional covariates attenuated some of the observed associations between plasma proteins and SBs, particularly in the longitudinal analyses. While this adjustment probably reduced confounding, it may also have introduced overfitting given the relatively small number of cases. In addition, medication use during follow-up was not comprehensively captured, limiting our ability to assess dynamic effects. Future studies incorporating longitudinal tracking of medication use and its impact on both plasma protein levels and SBs would improve causal inference and clarify temporal relationships. Finally, although we controlled for socioeconomic status, BMI, smoking and multiple categories of chronic medical and psychiatric disorders, other lifestyle and environmental factors, such as diet, physical activity and sleep, were not systematically accounted for. These factors may influence systemic inflammatory profiles and SB risk, and their integration will be critical in future work.

Methods

Study population

The UK Biobank is a population-based cohort initiated between 2006 and 2010 in the United Kingdom (http://www.ukbiobank.ac.uk). Over 500,000 participants aged 37–73 years were recruited at baseline. Comprehensive health information was collected for each participant, including lifestyle factors, health conditions, physical measures, biological samples, imaging data and genotyping. Long-term follow-up has been successfully established and is currently under way. The UK Biobank has received research tissue bank approval from the North West Multi-centre Research Ethics Committee, and informed consent was obtained from all participants. This study used the UK Biobank Resource under application number 19542.

Proteomic profiling

Blood samples were collected from over 50,000 participants during their baseline recruitment visit. Proteomic profiling was performed on blood plasma samples using the Olink Explore 3072 platform by Olink Analysis Service in Sweden. This platform captured 2,923 unique protein analytes, measured across eight panels, including cardiometabolic, cardiometabolic II, inflammation, inflammation II, neurology, neurology II, oncology and oncology II. Protein measurements were provided as normalized protein expression values, following stringent

quality control protocols as recommended by the manufacturer[42]. For further details on UK Biobank proteomics participant selection, sample handling and quality control of protein data, refer to ref. 43. Three proteins were excluded from the analysis due to more than 50% missing values. After exclusions, a total of 2,920 unique proteins were included in our analysis.

SBs definitions

Cases of SBs were identified by integrating information from hospital inpatient records and death register data, with a diagnosis code of X60–X84 according to ICD-10 (https://icd.who.int/browse10/2019/en#/X60-X84), encompassing suicide attempt and death by suicide[2]. Cases were classified into two groups depending on whether the behavior occurred earlier or later than the time of the baseline assessment (pre or post blood sample collection), with the former used for cross-sectional analyses and the latter for prospective longitudinal analyses.

In the cross-sectional study, the control group comprised individuals who had not attempted suicide before the baseline assessment. In the longitudinal study, individuals who had not attempted suicide before the baseline assessment were followed until the earliest recorded date of SBs diagnosis, or the date of death, or censoring (that is, 1 September 2023), whichever occurred first. Cases and controls were identified according to the occurrence of SBs during the follow-up period. To avoid misclassification, individuals with ICD-10 diagnosis codes for X40–X49 (accidental poisoning by and exposure to noxious substances), Y10–Y34 (event of undetermined intent), Y87.0 (sequelae of intentional self-harm) and Y87.2 (sequelae of events of undetermined intent) were excluded from the control group due to external illness or death potentially related to SBs. In total, 268 cases (all suicide attempts) and 52,662 controls with plasma proteomic data were included in the cross-sectional analysis, while 202 cases (83% suicide attempts and 17% deaths by suicide) and 52,303 controls were included in the longitudinal analysis.

Phenotypes

Structural MRI phenotypes. The UK Biobank collected multimodal brain imaging from approximately 40,000 participants using a standard Siemens Skyra 3 T scanner (VD13A SP4) equipped with a standard Siemens 32-channel head coil. All quality checks and data pre-processing procedures were conducted by the UK Biobank. Detailed information regarding acquisition protocols, image processing pipelines, image data files and imaging-derived phenotypes of brain structure is available on the UK Biobank website (https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf). Structure MRI encompassed primarily two categories: gray matter volumes derived from T1-weighted imaging data and white matter microstructures obtained from diffusion-weighted imaging data. A total of 68 cortical gray matter volumes defined by the Desikan–Killiany atlas (category 192) and 14 subcortical volumes defined by the ASEG atlas (category 190) were included in our analyses. Measures of white matter microstructure, including fractional anisotropy, MD, intracellular volume fraction, orientation dispersion and ISOVF, which were available for 27 major tracts mapped across the brain, were used (category 135).

Behavior-related phenotypes. Thirteen behavior-related risk factors were included to quantify the pathway interlinking behavior-related phenotypes, proteins, protein networks and SBs. These phenotypes were selected according to their strong associations with SBs risk[5,44], including age at the first intercourse (data field 2139), smoking (data field 20160), BMI (data field 21001), neuroticism (data field 20127), loneliness (data field 2020), risk-taking (data field 2040), cumulative traumatic events (category 145) and childhood adversity (data field 20487–20491, five items from the UKB Mental Health Questionnaire, including five specific subtypes: physical neglect, emotional neglect, sexual abuse, physical abuse and emotional abuse).

Statistical analysis

Logistic regression analysis. Logistic regression analyses were conducted using the R-implemented glm function on 2,920 unique plasma proteins to identify those associated with SBs in cross-sectional data (N_{case} = 268, $N_{control}$ = 52,662). These analyses were adjusted for potential confounders, including sex, age, assessment center, ethnicity (categorized as white and others), socioeconomic status (as measured by the Townsend deprivation index), BMI, protein batch, the time lag between blood sample collection and proteomic testing, smoking status (categorized as ever smoker and never smoker) and baseline diagnoses of chronic medical or psychiatric disorders. The baseline disorders comprised 37 conditions grouped into six categories: psychiatric, nervous system, cardiovascular, endocrine system, immune system and cancer (see Supplementary Table 2 for details). Significant associations were determined with an FDR-corrected two-sided P value < 0.05. Similarly, logistic regression analyses were performed for seven protein network expressions, adjusting for sex, age, assessment center, ethnicity, socioeconomic status, BMI, smoking status and six categories of baseline disease diagnoses. An FDR-corrected two-sided P value < 0.05 was also considered significant. Note that, for all analyses, participants with missing proteomic or covariate data were excluded from the relevant analyses, and no imputation was applied.

To examine the potential influence of medication use, additional logistic regression analyses were performed for both plasma proteins and co-regulated protein networks, with five categories of baseline medication use added as confounders. These categories, psychotropic drugs, antiepileptics, opioids, cardiometabolic drugs and steroids/immune modulators, were classified according to the ATC system at the third level[45] (see Supplementary Table 3 for details). Results from these additional analyses are summarized in the Supplementary Information (page 4) and reported in Supplementary Data Table 13.

Cox proportional hazards model. Cox proportional hazards models were conducted using the survival package in R to quantify longitudinal associations between the 2,920 plasma proteins and incident SBs (N_{case} = 202, $N_{control}$ = 52,303). Time 0 was defined as the date of blood draw at baseline, which served as the reference point for prospective analyses. A fully adjusted model was utilized, incorporating the same covariates as in the logistic regression analyses, including sex, age, assessment center, ethnicity, socioeconomic status, BMI, protein batch, the time lag between blood sample collection and proteomic testing, smoking status and six categories of baseline disease diagnoses. Statistical significance was defined as an FDR-corrected two-sided P value < 0.05. In sensitivity analyses, we excluded individuals who had attempted suicide within the first year of the baseline examination and performed the same association analyses. In addition, similar Cox models were employed to examine the longitudinal associations between the seven network expressions and incident SB, adjusting for sex, age, assessment center, ethnicity, socioeconomic status, BMI, smoking status and six categories of baseline disease diagnoses. Finally, additional Cox models incorporating five categories of medication use as confounders were conducted for both plasma proteins and co-regulated protein networks to assess the potential influence of medication use, and the results are reported in the Supplementary Information (page 4).

Co-regulated protein networks. The co-regulated protein networks (modules) were constructed from the full set of 2,920 plasma proteins across 53,026 participants using the Bioconductor R package Netboost[46]. Netboost serves as a dimensionality reduction tool that enhances the well-established methodology of weighted gene co-expression network analysis[16,17]. In accordance with previous network-based proteomic analyses[47,48], we set a minimum module size of 20 and employed a Spearman filter method with a soft power of β = 2,

determined through an analysis of network topology (considering scale independence and mean connectivity). Further computational details are provided in ref. 48 and in the Supplementary Information (page 5).

Correlation analysis. The R-implemented lm function was used to examine lineal associations between plasma protein levels and brain structural measurements. Each association test included covariates for sex, age, assessment center, ethnicity, socioeconomic status, BMI, protein batch, time lag, smoking status and six categories of baseline disease diagnoses, with total intracranial volume included as an additional covariate for analyses of gray matter volumes. Statistical significance was assessed using FDR correction across protein–gray matter volume pairs (421 proteins × 82 gray matter volumes = 34,522 associations) and protein–white matter microstructure pairs (421 proteins × 135 white matter microstructures = 56,835 associations). Similar analyses were conducted to assess the associations between SBs-related co-regulated protein network expressions and brain structural measurements, except for adjusting for protein batch and time lag.

Two-sample and two-step MR. The TwosampleMR package in R was used to perform bidirectional two-sample MR analyses on SBs-associated proteins, co-regulated protein networks and SBs. Genome-wide association study (GWAS) summary statistics for SBs were obtained from a European ancestry SB GWAS meta-analysis in a previous study[49], but excluding UK Biobank data to prevent sample overlap (33,353 cases and 444,626 controls; see details in page 6 of the Supplementary Information).

GWAS summary data for the 421 SBs-associated proteins were provided by the UK Biobank Pharma Proteomics Project consortium, as reported previously ($N > 50{,}000$)[43]. In addition, GWAS summary data for SBs-associated co-regulated protein networks (modules 4, 5 and 7) were generated from the UK Biobank samples used to compute co-regulated protein networks ($N = 53{,}026$). These GWAS analyses employed linear regression in PLINK 2.0[50]. Further criteria on data exclusion, quality control, ancestry checks and relatedness adjustments are available in the Supplementary Information (page 7).

For causal testing of proteins and co-regulated protein networks on SBs, we selected genetic instruments at a P-value threshold of 5×10^{-8}. Each exposure's significant SNPs were clumped using a distance of 1,000 kb and a maximum linkage disequilibrium of $r^2 = 0.01$, according to European ancestry reference data from the 1000 Genomes Project[51]. SNP effect data on both exposure and outcome were harmonized by aligning effect alleles before conducting the MR analysis. Note that 17 proteins were excluded from analysis due to having fewer than three SNPs. For the causal impact of SBs on proteins and co-regulated protein networks, we chose genetic instruments from the SBs GWAS at a P-value threshold of 1×10^{-6}. These SNPs underwent clumping with the same parameters, resulting in 28 independent genetic instruments. These SNPs were then identified within the GWAS summary statistics for each outcome, and those that were not present in both GWAS datasets were removed before harmonization. The threshold of 1×10^{-6} was set to maintain a sufficient number of genome-wide significant hits (three or more) for analysis in line with MR method requirements, such as MR–Egger regression, as done in previous studies.

We utilized primarily the IVW method for MR analysis. Sensitivity analyses evaluated potential horizontal pleiotropy using the MR–Egger intercept and assessed genetic instrument heterogeneity via Cochran's Q test. We also utilized a multiplicative random-effects model to confirm any factors showing potential heterogeneity in genetic instruments. Bonferroni correction was applied across all examined proteins and co-regulated protein networks.

A two-step MR analysis[52] was performed to explore potential causal pathways linking selected behavior-related risk factors, plasma proteins (GGH) and SBs. Detailed methods are available in the Supplementary Information (page 8).

Machine-learning models. We employed LightGBM[53] to build a model aimed at predicting whether a participant falls into class 0 (individuals without SBs) or class 1 (individuals with SBs). Briefly, all 2,920 plasma proteins, along with or without demographic information (including age and sex), were considered as predictors. The model was developed using the derived dataset (randomly split 2/3 dataset) and evaluated in a retained replication (remaining 1/3 dataset). Models were trained under a tenfold cross-validation strategy to find optimized proteins and hyperparameters. Evaluation was performed using bootstrap with 1,000 iterations. Model performance was evaluated using discrimination measured by the AUC, which ranges from 0.5 (indicating a non-informative model) to 1 (indicating a perfect discriminating model). Using this approach, we built four machine-learning models that included either protein measurements alone or protein measurements combined with demographic information (age and sex) for cross-sectional ($N_{case} = 268$, $N_{control} = 52{,}662$) and longitudinal ($N_{case} = 202$, $N_{control} = 52{,}303$) datasets, respectively. In addition, we developed machine-learning models that integrated protein measurements with demographic variables (age and sex) and BMI in both cross-sectional and longitudinal datasets.

Enrichment analysis, protein–protein interaction network analysis and druggability assessment. Functional annotations of SBs-related proteins were performed on the basis of the plasma proteins significantly associated with SBs in the cross-sectional analysis to characterize their biological relevance. Detailed methods are provided in the Supplementary Information (pages 9–10).

Reporting summary

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

Data availability

The data used in this study are available from UK Biobank (https://www.ukbiobank.ac.uk) with restrictions applied. Data were used under license and are thus not publicly available. Researchers can apply for access to the UK Biobank data via the Access Management System (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access). Publicly available summary-level data were obtained from the following sources: the GWAS summary statistics for suicide attempts were obtained for this specific study from the PGC SUI Data Access Committee under a data use agreement that does not permit public redistribution (ref. 49). Researchers interested in these specific data can apply directly to the committee for access at https://pgc.unc.edu/for-researchers/data-access-committee/data-access-portal/. GWAS summary data for plasma proteins can be downloaded from the UK Biobank Pharma Proteomics Project (http://ukb-ppp.gwas.eu) (ref. 43). European ancestry reference data from the 1000 Genomes Project can be found at https://github.com/getian107/PRScsx?tab=readme-ov-file (ref. 51).

Code availability

The analyses in this study utilized a range of software and packages. Statistical modeling, including logistic regression (base R), linear regression (base R), Cox proportional hazards models (survival v3.6.4), co-regulated protein network analysis (Netboost v2.18.1) and Mendelian randomization (TwoSampleMR package v0.5.6), was conducted in R (v4.2.3). GWAS was carried out with PLINK 2.0. Machine-learning models were developed using the LightGBM library (v3.3.2) in Python 3.9. Furthermore, several online platforms were employed for functional enrichment analyses: the Metascape platform for pathway enrichment, the FUMA GENE2FUNC module for tissue-specific enrichment,

the TRRUST database for transcriptional regulator enrichment and the STRING database for protein–protein interaction network analysis. Druggability assessment was conducted using the GREP platform. All primary code used in this study has been made publicly accessible via the GitHub repository (https://github.com/beimagic/Suicide_PlasmaProteins).

References

1. *Preventing Suicide: A Global Imperative* (WHO, 2014).
2. Monson, E. T. et al. Defining and assessing international classification of disease suicidality phenotypes for genetic studies. *Psychiatry Res.* **353**, 116760 (2025).
3. Hawton, K. & Pirkis, J. Preventing suicide: a call to action. *Lancet Public Health* **9**, e825–e830 (2024).
4. Turecki, G. et al. Suicide and suicide risk. *Nat. Rev. Dis. Primers* **5**, 74 (2019).
5. Zhang, B. et al. Identifying behaviour-related and physiological risk factors for suicide attempts in the UK Biobank. *Nat. Hum. Behav.* **8**, 1784–1797 (2024).
6. Pigoni, A. et al. Machine learning and the prediction of suicide in psychiatric populations: a systematic review. *Transl. Psychiatry* **14**, 140 (2024).
7. Dimitrov, D. S. Therapeutic Proteins. In *Therapeutic Proteins: Methods and Protocols* (eds Voynov, V. & Caravella, J. A.) 1–26, (Humana Press, 2012).
8. Williams, S. A. et al. Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* **25**, 1851–1857 (2019).
9. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
10. Carrasco-Zanini, J. et al. Proteomic signatures improve risk prediction for common and rare diseases. *Nat. Med.* **30**, 2489–2498 (2024).
11. Guo, Y. et al. Plasma proteomic profiles predict future dementia in healthy adults. *Nat. Aging* **4**, 247–260 (2024).
12. Sandberg, J. V. et al. Proteins associated with future suicide attempts in bipolar disorder: a large-scale biomarker discovery study. *Mol. Psychiatry* **27**, 3857–3863 (2022).
13. Lengvenyte, A. et al. Associations of potential plasma biomarkers with suicide attempt history, current suicidal ideation and subsequent suicidal events in patients with depression: a discovery study. *Brain Behav. Immun.* **114**, 242–254 (2023).
14. Kappelmann, N. et al. Dissecting the association between inflammation, metabolic dysregulation, and specific depressive symptoms: a genetic correlation and 2-sample Mendelian randomization study. *JAMA Psychiatry* **78**, 161–170 (2021).
15. Zhao, H. et al. Identifying novel proteins for suicide attempt by integrating proteomes from brain and blood with genome-wide association data. *Neuropsychopharmacology* **48**, 1255–1265 (2024).
16. Schlosser, P. et al. Netboost: boosting-supported network analysis improves high-dimensional omics prediction in acute myeloid leukemia and Huntington's disease. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **18**, 2635–2648 (2020).
17. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* **9**, 559 (2008).
18. Ducasse, D., Olie, E., Guillaume, S., Artero, S. & Courtet, P. A meta-analysis of cytokines in suicidal behavior. *Brain Behav. Immun.* **46**, 203–211 (2015).
19. Black, C. & Miller, B. J. Meta-analysis of cytokines and chemokines in suicidality: distinguishing suicidal versus nonsuicidal patients. *Biol. Psychiatry* **78**, 28–37 (2015).
20. Ganança, L. et al. The role of cytokines in the pathophysiology of suicidal behavior. *Psychoneuroendocrinology* **63**, 296–310 (2016).
21. Erhardt, S. et al. Connecting inflammation with glutamate agonism in suicidality. *Neuropsychopharmacology* **38**, 743–752 (2013).
22. Miná, V. et al. The influence of inflammatory cytokines in physiopathology of suicidal behavior. *J. Affect. DIsord.* **172**, 219–230 (2015).
23. Fields, R. D. & Itoh, K. Neural cell adhesion molecules in activity-dependent development and synaptic plasticity. *Trends Neurosci.* **19**, 473–480 (1996).
24. Gumbiner, B. M. Cell adhesion: the molecular basis of tissue architecture and morphogenesis. *Cell* **84**, 345–357 (1996).
25. Vawter, M. P. Dysregulation of the neural cell adhesion molecule and neuropsychiatric disorders. *Eur. J. Pharmacol.* **405**, 385–395 (2000).
26. Schmaal, L. et al. Imaging suicidal thoughts and behaviors: a comprehensive review of 2 decades of neuroimaging studies. *Mol. Psychiatry* **25**, 408–427 (2020).
27. Zhao, J., Lucassen, P. J. & Swaab, D. F. Suicide is a confounder in postmortem studies on depression. *Biol. Psychiatry* **86**, e37–e40 (2019).
28. Deininger-Czermak, E. et al. Magnetic resonance imaging of regional gray matter volume in persons who died by suicide. *Mol. Psychiatry* **30**, 1029–1033 (2024).
29. Zhang, B. et al. Roles of the medial and lateral orbitofrontal cortex in major depression and its treatment. *Mol. Psychiatry* **29**, 914–928 (2024).
30. Rolls, E. T. Emotion, motivation, decision-making, the orbitofrontal cortex, anterior cingulate cortex, and the amygdala. *Brain Struct. Funct.* **228**, 1201–1257 (2023).
31. Galivan, J. et al. Glutamyl hydrolase: pharmacological role and enzymatic characterization. *Pharmacol. Ther.* **85**, 207–215 (2000).
32. Liwinski, T. & Lang, U. E. Folate and its significance in depressive disorders and suicidality: a comprehensive narrative review. *Nutrients* **15**, 3859 (2023).
33. Kim, J.-M. et al. Prediction of suicidality according to serum folate levels in depressive patients receiving stepwise pharmacotherapy. *Front. Psychiatry* **12**, 747228 (2021).
34. Gibbons, R. D., Hur, K., Lavigne, J. E. & Mann, J. J. Association between folic acid prescription fills and suicide attempts and intentional self-harm among privately insured US adults. *JAMA Psychiatry* **79**, 1118–1123 (2022).
35. Kim, S. et al. γ-Glutamyl hydrolase modulation and folate influence chemosensitivity of cancer cells to 5-fluorouracil and methotrexate. *Br. J. Cancer* **109**, 2175–2188 (2013).
36. Lopresti, A. L. & Drummond, P. D. Obesity and psychiatric disorders: commonalities in dysregulated biological pathways and their implications for treatment. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **45**, 92–99 (2013).
37. Blasi, F. & Carmeliet, P. uPAR: a versatile signalling orchestrator. *Nat. Rev. Mol. Cell Biol.* **3**, 932–943 (2002).
38. Ventorp, F., Gustafsson, A., Träskman-Bendz, L., Westrin, Å & Ljunggren, L. Increased soluble urokinase-type plasminogen activator receptor (suPAR) levels in plasma of suicide attempters. *PLoS ONE* **10**, e0140052 (2015).
39. Smith, H. W. & Marshall, C. J. Regulation of cell signalling by uPAR. *Nat. Rev. Mol. Cell Biol.* **11**, 23–36 (2010).
40. Murphy, J. et al. Associations between soluble urokinase plasminogen activator receptor (suPAR) concentration and psychiatric disorders—a systematic review and meta-analysis. *Brain Behav. Immun.* **120**, 327–338 (2024).
41. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
42. Eldjarn, G. H. et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023).
43. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).

44. Mullins, N. et al. Dissecting the shared genetic architecture of suicide attempt, psychiatric disorders, and known risk factors. *Biol. Psychiatry* **91**, 313–327 (2022).
45. Wu, Y. et al. Genome-wide association study of medication-use and associated disease in the UK Biobank. *Nat. Commun.* **10**, 1891 (2019).
46. Schlosser, P., Knaus, J. & Loewenstein, Y. Netboost v2.18.1 (Bioconductor, 2021); https://doi.org/10.18129/B9.bioc.netboost
47. Emilsson, V. et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769–773 (2018).
48. Walker, K. A. et al. Proteomics analysis of plasma from middle-aged adults identifies protein markers of dementia risk in later life. *Sci. Transl. Med.* **15**, eadf5681 (2023).
49. Docherty, A. R. et al. GWAS meta-analysis of suicide attempt: identification of 12 genome-wide significant loci and implication of genetic risks for specific health factors. *Am. J. Psychiatry* **180**, 723–738 (2023).
50. Chang, C. PLINK 2.0 alpha (cog-genomics, 2026); https://www.cog-genomics.org/plink/2.0/
51. Consortium, G. P. A global reference for human genetic variation. *Nature* **526**, 68 (2015).
52. Vabistsevits, M. et al. Deciphering how early life adiposity influences breast cancer risk using Mendelian randomization. *Commun. Biol.* **5**, 337 (2022).
53. Ke, G. et al. Lightgbm: a highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **30**, 3149–3157 (2017).

Acknowledgements

This study used the UK Biobank Resource under application number 19542. We thank all participants and researchers from the UK Biobank. This work was partly supported by the grant from the Natural Science Foundation of Shanghai (number 25ZR1404005 to B.Z.) and the National Key R&D Program of China (number 2025YFC2511300 to B.Z.). This study was partly supported by grants from the National Science and Technology Major Project of China (number 2025ZD0546300 to W.C.), the National Key R&D Program of China (number 2023YFC3605400 to W.C.), the National Natural Science Foundation of China (number 82472055 and number 62433008 to W.C.), the Shanghai Pilot Program for Basic Research—Fudan University 21TQ1400100 (number 25TQ010, to W.C.), and Shanghai Science and Technology Commission Program (number 23JS1410100 to W.C.). This work was partly supported by the grant from the 111 Project (B18015 to J.F.), Zhangjiang Lab and Shanghai Center for Brain Science and Brain-Inspired Technology.

Author contributions

B.Z., W.C. and J.F. conceived and designed the experiment. B.Z. did the analyses with support from J.Y., P.R., Y.L. and W.Z. B.Z. drafted the manuscript with contributions from E.T.R., B.J.S. and F.L. and comments from all other authors. B.Z., J.Y. and W.C. directly accessed and verified the data reported in the article. B.Z., W.C. and J.F. were responsible for the decision to submit the manuscript to the journal. All authors read and approved the final manuscript. All authors confirm that they had full access to all the data in the study and accept responsibility to submit for publication.

Competing interests

All authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44220-025-00582-5.

Correspondence and requests for materials should be addressed to Jianfeng Feng or Wei Cheng.

Peer review information *Nature Mental Health* thanks Livia A. Carvalho, Aiste Lengvenyte and the other, anonymous, reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

[1]Brain Health Institute, National Center for Mental Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine and School of Psychology, Shanghai, China. [2]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. [3]State Key Laboratory of Brain Function and Disorders and MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. [4]Department of Computer Science, University of Warwick, Coventry, UK. [5]Department of Psychiatry, University of Cambridge, Cambridge, UK. [6]Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK. [7]Developmental and Behavioural Pediatric Department and Child Primary Care Department, Ministry of Education–Shanghai Key Laboratory for Children's Environmental Health, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China. [8]School of Data Science, Fudan University, Shanghai, China. [9]Fudan ISTBI-ZJNU Algorithm Centre for Brain-Inspired Intelligence, Zhejiang Normal University, Jinhua, China. [10]Department of Neurology and National Center for Neurological Disorders, Huashan Hospital, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Shanghai Medical College, Fudan University, Shanghai, China. [11]Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Fudan University, Ministry of Education, Shanghai, China. [12]These authors contributed equally: Bei Zhang, Jia You. ✉e-mail: jianfeng64@gmail.com; wcheng@fudan.edu.cn

nature portfolio

Corresponding author(s): Jianfeng Feng
Wei Cheng

Last updated by author(s): Nov 14, 2025

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: The data for this study were sourced exclusively from the UK Biobank. Consequently, no software for data collection were required.

Data analysis	R Environment (v4.2.3): Base glm function for logistic regression analysis. Base lm function for linear regression analysis. survival package (v3.6.4) for Cox proportional hazards regression models. Netboost package (v1.0.0) for co-regulated protein network analysis. TwoSampleMR package (v0.5.6) for univariable and multivariable Mendelian Randomisation analyses. Python Environment (v3.9): LightGBM library (v3.3.2) for the development of machine learning models. Standalone Software & Web Platforms: PLINK 2.0 for GWAS analysis. Metascape platform for pathway enrichment analysis. FUMA GENE2FUNC module for tissue-specific enrichment analysis. TRRUST database for transcriptional regulators enrichment analysis. STRING database for protein-protein interaction network analysis. GREP platform for druggability assessment. Code Repository: The custom analysis code generated in this study is publicly available without restrictions in the GitHub repository: https://github.com/beimagic/Suicide_PlasmaProteins.

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 individual-level data from UK Biobank (https://www.ukbiobank.ac.uk) that support the findings of this study are not publicly available due to privacy and ethical restrictions. Access to this data requires approval through the UK Biobank Access Management System (https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access). This study used the UK Biobank Resource under application number 19542.
Publicly available summary-level data were obtained from the following sources:
GWAS summary data for plasma proteins: The UK Biobank Pharma Proteomics Project (http://ukb-ppp.gwas.eu).
European ancestry reference data: The 1000 Genomes Project (https://github.com/getian107/PRScsx).
The GWAS summary statistics for suicide attempts were obtained for this specific study from the PGC SUI Data Access Committee under a data use agreement that does not permit public redistribution. Researchers interested in these specific data must apply directly to the committee for access at https://pgc.unc.edu/for-researchers/data-access-committee/data-access-portal/. The PGC data access committee typically provides an initial response to data access requests within 1-2 weeks.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	The study included both male and female participants from the UK biobank. Sex (Field ID 31) in the UK Biobank was determined based on self-reporting data via a questionnaire. Summary statistics on sex distributions were reported in Table 1 of Supplementary Materials. All statistical models were adjusted for sex to account for its potential influence.
Reporting on race, ethnicity, or other socially relevant groupings	The study cohort, drawn from the UK Biobank, was predominantly white (>93%) in both the cross-sectional and longitudinal datasets. We observed no significant difference in ethnicity distribution between the suicidal cases and controls, as detailed in Table 1 of Supplementary Materials. all association analyses were adjusted for ethnicity to account for this potential confounder.
Population characteristics	In the cross-sectional analysis, the 268 cases of suicidal behaviour had a higher proportion of females (59.33%) and were younger (mean age 52.09 ± 7.81 years) compared to the 52,662 controls (53.89% female; mean age 56.83 ± 8.21 years). A similar age difference was observed in the longitudinal dataset, which included 202 cases (47.03% female; mean age 54.55 ± 8.73 years) and 52,303 controls (53.91% female; mean age 56.83 ± 8.20 years). Further demographic details are provided in Table 1 of Supplementary Materials. All statistical analyses were adjusted for age and sex.
Recruitment	The UK Biobank is a prospective, population-based cohort that recruited more than 500,000 participants aged 37 - 73 years who attended 1 of 22 assessment centers across the United Kingdom between 2006 and 2010. The assessment visits comprised interviews and questionnaires covering lifestyles and health conditions, physical measures, biological samples, imaging, and genotype data.
Ethics oversight	UK Biobank has received ethical approval from the North West Multi-centre Research Ethics Committee (MREC, https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics), and informed consent through electronic signature was obtained from study participants. This study utilized the UK Biobank Resource under application number 19542.

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	Sample sizes were not predetermined using statistical methods. We utilized the entire available dataset with plasma protein data from the UK Biobank. In total, 268 cases (all suicide attempts) and 52,662 controls with plasma proteomic data were included in the cross-sectional analysis, while 202 cases (83% suicide attempts and 17% deaths by suicide) and 52,303 controls were included in the longitudinal analysis.
Data exclusions	Three proteins were excluded from the analysis due to more than 50% missing values.
Replication	All available data were used to maximize the statistical power of the analysis. No other cohort of this magnitude available for replication at this stage.
Randomization	All plasma protein associations were adjusted for sex, age, assessment centres, ethnicity, socioeconomic status, BMI, protein batch, the time lag between blood sample collection and proteomic testing, smoking status, and six categories of baseline disease diagnoses. All co-regulated protein network associations were adjusted for sex, age, assessment centres, ethnicity, socioeconomic status, BMI, smoking status, and six categories of baseline disease diagnoses.
Blinding	Blinding was not applicable to this study as this study is observational.

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
☒	☐ Plants

Methods

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

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

Magnetic resonance imaging

Experimental design

Design type	Structural MRI and diffusion MRI

Design specifications

The UK Biobank designed the imaging acquisition protocols including 6 modalities, covering structural, diffusion and functional imaging. The collection order is T1-weighted structural image, resting-state functional MRI, task functional MRI, T2-weighted FLAIR structural image, diffusion MRI and susceptibility-weighted imaging. The T1-weighted structural MRI data was acquired using straight sagittal orientation for 5 minutes. The diffusion MRI data was acquired for 7 minutes (including 36 seconds phase-encoding reversed data).

Behavioral performance measures

We used the T1-weighted structural imaging and diffusion imaging, which do not require task performance.

Acquisition

Imaging type(s)

T1-weighted structural MRI and diffusion MRI

Field strength

3T

Sequence & imaging parameters

3D MPRAGE sequence was used for T1-weighted MRI data and SE-EPI sequence was used for diffusion MRI data. The detailed scanning parameters is described in the references (Alfaro-Almagro, NeuroImage (2018); Smith et al., NatureNeuroscience (2021); Miller et al., Nature Neuroscience (2016) and Elliott et al., Nature (2018)),as well as the online refs (https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf).

Area of acquisition

Whole brain

Diffusion MRI ☒ Used ☐ Not used

Parameters

Diffusion MRI data in the UK Biobank is obtained with two b-values (b = 1,000 and 2,000 s/mm2) at a spatial resolution of 2 mm using a multiband acceleration factor of 3, which allows for the acquisition of three slices simultaneously. For the two diffusion-weighted shells, 50 distinct diffusion-encoding directions were acquired (and all 100 directions are distinct). The diffusion preparation is a standard ("monopolar") Stejskal-Tanner pulse sequence. This enables higher SNR due to a shorter echo time (TE=92ms) than a twice-refocused ("bipolar") sequence. This improvement comes at the expense of stronger eddy current distortions, which are removed in the image processing pipeline.

Preprocessing

Preprocessing software

The gray matter volume was computed with FreeSurfer v6.0.0. The diffusion MRI was processed with FSL.

Normalization

The gray matter volume from T1w MRI images was measured in subject space,rather than standard space, so the T1w images were not normalized. The diffusion weighted images were normalized with FNIRT-based nonlinear warping in FSL. Specifically, although the Eddy and BEDPOSTx outputs are in the space and resolution of the (GDC-unwarped) native diffusion data space, the nonlinear transformation between subject space and 1mm MNI standard space is used to create tractography results in 1mm standard space. Detailed processing can be found at https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf

Normalization template

As mentioned above, the T1w images were in subject space and dMRI images were normalized to FMRIB58_FA_1mmstandard space.

Noise and artifact removal

Details are available at https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf or the references (Alfaro-Almagro, NeuroImage (2018); Smith et al., Nature Neuroscience (2021); Miller et al., Nature Neuroscience (2016) and Elliott et al., Nature (2018)). Briefly,the dMRI data is corrected for eddy currents and head motion, and has outlier-slices (individual slices in the 4D) corrected.

Volume censoring

Detailed in Alfaro-Almagro, NeuroImage (2018); Smith et al., Nature Neuroscience (2021); Miller et al., Nature Neuroscience (2016) and Elliott et al., Nature (2018).

Statistical modeling & inference

Model type and settings

The R-implemented "lm" function was used to examine lineal associations between plasma proteins levels and brain structural measurements. Each association test included covariates for sex, age, assessment centres, ethnicity, socioeconomic status, BMI, protein batch, time lag, smoking status, and six categories of baseline disease diagnoses, with total intracranial volume (TIV) included as an additional covariate for analyses of grey matter volumes. Similar analyses were conducted to assess the associations between co-regulated protein networks expressions and brain structural measurements, except for adjusting for protein batch and time lag.

Effect(s) tested

Neither ANOCA nor factorial designs were used.

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Anatomical location(s)

A total of 68 cortical grey matter volumes defined by the Desikan-Killiany atlas and 14 subcortical volumes defined by the ASEG atlas were included. Measures of white matter microstructure, including FA (fractional anisotropy), MD (mean diffusivity), ICVF (intra-cellular volume fraction), OD (orientation dispersion), and ISOVF (isotropic volume fraction), which were available for 27 major tracts mapped across the brain were used.

Statistic type for inference (See Eklund et al. 2016)	ROI-wise statistical inference was made.
Correction	Statistical significance was assessed using FDR correction across protein-grey matter volume pairs, protein-white matter microstructure pairs, network-grey matter volume pairs, and network-white matter microstructure pairs, respectively.

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis