nature human behaviour

Article

https://doi.org/10.1038/s41562-024-01903-x

Identifying behaviour-related and physiological risk factors for suicide attempts in the UK Biobank

Received: 22 October 2023

Accepted: 29 April 2024

Published online: 02 July 2024

Check for updates

Bei Zhang[1,2], Jia You[1,2], Edmund T. Rolls[1,3,4], Xiang Wang[5,6,7], Jujiao Kang[1,2], Yuzhu Li[1,2], Ruohan Zhang[4], Wei Zhang[1,2], Huifu Wang[8], Shitong Xiang[1,2], Chun Shen[1,2], Yuchao Jiang[1,2], Chao Xie[1,2], Jintai Yu[1], Wei Cheng[1,2,9,10] ✉ & Jianfeng Feng[1,2,4,9,10] ✉

Suicide is a global public health challenge, yet considerable uncertainty remains regarding the associations of both behaviour-related and physiological factors with suicide attempts (SA). Here we first estimated polygenic risk scores (PRS) for SA in 334,706 UK Biobank participants and conducted phenome-wide association analyses considering 2,291 factors. We identified 246 (63.07%) behaviour-related and 200 (10.41%, encompassing neuroimaging, blood and metabolic biomarkers, and proteins) physiological factors significantly associated with SA-PRS, with robust associations observed in lifestyle factors and mental health. Further case–control analyses involving 3,558 SA cases and 149,976 controls mirrored behaviour-related associations observed with SA-PRS. Moreover, Mendelian randomization analyses supported a potential causal effect of liability to 58 factors on SA, such as age at first intercourse, neuroticism, smoking, overall health rating and depression. Notably, machine-learning classification models based on behaviour-related factors exhibited high discriminative accuracy in distinguishing those with and without SA (area under the receiver operating characteristic curve 0.909 ± 0.006). This study provides comprehensive insights into diverse risk factors for SA, shedding light on potential avenues for targeted prevention and intervention strategies.

Suicide is a major public health concern, encompassing a spectrum of suicidal ideation, plans and attempts, with suicide attempts (SA) being a leading cause of death worldwide[1]. While previous research has explored various factors associated with attempted suicide[2–4], critical research gaps remain.

The existing literature on SA has generally focused on identifying a limited set of hypothesized factors, such as psychiatric disorders (that is, depression), personality and psychological characteristics (that is, hopelessness) and social and family factors (that is, low social support and stressful life events), often within relatively small clinical samples[4–10]. This narrow focus may result in additional risk factors being overlooked or unknown. Likewise, the relative impact of multi-faceted risk factors for SA remains largely unknown. Certain factors might be statistically significant when studied alone, yet they may not demonstrate strong resilience when assessed in conjunction with other variables[11]. A more comprehensive examination of a broader range of factors within a large-scale population-based study could help confirm existing relationships and reveal new potential targets for intervention.

A full list of affiliations appears at the end of the paper. ✉e-mail: wcheng@fudan.edu.cn; jianfeng64@gmail.com

https://doi.org/10.1038/s41562-024-01903-x

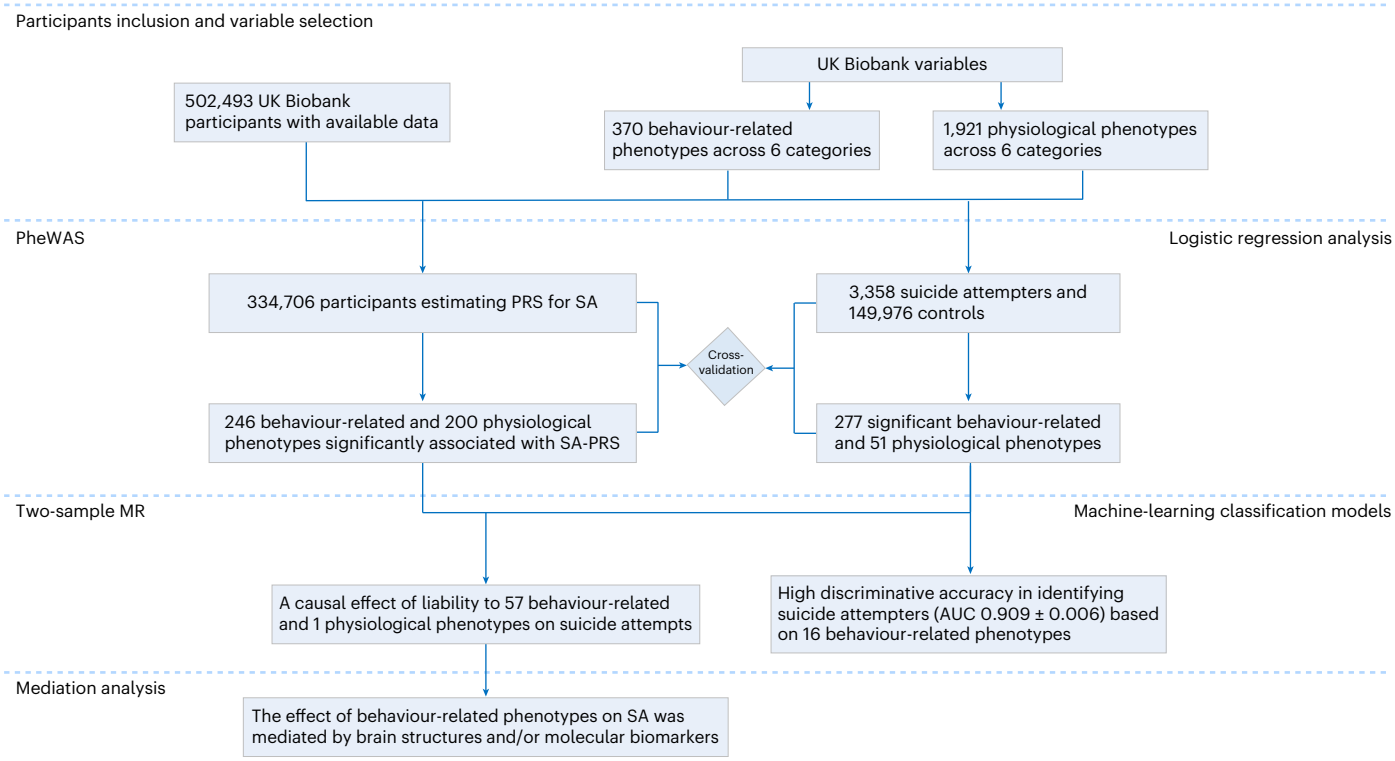

Fig. 1 | Overview of the study. Analytical procedures to identify behaviour-related and physiological risk factors associated with SA in the UK Biobank.

Moreover, recent suicide-related models[5,12–14], such as the biopsychosocial model[5], highlighted the complexity of suicide risk development, implicating a broad range of contributors from biological, psychological, clinical, social and environmental factors. However, prior studies have predominately emphasized the influence of behaviour-related risk factors on SA[2,3]. Recent genome-wide association studies (GWASs) have suggested a polygenic architecture underlying SA[10,15,16], implicating the potential role of genetic associations (for example, polygenic risk scores (PRS)) in identifying risk factors[17]. To gain a more comprehensive understanding, it is crucial to investigate both behaviour-related and physiological systems concurrently, shedding light on their complex interplay. Last, very few studies, to our knowledge, have evaluated the causal relationships between risk factors and SA[18], highlighting the need for a deeper investigation into underlying causal mechanisms.

The UK Biobank provides an unparalleled opportunity to narrow these gaps, given that it collects rich information across multiple domains with a large sample size. Advanced analytical approaches, such as phenome-wide association study (PheWAS)[19] and Mendelian randomization (MR) analyses[20], can be employed to rigorously assess the putative associations and reduce false positive findings[21–23]. The PheWAS, in particular, offers valuable insights for identifying risk factors that probably reflect underlying causal relationships, as it is less constrained by prior assumptions, leverages solid biological knowledge from birth, making it less susceptible to confounding and reverse causality, and has the potential to identify risk factors in the early stages of disease development[24–26].

In this article, harnessing the power of the UK Biobank, we aimed to systematically assess and rank both behaviour-related and physiological risk factors for SA using several data-driven approaches. Specifically, by applying PRS for SA as a proxy for SA risk, we first conducted PheWAS to identify factors associated with genetic predisposition to SA comprehensively. Subsequently, logistic regression analyses for individuals who have attempted suicide and a control group were also performed within the case–control framework. Moreover, we employed bidirectional MR to evaluate potential causal relationships between SA and the identified associated risk factors observed in the PheWAS and logistic regression analyses. Finally, machine-learning models were developed to ascertain the effectiveness of risk factors in distinguishing those with and without SA (see Fig. 1 for an overview of the study).

Results

Identifying behaviour-related risk factors for SA in PheWAS

Of the 334,706 individuals included in the calculation of PRS for SA (SA-PRS) and the subsequent PheWAS, 53.59% were female, and the mean (standard deviation, s.d.) age was 56.91 (7.99) years (see Supplementary Table 2 for detailed demographic characteristics). In the PheWAS analysis, we considered a total of 2,291 factors spanning 12 categories encompassing both behaviour-related and physiological phenotypic types, as outlined in Table 1 and Supplementary Data Table 1. Sample sizes for each category are summarized in Table 1. In total, we found 246 (63.07% of those examined) behaviour-related phenotypes and 200 (10.41%) physiological phenotypes that showed significant associations with SA-PRS after Bonferroni correction for multiple comparisons ($P < 1.94 \times 10^{-5}$ (0.05/2,576), including multicategorical variables). A list of significant results from the PheWAS analysis is presented in Supplementary Data Table 2.

Among the examined behaviour-related factors, 246 out of 370 showed significant associations with SA-PRS after Bonferroni correction ($P < 1.94 \times 10^{-5}$ (0.05/2,576); Fig. 2a). These significant factors included 11 sociodemographic, 43 lifestyle factors, 11 early-life and family-history factors, 36 physical measures, 16 cognitive functions and 129 mental health (136 associations, including multiple categorical associations) (absolute $\beta > 0.006$, β are standardized regression coefficients throughout).

Higher SA-PRS was associated with greater Townsend deprivation index, lower household income, less likely to own the accommodation

https://doi.org/10.1038/s41562-024-01903-x

Table 1 | Summary of number of phenotypes and sample sizes within a category

Category	Number of factors	Sample size	
		PheWAS (mean [range])	Logistic regression analyses (mean SA cases versus controls)
Behaviour-related phenotypes	370		
Sociodemographic	13	238,542 [30,537–334,258]	2,338 versus 95,036
Lifestyle factors	73	254,436 [1,208–334,616]	2,826 versus 117,533
Early-life and family-history factors	17	248,230 [74,698–334,201]	2,653 versus 112,941
Mental health	177	102,734 [780–333,488]	2,486 versus 101,892
Physical measures	64	226,815 [7,110–334,072]	2,501 versus 102,710
Cognitive functions	26	101,490 [1,999–334,335]	1,082 versus 47,363
Physiological phenotypes	1,921		
Regional grey matter volumes	94	27,092 [27,092–27,092]	599 versus 26,048
White matter microstructure	135	27,880 [27,880–27,881]	602 versus 26,599
Blood count	31	323,351 [319,508–324,755]	3,375 versus 143,089
Blood biochemistry	30	287,541 [28,479–319,079]	2,987 versus 127,229
NMR metabolomics	168	80,521 [77,032–80,567]	851 versus 35,251
Proteomic phenotypes	1,463	34,329 [9,365–35,307]	345 versus 15,261

A total of 370 behaviour-related phenotypes (6 categories) and 1921 physiological phenotypes (6 categories, including 229 neuroimaging variables (2 categories), 229 blood and metabolic biomarkers (3 categories), and 1,463 unique proteins) are included. Where there are multiple sample sizes in a category, the mean sample size (*N*) is presented. NMR, nuclear magnetic resonance.

that they live in, heavy manual and shift jobs, and younger age completing full-time education (absolute $\beta > 0.017$, $P_{BOF} = 3.81 \times 10^{-156}$ to 2.22×10^{-9}). Besides, SA-PRS was significantly associated with multiple lifestyle factors, reporting that sexual factors (for example, lifetime number of sexual partners), sleep problems (for example, insomnia), smoking behaviours (for example, both current and ever smoking), alcohol consumption (for example, alcohol intake frequency), diet (for example, salt added to food), electronic device use (for example, usage of mobile phone and playing computer games) and physical activity (for example, time spent watching television) were positively correlated with SA-PRS ($\beta = 0.016$ to 0.153, $P_{BOF} = 8.13 \times 10^{-148}$ to 0.015). In contrast, some sexual factors and diet, such as age at first intercourse, and cereal, cheese and dried fruit intake, were negatively associated with SA-PRS ($\beta < -0.015$, $P_{BOF} < 0.032$).

Moreover, Bonferroni-significant associations were also identified between higher SA-PRS and maternal smoking around birth, adoption and more siblings ($\beta = 0.040$ to 0.119, $P_{BOF} = 4.24 \times 10^{-79}$ to 6.79×10^{-12}), as well as a range of disease factors such as cancer, diabetes and body mass index (BMI; $\beta = -0.090$ to 0.099, $P_{BOF} = 5.17 \times 10^{-183}$ to 0.038), with the most significant item being poorer overall health ratings. SA-PRS was also related to cognitive functions, such as fluid intelligence, prospective memory, trail making, pairs matching and reaction time (absolute $\beta = -0.085$ to 0.074, $P_{BOF} = 7.39 \times 10^{-54}$ to 0.021). Better performance on all cognitive tests was associated with lower SA-PRS. Substantial significant associations were identified in self-reported psychological traits, including the sum of traumatic events, depressive symptoms, anxiety symptoms, mental distress, subjective wellbeing, psychotic experiences, mania, neuroticism score and so on (absolute $\beta > 0.021$, $P_{BOF} = 4.41 \times 10^{-229}$ to 0.045). Of these, in addition to happiness and subjective wellbeing, more severe symptoms were associated with higher SA-PRS.

Identifying physiological risk factors for SA in PheWAS

The SA-PRS demonstrated significant associations with 200 out of 1,921 physiological phenotypes that survived Bonferroni correction ($P < 1.94 \times 10^{-5}$ (0.05/2,576), absolute $\beta = 0.008$ to 0.050). These associations comprised 20 neuroimaging phenotypes (including 13 grey matter volumes and 7 white matter microstructures), 76 blood and metabolic biomarkers (encompassing 18 blood cell counts, 14 blood biochemistries and 44 nuclear magnetic resonance (NMR) metabolites) and 104 proteins. Detailed results of the PheWAS analysis are presented in Supplementary Data Table 2.

To start with the neuroimaging phenotypes, lower grey matter volumes were associated with higher SA-PRS in 13 out of 94 brain regions following Bonferroni correction ($\beta = -0.023$ to -0.019, $P_{BOF} = 5.10 \times 10^{-4}$ to 0.044), including the right insula, amygdala, ventromedial prefrontal cortex (vmPFC), medial and lateral orbitofrontal cortex (OFC), supramarginal gyrus, left superior temporal gyrus, bilateral middle cingulate gyri, Heschl's gyrus and Rolandic operculum (Fig. 3a). Besides, Bonferroni-significant associations were also found between SA-PRS and 7 out of 135 white matter microstructures (Fig. 3b). Decreased fractional anisotropy (FA, such as in forceps minor, medial lemniscus, and posterior and superior thalamic radiation, $\beta = -0.033$ to -0.026, $P_{BOF} = 3.87 \times 10^{-5}$ to 0.031) and increased orientation dispersion (OD, medial lemniscus, $\beta = 0.026$, $P_{BOF} = 0.020$) were associated with higher SA-PRS.

For blood and metabolic biomarkers (Fig. 4a), all 18 significant blood cell counts associations showed an identical effect direction with SA-PRS ($\beta = 0.008$ to 0.029, $P_{BOF} = 1.46 \times 10^{-55}$ to 0.005) except for monocyte percentage. That is, higher white blood cell counts (for example, leukocytes and lymphocyte count) and red blood cell index (for example, light scatter reticulocyte count) were associated with higher SA-PRS ($\beta = 0.008$ to 0.029, $P_{BOF} = 1.46 \times 10^{-55}$ to 4.95×10^{-3}). In addition, for blood biochemistry, higher C-reactive protein, triglycerides and gamma glutamyltransferase ($\beta = 0.009$ to 0.029, $P_{BOF} = 2.21 \times 10^{-64}$ to 1.65×10^{-3}), and lower IGF-1, vitamin D and high-density lipoprotein (HDL) cholesterol ($\beta = -0.020$ to -0.008, $P_{BOF} = 2.17 \times 10^{-28}$ to 1.45×10^{-3}) demonstrated the most significant correlations with higher SA-PRS. Similarly, SA-PRS was positively correlated with inflammation variables (for example, glycoprotein acetyls) and triglyceride variables (for example, triglycerides in small HDL) but negatively correlated with cholesterol variables (for example, cholesterol in very large HDL) within NMR metabolomics ($\beta = -0.024$ to 0.027, $P_{BOF} = 4.06 \times 10^{-11}$ to 0.038).

Of all 1463 unique proteins, 104 showed significant associations with SA-PRS after Bonferroni correction (absolute $\beta = 0.015$ to 0.050, $P_{BOF} = 1.50 \times 10^{-17}$ to 0.049; Fig. 4b,c). Relative to proteins in other panels, associations of inflammatory proteins with SA-PRS were more pronounced, with BTN3A2 and CXCL17 showing the most significant associations ($\beta = 0.036$ to 0.050, $P_{BOF} = 1.50 \times 10^{-17}$ to 2.15×10^{-11}).

Cross-validation of risk factors in case–control analyses

To validate the risk factors identified in PheWAS, we performed logistic regression analyses for all phenotypes on a dataset comprising 3,558 SA cases and 149,976 controls (56.28% female; mean [s.d.] age, 56.03 [7.72] years). The demographic characteristics of the two groups are detailed in Supplementary Table 2, and the mean sample sizes for each category are reported in Table 1. Our logistic regression results revealed that 277 behaviour-related phenotypes (334 associations,

https://doi.org/10.1038/s41562-024-01903-x

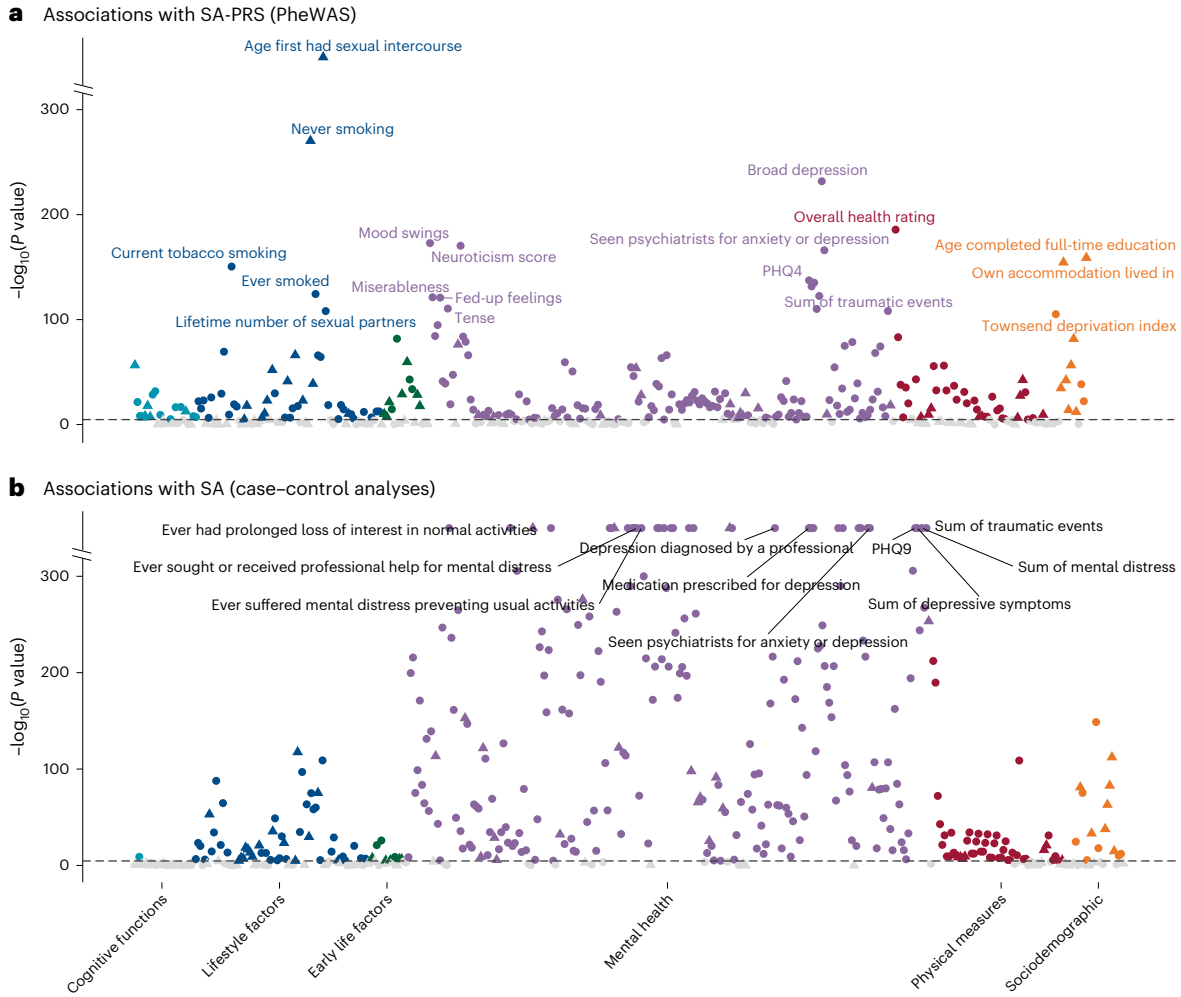

Fig. 2 | Significance plots for behaviour-related phenotypes associated with PRS for SA in PheWAS and associated with SA in case–control analyses. **a**, Significance plots for behaviour-related phenotypes associated with PRS for SA in PheWAS. Triangles indicate a negative association with PRS for SA, whereas dots represent a positive association. The top 22 significant phenotypes are annotated ($P < 1.94 \times 10^{-5}$ (0.05/2,576)). Note that higher overall health rating scores indicated poorer health in the UK Biobank (Field ID 2178). **b**, Significance plots for behaviour-related phenotypes associated with SA in case–control analyses. Triangles indicate a negative association with SA, whereas dots represent a positive association. The top 10 significant phenotypes are annotated ($P < 2.12 \times 10^{-5}$ (0.05/2,360)). In both plots, each dot or triangle corresponds to a specific phenotype, and their respective colours denote their corresponding categories. The *x* axis illustrates six distinct categories, and the *y* axis represents the $-\log_{10}$ of uncorrected *P* values of the two-sided test. Dashed black lines denote the Bonferroni threshold for multiple comparisons ($\alpha = 0.05$). The complete significant results are detailed in Supplementary Data Tables 2 and 3. PHQ4, Patient Health Questionnaire 4 and its four subitems; PHQ9, Patient Health Questionnaire 9.

including multiple categorical associations) reached significance after Bonferroni correction ($P < 2.12 \times 10^{-5}$ (0.05/2,360), $\beta = -4.339$ to 8.940). Of all behaviour-related categories, mental health stands out as the most significant (Fig. 1b). Notably, about 83% of significant behaviour-related associations identified in PheWAS also remained significant in the logistic regression analyses, highlighting the consistency of PRS in identifying behaviour-related risk factors with case–control investigations.

In addition, our logistic regression analyses identified that 51 physiological phenotypes reached statistical significance following Bonferroni correction ($P < 2.12 \times 10^{-5}$ (0.05/2,360)), including 29 blood and metabolic biomarkers (11 blood cell counts, 17 blood biochemistries and 1 NMR metabolite, $\beta = -2.386$ to 8.056, $P_{BOF} = 3.73 \times 10^{-22}$ to 0.041), and 22 proteins ($\beta = -0.912$ to 0.845, $P_{BOF} = 1.74 \times 10^{-8}$ to 0.037). Although the significant physiological factors identified in logistic regression analyses were somewhat less consistent in PheWAS (37 duplicates), the top significant factors within a category displayed similarity. For instance, proteins such as CXCL17, blood biochemistry markers such as triglycerides, C-reactive protein, gamma glutamyltransferase and IGF-1, and blood cell measures such as neutrophil and leukocyte count emerged as top significant phenotypes in their respective categories in both the PheWAS and logistic regression analyses. A comprehensive list of significant results in the logistic regression analyses is presented in Supplementary Data Table 3.

MR on risk factors and SA

To explore the potential causal effects of phenotypes on SA, we first conducted MR analyses on risk factors that were both significant in PheWAS and case–control analyses and SA. Results obtained using the inverse variance weighted (IVW) method showed that, out of the 198 examined factors, 86 behaviour-related factors reached significance at $P_{IVW} < 0.05$ and 57 remained significant after Bonferroni correction ($P_{IVW} < 2.53 \times 10^{-4}$ (0.05/198); Fig. 5). These Bonferroni-corrected factors consistently exhibited the same direction of effect observed in the PheWAS and case–control analyses. Specifically, these factors consisted of four sociodemographic factors (odds ratio $(OR)_{IVW} = 0.521$ to 2.041, $P_{BOF} = 1.77 \times 10^{-5}$ to 0.032), including less average household income, higher Townsend deprivation index, more numbers of vehicles in household, and heavy manual or physical

https://doi.org/10.1038/s41562-024-01903-x

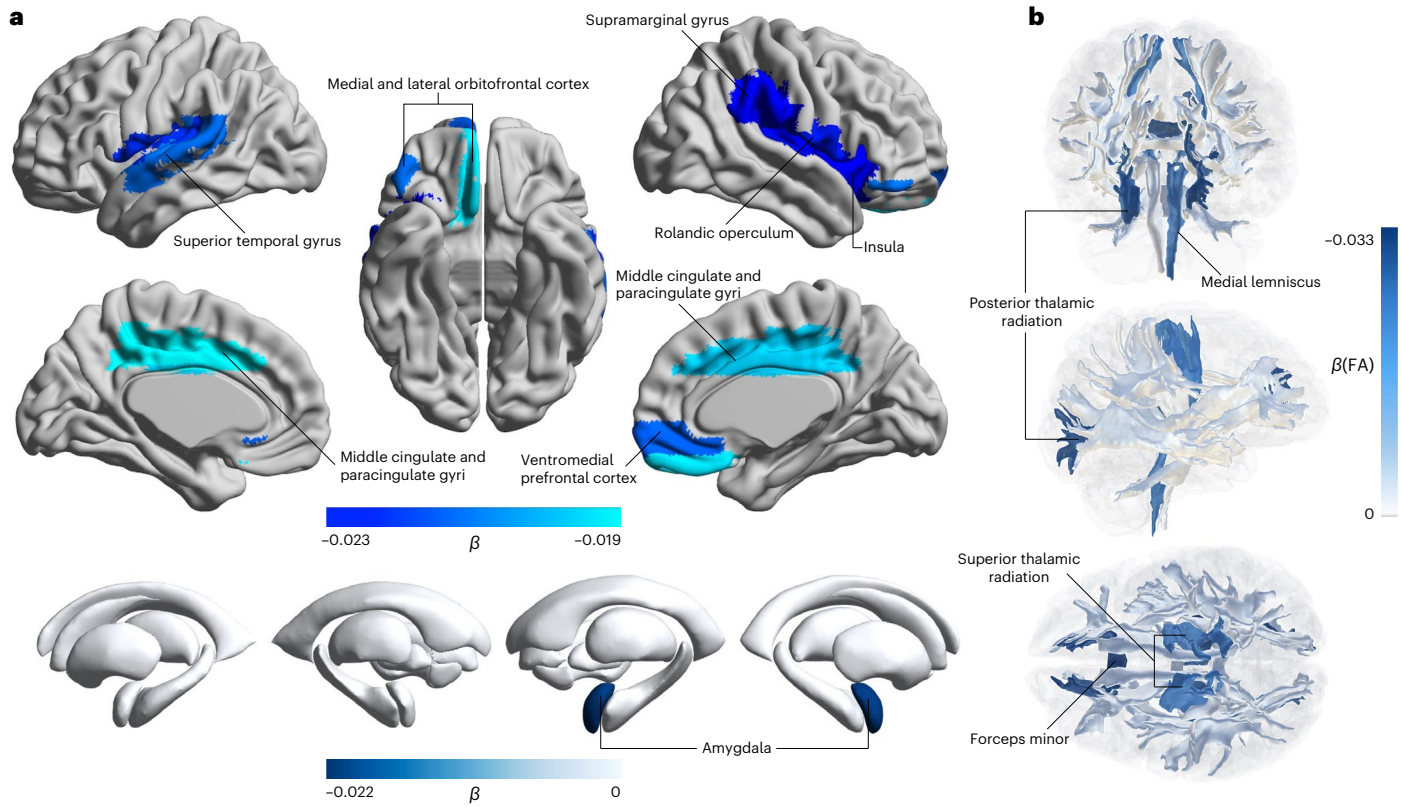

Fig. 3 | Maps of significant associations between PRS for SA and neuroimaging phenotypes. a, Lower grey matter volumes of cortical (top) and subcortical (bottom) regions are significantly associated with higher PRS for SA. These brain regions survived Bonferroni correction for multiple comparisons ($P < 1.94 \times 10^{-5}$ (0.05/2,576)) in the two-sided test for linear regression between PRS for SA and each phenotype. **b**, Decreased white matter microstructure in FA are significantly associated with higher PRS for SA. These brain regions survived Bonferroni correction for multiple comparisons ($P < 1.94 \times 10^{-5}$ (0.05/2,576)) in the two-sided test for linear regression between PRS for SA and each phenotype. In both plots, the shading represents the standardized effect size (β), with darker shading indicating larger absolute of β values.

jobs, 11 measures in lifestyle (OR_{IVW} = 0.239 to 5.888, $P_{BOF} = 8.32 \times 10^{-81}$ to 0.045), including insomnia, smoking, sexual factors (age at first intercourse and lifetime number of sexual partners), alcohol intake and diets (cereal intake and salt added to food), 18 physical measurements (OR_{IVW} = 0.746 to 3.370, $P_{BOF} = 2.83 \times 10^{-31}$ to 0.040) and 24 measures in mental health (OR_{IVW} = 1.218 to 53.521, $P_{BOF} = 5.84 \times 10^{-36}$ to 0.013). Important physical measurements included overall health rating (poorer health), higher BMI, long-standing illness, disability or infirmity, and higher body fat percentage. Among mental health variables, neuroticism had the most significant effect, followed by depression, seeing psychiatrists for anxiety or depression, sum of traumatic events, loneliness and so on.

In addition, we also found nine physiological phenotypes as potential causal factors for SA at $P_{IVW} < 0.05$, but only one biomarker remained significant after Bonferroni correction: white blood cell (leukocyte) count (OR_{IVW} = 1.110, $P_{BOF} = 3.74 \times 10^{-4}$). For the above Bonferroni-significant results, no factors showed substantial horizontal pleiotropy (as indicated by $P_{Intercept}$ for all MR–Egger intercept >0.01). For risk factors showing potential single-nucleotide polymorphisms (SNPs) heterogeneity, we employed an additional multiplicative random-effects model for validation. This validation approach confirmed the retention of all significant factors ($P_{BOF} < 0.045$). Additionally, results derived using the weighted median method were overall similar to those identified using the IVW method. Further details regarding forward MR results are presented in Supplementary Data Table 4.

Except for the phenotypes described above, we also performed MR between SA and factors that were significant only in PheWAS or case–control analyses. We identified nine additional factors as causal factors for SA after Bonferroni correction ($P_{IVW} < 1.83 \times 10^{-4}$ (0.05/273)), including younger age completing full-time education (OR_{IVW} = 0.575, $P_{BOF} = 1.04 \times 10^{-9}$), higher fluid intelligence score (OR_{IVW} = 0.921, $P_{BOF} = 3.64 \times 10^{-4}$) and BTN3A2 protein ($OR_{IVW}$ = 1.052, P_{BOF} = 0.038). For a comprehensive overview of the forward MR results obtained from the three different methods and sensitivity analyses for all tested factors, please refer to Supplementary Data Table 4.

Conversely, we identified potential causal effects of SA on 55 factors that survived Bonferroni correction ($P_{IVW} < 2.63 \times 10^{-4}$ (0.05/190)). Among these, 29 showed substantial horizontal pleiotropy ($P_{Intercept} < 0.05$) but failed to reach significance in the MR–Egger regression method. Therefore, Bonferroni-corrected significant causal effects of SA on 26 phenotypes were identified (Supplementary Fig. 1). These factors included one physical measure (overall health rating, OR_{IVW} = 1.048, $P_{BOF} = 2.62 \times 10^{-5}$), 6 lifestyle factors (for example, alcohol usually taken with meals; OR_{IVW} = 0.966 to 1.064, $P_{BOF} = 3.74 \times 10^{-10}$ to 0.045), 13 measures in mental health (for example, risk-taking and sum of psychotic experiences; OR_{IVW} = 0.969 to 1.088, $P_{BOF} = 1.54 \times 10^{-5}$ to 0.019), 5 sociodemographic (for example, average total household income; OR_{IVW} = 0.941 to 1.039, $P_{BOF} = 5.91 \times 10^{-5}$ to 0.005) and 1 protein (alkaline phosphatase, placental type (ALPP); OR_{IVW} = 1.082, P_{BOF} = 0.029). These factors remained significant in the multiplicative random-effects models ($P_{BOF} < 0.046$). Of the 271 factors in the additional analyses, four factors survived Bonferroni correction as causal outcomes of SA ($P_{IVW} < 1.83 \times 10^{-4}$ (0.05/271)), including less work satisfaction (OR_{IVW} = 1.048, P_{BOF} = 0.002) and talking therapies to treat depression (OR_{IVW} = 1.021, P_{BOF} = 0.029). The directional results

https://doi.org/10.1038/s41562-024-01903-x

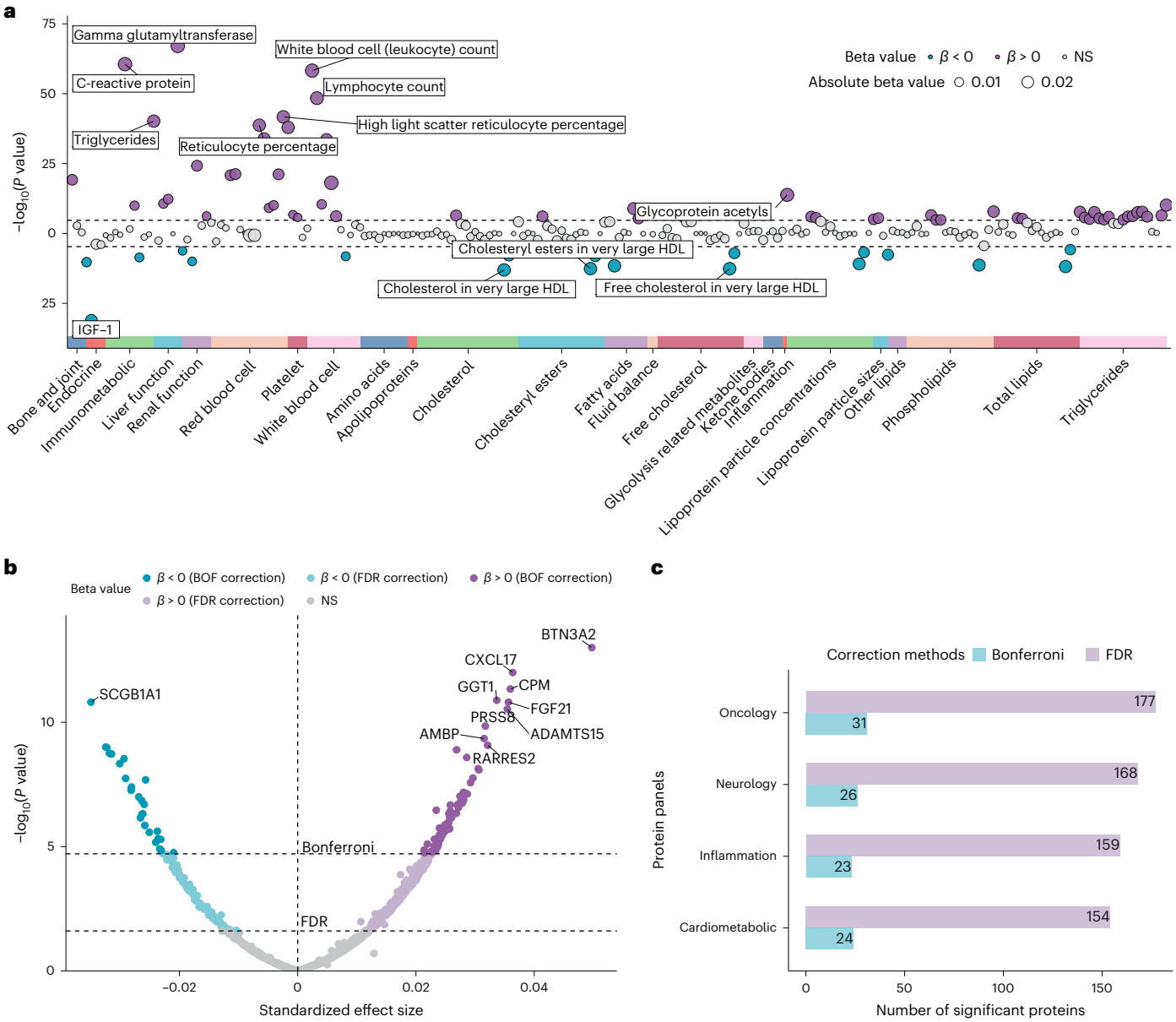

Fig. 4 | Significant associations between blood and metabolic biomarkers and proteins and PRS for SA. a, A Manhattan plot of associations of blood and metabolic biomarkers with PRS for SA. The *x* axis represents 24 different classifications, and the *y* axis represents the $-\log_{10}$ of uncorrected *P* values resulting from the two-sided test for linear regression between PRS for SA and each phenotype. The purple dots at the top represent a positive association with PRS for SA (standardized effect size $\beta > 0$), and the blue dots at the bottom represent a negative association (standardized effect size $\beta < 0$). The dashed lines indicate thresholds to survive Bonferroni correction for multiple comparisons ($\alpha = 0.05$). The top four significant phenotypes of each category are annotated in the figure ($P < 1.94 \times 10^{-5}$ (0.05/2,576)). **b**, A significance plot for proteins associated with PRS for SA. The *x* axis represents the standardized effect size (β), and the *y* axis represents $-\log_{10}$ of uncorrected *P* values resulting from the two-sided test for linear regression between PRS for SA and each phenotype. The dashed lines indicate the threshold to survive Bonferroni (BOF) and false discovery rate (FDR) correction, respectively. The top ten significant phenotypes are annotated in the figure ($P < 1.94 \times 10^{-5}$ (0.05/2,576)). **c**, The number of significant proteins at four different proteomic panels that survived Bonferroni and FDR correction for multiple comparisons (adjusted *P* value <0.05), respectively.

obtained using three methods and sensitivity analyses are detailed in Supplementary Data Table 5.

Machine-learning classification models for SA

To further assess the effectiveness of risk factors in differentiating individuals with and without SA, we developed machine-learning classification models for SA based on behaviour-related, metabolic and proteomic variables, respectively. Taking the behaviour-related SA classification model as an example, as depicted in Fig. 6a, a set of top 30 behaviour-related predictors were exhibited and sorted by their importance to the SA classification task. Following a sequential forward selection strategy, the top 16 variables were chosen as the final predictors for the machine-learning model development. Among these, the most important predictors were suffering mental distress preventing usual activities, feelings of worthlessness during the worst period of depression, age at first episode of depression, seeing psychiatrists for anxiety or depression, and thoughts of death during depression (Fig. 6b). Most of these important predictors were derived from the self-reported questionnaires. For instance, the predictor related to suffering mental distress was obtained through a specific question: 'In your life, have you suffered from a period of mental distress that prevented you from doing your usual activities?'. Further details regarding the

https://doi.org/10.1038/s41562-024-01903-x

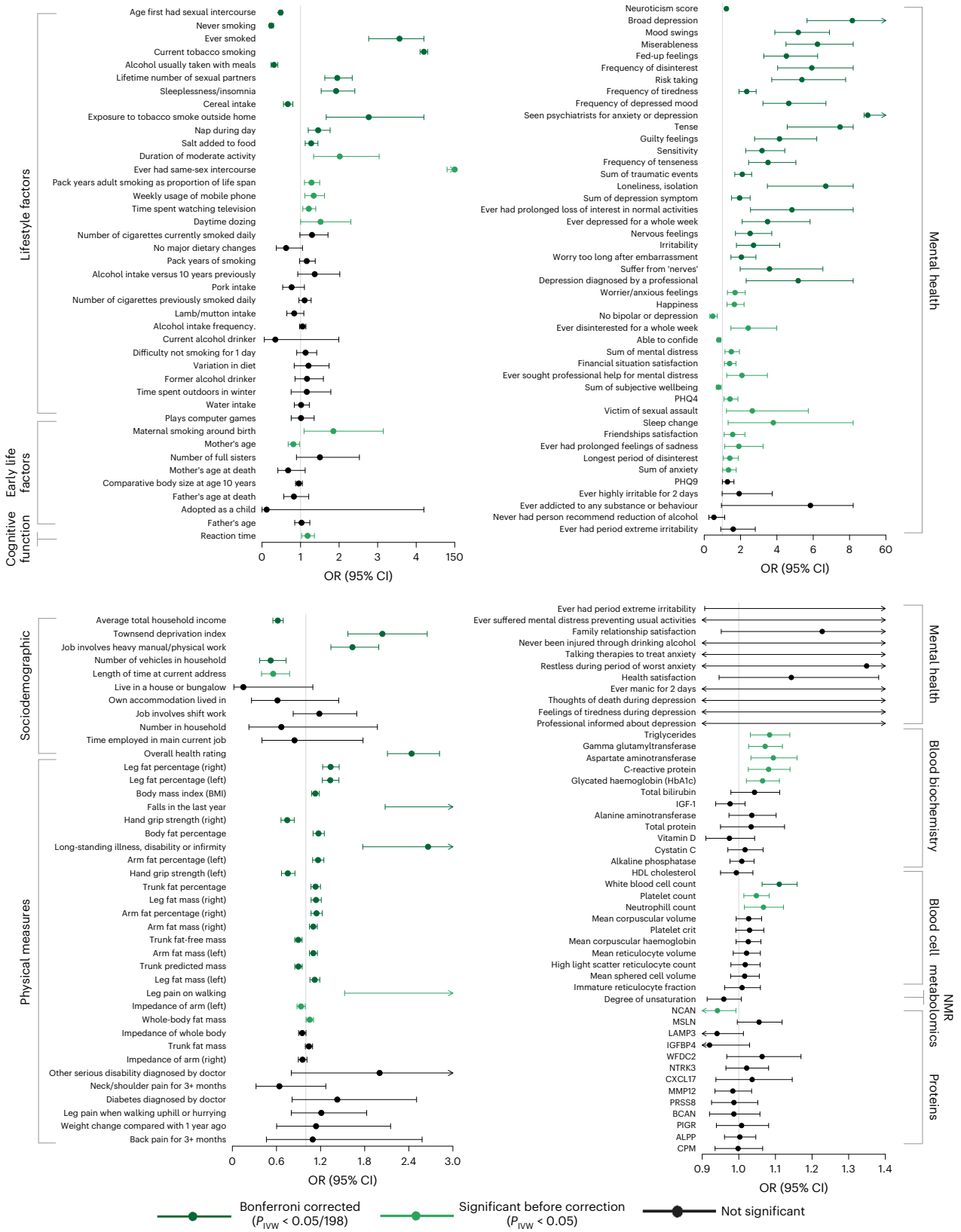

Fig. 5 | MR analysis between risk factors and SA. The primary method employed was the IVW method. Dark-green colour represents phenotypes with causal effects on SA (Bonferroni-corrected, $P_{IVW} < 2.53 \times 10^{-4}$ (0.05/198)). Light-green colour represents phenotypes with causal effects on SA at $P_{IVW} < 0.05$. The dots represent ORs, and horizontal lines indicate corresponding 95% confidence intervals. The sample size for the GWAS summary statistics of SA included 33,353 SA cases and 444,626 controls, and sample sizes for GWAS summary statistics of each phenotype are presented in Supplementary Data Table 4. The top 55 mental health risk factors are shown in the figure, and details of the remaining 25 non-significant mental health factors are presented in Supplementary Data Table 4. Note that, in the UK Biobank, higher scores for happiness (Field ID 4526), financial situation satisfaction (Field ID 4581) and friendship satisfaction (Field ID 4570) indicated less happiness or satisfactions. CI, confidence interval; PHQ4, Patient Health Questionnaire 4; PHQ9, Patient Health Questionnaire 9.

https://doi.org/10.1038/s41562-024-01903-x

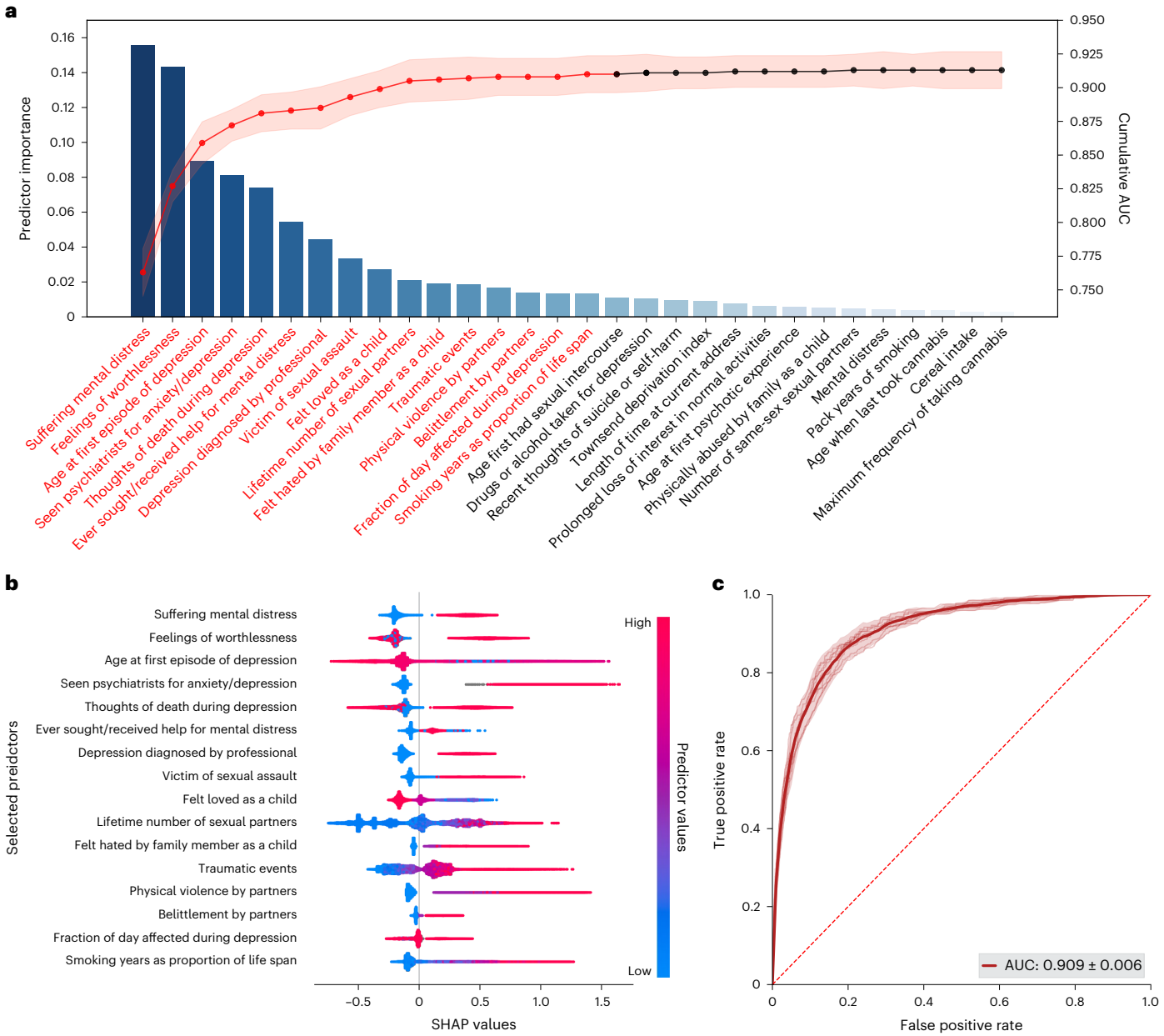

Fig. 6 | Predictor selection, SHAP visualization and performance of machine-learning classification model based on behaviour-related phenotypes. **a**, Sequential forward selection of behaviour-related variables. The bar chart displays variables sorted by importance in the SA classification task. The line chart shows the AUCs (right axis) on the inclusion of variables one by each iteration, and shaded regions represent corresponding 95% confidence intervals derived from cross-validation. The top 16 predictors (coloured in red) were finally selected for model development. **b**, SHAP visualization of selected predictors. Each participant is represented as a data point. The width of the horizontal bars reflects their impact on model predictions, with a wider range indicating a larger impact. The colour of the horizontal bars represents the magnitude of predictors, which was coded in a gradient from blue (low) to red (high), shown as the colour bar on the right-hand side. The *x* axis indicates the likelihood of developing SA (right) or being healthy (left). Using the predictor 'suffering mental distress' as an example, those with more mental distress (coloured red) are more likely to attempt suicide (right side), while those with less mental distress (coloured blue) tend not to attempt suicide (left side). We leveraged all analysed individuals by aggregating samples from each of the testing sets within cross-validation ($N = 153{,}534$). **c**, AUC plots for the behaviour-related phenotype-based SA classification model. The shaded regions represent 95% confidence intervals derived from cross-validation.

acquisition of the top 16 predictors are presented in Supplementary Data Table 6. Overall, the area under the receiver operating characteristic curve (AUC) of the behaviour-related SA classification model reached 0.909 ± 0.006 (Fig. 6c), suggesting a strong discrimination performance. The optimal cutoff value for detecting future SA events was 0.02, determined upon the achievement of the largest Youden index. Under this threshold, the behaviour-related model achieved precision, recall, accuracy, specificity and *F*1 scores of 0.097 ± 0.007, 0.86 ± 0.012, 0.812 ± 0.016, 0.81 ± 0.016 and 0.174 ± 0.011, respectively. Additional relevant statistics calculated under other thresholds ranging from 0.05 to 0.5 at a step of 0.05 are provided in Supplementary Data Table 7. However, the AUCs for the metabolic and proteomic models were notably lower, at 0.601 ± 0.039 and 0.702 ± 0.040, respectively (Supplementary Figs. 2 and 3).

Mediation of behaviour-related risk factors on SA

To examine whether the effect of behaviour-related risk factors on SA could be mediated by brain structures and/or molecular biomarkers,

https://doi.org/10.1038/s41562-024-01903-x

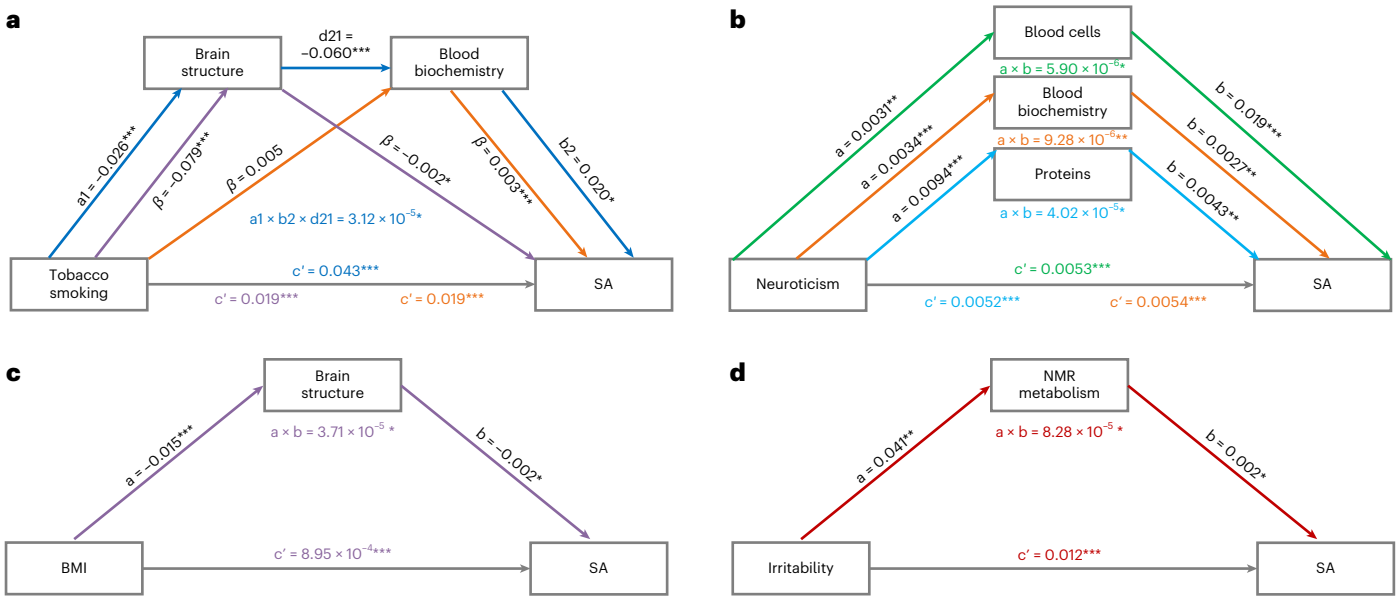

Fig. 7 | Mediation analysis of behaviour-related phenotypes (predictors) on SA (dependent variable) via brain structure and/or molecular biomarkers (mediators). a, The indirect pathway of the effect of tobacco smoking on SA via brain structure and blood biochemistry was significant (indirect effect: $a1 \times b2 \times d21 = 3.12 \times 10^{-5}$, $P = 0.046$, coloured blue). Meanwhile, brain structure (coloured purple), but not blood biochemistry (coloured orange), significantly mediated the association of tobacco smoking with SA. **b**, The effect of neuroticism on SA was significantly mediated by blood cells (coloured green, indirect effect: $a \times b = 5.90 \times 10^{-6}$, $P = 0.017$), blood biochemistry (coloured orange, indirect effect: $a \times b = 9.28 \times 10^{-6}$, $P = 0.004$) and proteins (coloured light blue, indirect effect: $a \times b = 4.02 \times 10^{-5}$, $P = 0.028$), respectively. **c**, Brain structure significantly mediated the effect of body mass index (BMI) on SA (indirect effect of 3.71×10^{-5}, $P = 0.027$, coloured purple). **d**, The association of irritability with SA was significantly mediated by NMR metabolites (indirect effect of 8.28×10^{-5}, $P = 0.048$, coloured red). These associations were estimated by the 10,000-iteration non-parametric bootstrap approach. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

we constructed four mediation models (Methods). We observed that the path from brain structure to blood biochemistry significantly mediated the effect of multiple factors on SA, including tobacco smoking (indirect $\beta = 3.12 \times 10^{-5}$, $P = 0.046$), ever-smoked (indirect $\beta = 3.94 \times 10^{-5}$, $P = 0.039$), seeing psychiatrists for anxiety/depression (indirect $\beta = 2.14 \times 10^{-5}$, $P = 0.047$) and numbers of vehicles in the household (indirect $\beta = -2.20 \times 10^{-5}$, $P = 0.046$; see Fig. 7a for an example of tobacco smoking) in the first mediation model, while none of the significant mediation effects was found via the path from molecular biomarkers to brain structure in the second mediation model. In the third and fourth models, when considering the mediator, separately, we found that the effect of multiple factors on SA was individually mediated by brain structure and/or molecular biomarkers. Specifically, the association of neuroticism with SA was mediated by blood biochemistry (indirect $\beta = 9.28 \times 10^{-6}$, $P = 0.004$), blood cells (indirect $\beta = 5.90 \times 10^{-6}$, $P = 0.017$) and proteins (indirect $\beta = 4.02 \times 10^{-5}$, $P = 0.028$), respectively (Fig. 7b). Similar mediations were identified for factors such as age at first intercourse, depression, sleeplessness, risk-taking and loneliness. The NMR metabolites significantly mediated the effect of irritability on SA (indirect $\beta = 8.28 \times 10^{-5}$, $P = 0.048$; Fig. 7d). Brain structure significantly mediated the effect of BMI on SA (indirect $\beta = 3.71 \times 10^{-5}$, $P = 0.027$; Fig. 7c). More details of the mediation of behaviour-related risk factors on SA are provided in Supplementary Data Table 8.

Gender-stratified analyses

To explore potential sex-specific risk factors associated with PRS for SA, we conducted a gender-stratified analysis in PheWAS. Our results identified 203 behaviour-related and 87 physiological phenotypes significantly associated with SA-PRS for women after Bonferroni correction for multiple comparisons ($\beta > 0.011$, $P_{BOF} = 2.73 \times 10^{-197}$ to 0.046). Similarly, for men, 199 behaviour-related and 73 physiological associations survived Bonferroni correction ($\beta > 0.011$, $P_{BOF} = 3.81 \times 10^{-138}$ to 0.049). Of these identified risk factors, 202 were significant in both genders, indicating a high degree of overlap in risk factors between men and women. However, 88 risk factors were female specific, and 70 were male specific.

In terms of female-specific risk factors, mental health ($n = 24$) and NMR metabolism ($n = 42$) exhibited higher numbers of associations. Notably, the most significant female-specific factors were trauma-related factors in the mental health category, including whether an individual has been sexually assaulted, physically abused by a partner or ex-partner, belittled by a partner or ex-partner and sexually molested as a child (absolute $\beta > 0.086$, $P_{BOF} = 6.14 \times 10^{-25}$ to 1.66×10^{-17}). On the other hand, male-specific risk factors included 28 proteins, 14 mental health variables, 8 physical measures and 5 grey matter volumes, among others. Specifically, the most significant behaviour-related factors were the number of people living together, alcohol consumption, fresh fruit intake and experiences of physical violence as victims of crime. Additionally, specific proteins ($\beta = -0.037$ to 0.045, $P_{BOF} = 2.49 \times 10^{-5}$ to 0.046), such as FGF21, F9 and ASGR1, showed significant associations with SA-PRS only in men. Further detailed reports of gender-stratified analyses are shown in Supplementary Data Tables 9 and 10.

Discussion

To the best of our knowledge, this represents the most extensive study so far that unveiled both associations and causal relationships between a wide array of behaviour-related and physiological factors and SA in adults. We systematically identified and ranked 246 behaviour-related and 200 physiological factors that displayed Bonferroni-significant associations with genetic predisposition to SA. Among these, behaviour-related measurements, particularly mental health and lifestyle factors, exhibited stronger associations. Additionally, these significant behaviour-related PRS associations were consistently identified in case–control analyses. Importantly, of the risk factors significantly associated with both PRS and SA, we further suggested

that 57 behaviour-related factors and 1 biomarker were causal factors for SA after Bonferroni correction. Our utilization of machine-learning classification models, anchored in behaviour-related factors (but not physiological factors), demonstrated a remarkable capacity to discriminate individuals with a history of SA accurately. This underscores the significance of behaviour-related factors such as suffering mental distress and feelings of worthlessness in the context of suicide risk.

The robust associations observed in behaviour-related factors, particularly those related to mental health and lifestyle measures, align with findings from previous studies, including one that employed machine-learning approaches to predict SA risk in children[27]. They demonstrated that predictive values for social-environmental, cognitive and clinical psychiatric factors were higher than for physiological traits like neuroimaging. We propose two plausible explanations for this observation. First, as highlighted in the biopsychosocial model of suicide[5], physiological measurements such as neuroimaging variables, as proximal factors, were more sensitive to whether suicidal behaviours are current or more distal[28]. In contrast, the assessment of SA in the UK Biobank relied on prior attempt history, thus placing greater emphasis on distal behaviour-related risk factors, such as lifestyle, early-life experiences and family history[29]. Second, previous studies have consistently highlighted a high prevalence of psychiatric disorders among individuals with a history of SA[30]. This, coupled with the shared genetic aetiology between SA and psychiatric disorders[16], underscores the salience of mental health as a key risk category.

Genetic epidemiology studies have consistently shown that major depressive disorder exhibits the largest genetic overlap with SA among all psychiatric disorders[15], and depressive symptom was the most significant predictor for SA[31]. Consistent with these findings, our results evidenced a potential causal effect of liability to depression (including depression diagnosed by a professional and sum of depression symptoms) on SA. This underscores the importance of addressing depression as a important risk factor for SA. In addition to depression, neuroticism has consistently emerged as a critical factor associated with SA, in both our study and previous research[31]. We further support that neuroticism was one of the most robust risk factors contributing to SA[32], which was potentially mediated by physiological measures including elevated blood cells, biochemistries and proteins. Interestingly, our machine-learning classification models emphasized that suffering from mental distress stands out as the most pivotal predictor for identifying individuals with a history of SA. This observation aligns with the hypothesis put forward by Shneidman (1993)[33] that psychological pain is a particularly important trigger for suicide behaviours.

While brain traits may not have been considered promising in predicting SA[29,34,35], our study uncovered lower grey volumes in several key brain regions related to higher SA-PRS, such as medial and lateral OFC, insula, amygdala and vmPFC. Notably, these findings are largely consistent with a recent review paper that highlighted the neurobiological substrates of SA, emphasizing a dysfunctional emotion network[36]. Among these brain regions, the insula, in particular, exhibited the most significant effects, which was consistent with previous findings reporting smaller insula volume in adult suicide attempters across various mental disorders[37–39]. Additionally, previous studies also provided evidence that SA individuals with bipolar disorders demonstrated lower grey matter volume in the lateral OFC[40]. SA individuals with major depressive disorder exhibited weaker activation in the lateral OFC during decision-making tasks[41] but greater activation in response to angry versus neutral faces[42], suggesting that the involvements of the insula and lateral OFC in SA may play dissociable roles in distinct behavioural impairments such as emotion and decision-making.

Our results also highlighted the potential important role of immune–inflammatory pathways in SA. For blood and metabolic biomarkers, increased white blood cell and lymphocyte counts, as well as immunometabolic factors such as C-reactive protein, emerged as robust associations with SA-PRS. Additionally, BTN3A2 and CXCL17, the proteins in the inflammation panel, were identified as the most significant protein associated with SA-PRS. Importantly, the white blood cell (leukocyte) count and BTN3A2 were important contributors for SA risk. These findings align with prior post-mortem investigations on depressed individuals who died by suicide, which revealed disruptions in glucocorticoid and inflammatory responses in the brain[43]. The convergence of evidence points to the important role of inflammatory profiles in contributing to SA, which could potentially be linked to a dysregulated stress response system and their relevant downstream effects[12].

While our study has many strengths, such as benefitting from a large population-based sample, comprehensively recorded factors, investigations of causal relationships and utilization of various data-driven approaches, there are several limitations to consider. First, the information available from the UK Biobank did not provide specific dates for the first SA. To capture the temporal dynamics and trajectory of differences before and after the onset of suicidal behaviours, longitudinal data and a high-risk participant group would be necessary. Second, participants included in this study are primarily middle- to late-aged British individuals of European ancestry. Factors such as ageing, the long-term consequences of early developmental deficits, and cultural influences may impact the variability in phenotypes within this specific age range, ancestry and region. Furthermore, it is crucial to acknowledge the inherent limitations associated with the UK Biobank dataset, particularly the presence of a 'healthy volunteer' bias[44]. This bias arises due to the voluntary nature of participation, potentially resulting in an overrepresentation of individuals who exhibit better health and are more inclined to participate in research endeavours. Consequently, the generalizability of our findings to broader populations may be constrained, and it would be of interest to address this in the future.

Methods

Study population

The UK Biobank is a population-based cohort comprising over 500,000 participants in the United Kingdom aged between 37 and 73 years and recruited between 2006 and 2010 (http://www.ukbiobank.ac.uk). SA were evaluated in two steps: first, participants were queried with a self-report question (Data-Field ID 20480, 'Have you deliberately harmed yourself, whether or not you meant to end your life?') to identify individuals who engaged in self-harm (n = 6,861) and those who did not (n = 149,976). For individuals reporting self-harm, a subsequent question (Data-Field ID 20483, 'Have you harmed yourself with the intention to end your life?') was asked to distinguish between SA (n = 3,558) and non-suicidal self-injury (n = 3,303). Individuals who responded with 'Yes' to both questions were categorized as SA cases (n = 3,558), while those who responded with 'No' to the first question were identified as controls (n = 149,976). Of the 3,558 participants who attempted suicide, an additional question (Data-Field ID 20482) indicated that 921 had made a cumulative total of three or more attempts, 673 had made two attempts, 1,924 had made one attempt and 40 had made at least one attempt but preferred not to disclose the specific number. 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.

PRS for SA

Genotype data were available for 502,493 participants in the UK Biobank v3 imputation. Detailed genotyping and quality control procedures performed by UK Biobank have been described in a previous publication[45]. Our study excluded SNPs with call rates <95%, minor allele frequency <0.5% and deviation from the Hardy–Weinberg equilibrium with $P < 1 \times 10^{-6}$. Participants with less than 5% missing rates, not outliers in heterozygosity, who had no sex chromosome aneuploidy, of British ancestry, and who had no more than ten putative third-degree relatives in the kinship table were selected. After the quality control, 9,910,057 SNPs and 337,138 participants remained for analysis.

PRS for SA were calculated on the basis of the summary statistics from a European ancestry SA GWAS meta-analysis involving 14 cohorts from a previous study[46]. Note that the initial publication included 15 cohorts for calculating European ancestry GWAS meta-analysis, representing the largest study on SA so far. To avoid sample overlap, we excluded data from the UK Biobank and generated a new GWAS meta-analysis with the remaining 14 cohorts, involving 33,353 SA cases and 444,626 controls. For further details of this GWAS summary statistics, refer to the study conducted by Docherty et al.[46] and Supplementary Table 1. PRS was calculated using PRS-CS[47], a method employing a Bayesian regression framework to apply continuous shrinkage priors on the effect sizes of SNPs in the PRS. These priors are adaptive to both the strength of their association signal in the discovery GWAS and the linkage disequilibrium (LD) structure from an external reference panel. Given that both the discovery GWAS and target datasets in our study are of European ancestry, LD between SNPs was estimated using the 1000 Genomes European reference panels[48]. We performed PRS-CS based on the PRS-CSx tool (https://github.com/getian107/PRScsx), using the auto models that learn the phi parameter from the discovery GWAS, as well as the default settings for other parameters. PLINK 2.0 (https://www.cog-genomics.org/plink/2.0/) was used to weigh posterior effect sizes of SNPs and to aggregate all SNPs into PRS for each individual in the UK Biobank sample. The PRS for suicide attempters was significantly higher than for controls. Note that, to minimize sample overlap between the PheWAS and the case–control analyses, individuals with SA (n = 2,432) were excluded from the subsequent PheWAS calculation, resulting in a total of 334,706 samples in identifying risk factors associated with genetic predisposition to SA. In addition to the PRS-CS, we also constructed the PRS using the clumping and thresholding (PRS-CT) approach. The PheWAS results derived from the PRS-CT approach were generally consistent with those resulting from the PRS-CS approach. For further details on PheWAS results from the PRS-CT approach, see Supplementary Information (p. 11) and Supplementary Data Table 13.

Phenotypes

Behaviour-related phenotypes. Behaviour-related phenotypes consist of six broad categories (Table 1), containing 370 variables in total. (1) Sociodemographic factors included items on education, employment and household information. (2) Lifestyle factors consisted of items on physical activity, sleep, smoking, alcohol consumption, diet, electronic device use, sun exposure and sexual factors. (3) Early-life and family-history factors contained early-life measures such as adoption and maternal smoking, and family-history variables such as parents' age and illnesses. (4) Mental health included self-reported symptoms of major psychiatric conditions conducted both at the assessment centres and online. We additionally summarized seven mental health symptoms, including anxiety, depressive symptoms, mania, mental distress, psychotic experience, trauma and wellbeing. Quantitative measures of these mental health symptoms were obtained by calculating an average score of the items used to assess each mental health symptom. (5) Physical measures consisted of measures such as blood pressure, arterial stiffness and hand-grip strength, as well as self-declared medical conditions such as cancers, operations, recent pains, heart and artery diseases and other major illnesses. Note that these measures were included here for convenience, as some may relate to behaviour, and are at a whole-person level rather than the physiological measures included elsewhere. (6) Cognitive functions contained a series of cognitive tests conducted at the assessment centres, including fluid intelligence, matrix pattern completion, paired associate learning, pairs matching, numeric memory, prospective memory, reaction time, symbol digit substitution, tower rearranging and trail making.

Note that all behaviour-related variables included were reorganized to a small extent on the basis of the framework of the UK Biobank showcase and the summarized data provided in a previous study[22]. Except for some mental health variables derived online, all behaviour-related phenotypes were obtained at the assessment centre concurrently with collecting physiological assessments. For more details on behaviour-related and subsequent physiological phenotypes, see Supplementary Information (pp. 3–4) and Supplementary Data Table 1.

Physiological phenotypes. Physiological phenotypes also contain six categories in three types, involving a total of 1,921 variables. First, neuroimaging phenotypes primarily involved two categories, including (1) regional grey matter volumes, containing 94 brain regions based on the AAL2 atlas[49], which were calculated in our laboratory on the basis of quality-controlled structural magnetic resonance imaging data, and (2) white matter microstructure, including FA, mean diffusivity, intra-cellular volume fraction, OD and isotropic volume fraction measures, which were available for 27 major tracts mapped across the brain provided by UK Biobank. Second, blood and metabolic phenotypes comprised three categories: (3) blood cells, including 31 measurements that were categorized into white blood cells, red blood cells and platelets; (4) blood biochemistry, including 30 biomarkers for liver function, renal function, endocrine, immunometabolism, and bone and joint; and (5) NMR metabolomics, which contains 168 directly measured metabolic biomarkers categorized into 15 systems. Moreover, (6) proteomic phenotypes considered a total of 1,463 unique proteins, which were distributed across four protein panels, including 369 cardiometabolic, 368 inflammation, 367 neurology and 368 oncology, with three proteins included in all four protein panels. More details of proteomic data can be observed in Sun et al.[50].

Statistical analysis

PheWAS. The PHESANT package in R was used to test the PheWAS associations. In this study, 2,291 risk factors involving 2,576 associations (including multicategorical variables) were tested. Sex, age, the first ten genetic principal components, genotyping array and assessment centres were covariates for all association tests. Scanner positions on the x, y and z axes were included as additional covariates for neuroimaging phenotypes to control for static-field heterogeneity, and total intracranial volume (TIV) was included as a covariate for regional grey matter volumes. To directly compare the results between linear and logistic regression models, standardized regression coefficients (β) were estimated as effect sizes. Two-sided statistical tests were applied in all analyses. All 2,576 tested associations were considered together and corrected using Bonferroni correction for multiple comparisons (α = 0.05). Gender-stratified analyses were also conducted to explore sex-specific risk factors using the same procedure.

Logistic regression analysis. Logistic regression analyses were performed using the R-implemented 'glm' function on all phenotypes (2,291 risk factors involving 2,360 associations) for 3,558 SA cases and 149,976 controls to quantify significant associations within the case–control framework. All analyses were adjusted for sex, age and assessment centres, with TIV as an additional covariate for regional grey matter volumes and scanner positions as additional covariates for neuroimaging phenotypes. Bonferroni correction was applied across all 2,360 associations for multiple comparisons (α = 0.05).

MR. The TwosampleMR package in R was used to perform bidirectional two-sample MR analyses on significant phenotypes observed in PheWAS and case–control analyses and SA. GWAS summary statistics for SA were the one used to generate the PRS, as described earlier. GWAS summary data for behaviour-related and blood and metabolic variables were acquired from the Medical Research Council Integrative Epidemiology Unit OpenGWAS database (https://gwas.mrcieu.ac.uk/). GWAS summary data for protein variables were derived from the UK

Biobank Pharma Proteomics Project (UKB–PPP)[51]. For neuroimaging, proteomic and summarized seven mental health variables, GWAS analyses were performed in the UK Biobank samples used for PheWAS. The GWAS was conducted using linear regression in PLINK 2.0 (https://www.cog-genomics.org/plink/2.0/) for each phenotype. All exclusion criteria, genetic data quality check, ancestry control and relatedness removal remain the same as in the generation of the PRS. Sex, age, the first ten genetic principal components and assessment centres were covariates for all variables, and TIV was included as an additional covariate for regional grey matter volumes. All continuous variables were scaled to a mean of 0 and a s.d. of 1 to obtain standardized estimates.

To test for causal effects of phenotypes on SA, we chose genetic instruments at a P threshold of 5×10^{-6}. Significant SNPs of each exposure were clumped with a distance of 1,000 kb and a maximum LD r^2 of 0.01 based on European ancestry reference data from the 1000 Genomes Project[48]. SNP effect data on both exposure and outcome were then harmonized to match the effect alleles before conducting the MR analyses. The threshold of 5×10^{-6} was chosen to select genetic instruments to ensure that the number of genome-wide significant hits for the examined phenotype GWAS was not less than 3 before harmonizing the two GWAS summary statistics, which was determined on the basis of the requirements of MR methods (that is, MR–Egger regression), similar to previous studies[22,52,53]. For the causal effects of SA on phenotypes, genetic instruments were selected from the GWAS summary statistics of SA at a P threshold of 5×10^{-6}. These SNPs underwent clumping with the same parameters, resulting in 71 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 harmonizing the two GWAS summary statistics. To draw reliable conclusions, we performed additional analyses to select genetic instruments with a P threshold of 5×10^{-7}. If the number of suitable genetic instruments was less than 3 under this threshold, we excluded the phenotype from MR analyses. Note that our additional analyses results also showed that the available results obtained using the P threshold of 5×10^{-7} were generally consistent with those using the 5×10^{-6} (see Supplementary Data Tables 11 and 12 for more details).

We primarily utilized the IVW method for MR analysis, given its capability to combine the ratio estimates from individual genetic variants using inverse variance weights, representing a powerful method that has been widely utilized in similar investigations[52,54]. Sensitivity analyses were conducted to assess potential horizontal pleiotropy by estimating the MR–Egger intercept, and to evaluate the global heterogeneity of the genetic instruments using Cochran's Q test. We also employed a (multiplicative) random-effects model to validate factors showing potential genetic instrument heterogeneity. Additionally, to ensure the robustness of our findings, we performed MR analyses using other methods, including weighted median[55] and MR–Egger regression[56].

Our analysis examined a total of 198 factors for their potential causal effects on SA, as well as the causal effects of SA on 190 factors (excluding early-life and family-history factors). We employed Bonferroni correction for multiple comparisons across all examined factors ($\alpha = 0.05$). Furthermore, to broaden our investigation, following the same procedure, we also conducted MR analyses of 273 risk factors that were significant in either PheWAS or case–control analyses on SA, as well as SA on 271 additional factors. This study adheres to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guideline, and a checklist is provided in Supplementary Table 3.

Machine-learning models. We implemented LightGBM[57] to construct a model that performs the classification task to determine whether a participant falls into class 0 (classified as non-suicide attempters) or class 1 (classified as suicide attempters). Briefly, predictors for the model (that is, behaviour-related variables and demographic information such as age and sex) were first determined by variable importance ranking and sequential forward selection. The proposed model was then developed by the ranked predictors based on SA cases ($n = 3{,}558$) and controls ($n = 149{,}976$). The model was trained and evaluated through a tenfold cross-validation. The performance was assessed using discrimination evaluated through the AUC. We built models for behaviour-related, NMR metabolic and proteomic phenotypes, respectively. We utilized Shapley additive explanations (SHAP)[58] plots to visualize the extent to which each predictor contributed to the target variable. Additional details can be found in Supplementary Information (p. 5).

Mediation analysis. Mediation analysis was performed using the lavaan package in R. Four mediation models were conducted. First, the serial mediation model was used to investigate whether the effect of behaviour-related phenotypes on SA was mediated by brain structures and molecular biomarkers. Two paths, including (1) behaviour-related phenotypes → brain structure → molecular biomarkers → SA and (2) behaviour-related phenotypes → molecular biomarkers → brain structure → SA were conducted. In addition, models as (3) behaviour-related phenotypes → molecular biomarkers → SA and (4) behaviour-related phenotypes → brain structure → SA were conducted to determine whether the effect could be mediated individually by a category. The four mediation models were applied to each behaviour-related factor that showed a causal effect on SA survived Bonferroni correction. Brain structure represented the mean values of grey matter volumes significantly associated with SA-PRS survived Bonferroni correction in PheWAS. Molecular biomarkers were the mean values of blood cells (or blood biochemistry, or NMR metabolites, or proteins) significantly associated with SA-PRS survived Bonferroni correction in PheWAS. Sex, age and assessment centres were covariates for all models. TIV was an additional covariate for models containing brain structure. We used a bootstrapping approach to establish generalizability across sites, address the unequal sample size of the groups, and correct for multiple comparisons. Total, direct and indirect associations were estimated by the 10,000-iteration non-parametric bootstrap approach.

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 the present study are available from UK Biobank with restrictions applied. Data were used under licence 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 UK Biobank-based summary statistics for the GWAS of behavioural-related risk factors can be obtained from the MRC IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/). GWAS summary data for protein variables can be downloaded from the UK Biobank Pharma Proteomics Project (https://www.synapse.org/#!Synapse:syn51364943/). GWAS summary data for SA can be applied via the PGC SUI Data Access Portal (https://pgc.unc.edu/for-researchers/data-access-committee/data-access-portal/). European ancestry reference data from the 1000 Genomes Project can be found via https://github.com/getian107/PRScsx?tab=readme-ov-file.

Code availability

For the analyses conducted in R (version 4.2.3), the PHESANT package (v1.1) was used to perform PheWAS, TwoSampleMR (v0.5.6) to perform MR analysis, base 'glm' function to perform logistic regression analysis, and lavaan (v0.6-16) to perform mediation analysis. PLINK 2.0 were used to calculate PRS and perform GWAS analysis. PRS-CSx tool (v1.1.0) based on Python 3.9 was used to estimate PRS score using the PRS-CS

method. LightGBM library (v3.3.2) based on Python 3.9 was used to develop the machine learning models. The primary code used in this study has been made publicly accessible through the GitHub repository (https://github.com/beimagic/Suicide_Risk_factors).

References

1. *Preventing Suicide: A Global Imperative* (World Health Organization, 2014).
2. Mars, B. et al. Predictors of future suicide attempt among adolescents with suicidal thoughts or non-suicidal self-harm: a population-based birth cohort study. *Lancet Psychiatry* **6**, 327–337 (2019).
3. Janiri, D. et al. Risk and protective factors for childhood suicidality: a US population-based study. *Lancet Psychiatry* **7**, 317–326 (2020).
4. Franklin, J. C. et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychol. Bull.* **143**, 187–232 (2017).
5. Turecki, G. et al. Suicide and suicide risk. *Nat. Rev. Dis. Primers* **5**, 74 (2019).
6. Brezo, J. et al. Predicting suicide attempts in young adults with histories of childhood abuse. *Br. J. Psychiatry* **193**, 134–139 (2008).
7. Millner, A. J., Robinaugh, D. J. & Nock, M. K. Advancing the understanding of suicide: the need for formal theory and rigorous descriptive research. *Trends Cogn. Sci.* **24**, 704–716 (2020).
8. Joo, Y. Y. et al. Association of genome-wide polygenic scores for multiple psychiatric and common traits in preadolescent youths at risk of suicide. *JAMA Netw. Open* **5**, e2148585–e2148585 (2022).
9. DeVille, D. C. et al. Prevalence and family-related factors associated with suicidal ideation, suicide attempts, and self-injury in children aged 9 to 10 years. *JAMA Netw. Open* **3**, e1920956–e1920956 (2020).
10. Strawbridge, R. J. et al. Identification of novel genome-wide associations for suicidality in UK Biobank, genetic correlation with psychiatric disorders and polygenic association with completed suicide. *EBioMedicine* **41**, 517–525 (2019).
11. Ioannidis, J. P. Neglecting major health problems and broadcasting minor, uncertain issues in lifestyle science. *JAMA* **322**, 2069–2070 (2019).
12. Oquendo, M. A. et al. Toward a biosignature for suicide. *Am. J. Psychiatry* **171**, 1259–1277 (2014).
13. Mann, J. J. & Rizk, M. M. A brain-centric model of suicidal behavior. *Am. J. Psychiatry* **177**, 902–916 (2020).
14. Turecki, G. & Brent, D. A. Suicide and suicidal behaviour. *Lancet* **387**, 1227–1239 (2016).
15. Mullins, N. et al. GWAS of suicide attempt in psychiatric disorders and association with major depression polygenic risk scores. *Am. J. Psychiatry* **176**, 651–660 (2019).
16. Ruderfer, D. M. et al. Significant shared heritability underlies suicide attempt and clinically predicted probability of attempting suicide. *Mol. Psychiatry* **25**, 2422–2430 (2020).
17. Goldman, D. Predicting suicide. *Am. J. Psychiatry* **177**, 881–883 (2020).
18. Orri, M. et al. A genetically informed study on the association of cannabis, alcohol, and tobacco smoking with suicide attempt. *Mol. Psychiatry* **26**, 5061–5070 (2021).
19. Richardson, T. G., Harrison, S., Hemani, G. & Davey Smith, G. An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. *eLife* **8**, e43657 (2019).
20. Burgess, S. et al. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* **30**, 543–552 (2015).
21. Lim, K. X. et al. Studying individual risk factors for self-harm in the UK Biobank: a polygenic scoring and Mendelian randomisation study. *PLoS Med.* **17**, e1003137 (2020).
22. Shen, X. et al. A phenome-wide association and Mendelian randomisation study of polygenic risk for depression in UK Biobank. *Nat. Commun.* **11**, 2301 (2020).
23. Chen, S. D. et al. A phenome-wide association and mendelian randomization study for Alzheimer's disease: a prospective cohort study of 502,493 participants from the UK biobank. *Biol. Psychiatry* **93**, 790–801 (2023).
24. Bush, W. S., Oetjens, M. T. & Crawford, D. C. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat. Rev. Genet.* **17**, 129–145 (2016).
25. Hebbring, S. J. The challenges, advantages and future of phenome-wide association studies. *Immunology* **141**, 157–165 (2014).
26. Denny, J. C., Bastarache, L. & Roden, D. M. Phenome-wide association studies as a tool to advance precision medicine. *Annu. Rev. Genomics Hum. Genet.* **17**, 353–373 (2016).
27. Van Velzen, L. S. et al. Classification of suicidal thoughts and behaviour in children: results from penalised logistic regression analyses in the Adolescent Brain Cognitive Development study. *Br. J. Psychiatry* **220**, 210–218 (2022).
28. Kim, D. J. et al. Examination of structural brain changes in recent suicidal behavior. *Psychiatry Res. Neuroimaging* **307**, 111216 (2021).
29. Auerbach, R. P., Chase, H. W. & Brent, D. A. The elusive phenotype of preadolescent suicidal thoughts and behaviors: can neuroimaging deliver on its promise? *Am. J. Psychiatry* **178**, 285–287 (2021).
30. Beautrais, A. L. et al. Prevalence and comorbidity of mental disorders in persons making serious suicide attempts: a case-control study. *Am. J. Psychiatry* **153**, 1009–1014 (1996).
31. Melhem, N. M. et al. Severity and variability of depression symptoms predicting suicide attempt in high-risk individuals. *JAMA Psychiatry* **76**, 603–613 (2019).
32. Brezo, J., Paris, J. & Turecki, G. Personality traits as correlates of suicidal ideation, suicide attempts, and suicide completions: a systematic review. *Acta Psychiatr. Scand.* **113**, 180–206 (2006).
33. Shneidman, E. S. *Suicide as Psychache: A Clinical Approach to Self-Destructive Behavior* (Jason Aronson, 1993).
34. Vidal-Ribas, P. et al. Multimodal neuroimaging of suicidal thoughts and behaviors in a US population-based sample of school-age children. *Am. J. Psychiatry* **178**, 321–332 (2021).
35. van Velzen, L. S. et al. Structural brain alterations associated with suicidal thoughts and behaviors in young people: results from 21 international studies from the ENIGMA Suicidal Thoughts and Behaviours consortium. *Mol. Psychiatry* **27**, 4550–4560 (2022).
36. Schmaal, L. et al. Imaging suicidal thoughts and behaviors: a comprehensive review of 2 decades of neuroimaging studies. *Mol. Psychiatry* **25**, 408–427 (2020).
37. Soloff, P. H. et al. Structural brain abnormalities and suicidal behavior in borderline personality disorder. *J. Psychiatr. Res.* **46**, 516–525 (2012).
38. Giakoumatos, C. I. et al. Are structural brain abnormalities associated with suicidal behavior in patients with psychotic disorders? *J. Psychiatr. Res.* **47**, 1389–1395 (2013).
39. Hwang, J.-P. et al. Cortical and subcortical abnormalities in late-onset depression with history of suicide attempts investigated with MRI and voxel-based morphometry. *J. Geriatr. Psychiatry Neurol.* **23**, 171–184 (2010).
40. Johnston, J. A. et al. Multimodal neuroimaging of frontolimbic structure and function associated with suicide attempts in adolescents and young adults with bipolar disorder. *Am. J. Psychiatry* **174**, 667–675 (2017).
41. Jollant, F. et al. Decreased activation of lateral orbitofrontal cortex during risky choices under uncertainty is associated with disadvantageous decision-making and suicidal behavior. *NeuroImage* **51**, 1275–1281 (2010).

42. Jollant, F. et al. Orbitofrontal cortex response to angry faces in men with histories of suicide attempts. *Am. J. Psychiatry* **165**, 740–748 (2008).
43. Sudol, K. & Mann, J. J. Biomarkers of suicide attempt behavior: towards a biological model of risk. *Curr. Psychiatry Rep.* **19**, 1–13 (2017).
44. 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).
45. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
46. 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).
47. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
48. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
49. Rolls, E. T., Joliot, M. & Tzourio-Mazoyer, N. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage* **122**, 1–5 (2015).
50. Sun, B. B. et al. Genetic regulation of the human plasma proteome in 54,306 UK Biobank participants. Preprint at *bioRxiv* https://doi.org/10.1101/2022.06.17.496443 (2022).
51. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
52. Zhang, Y. et al. Identifying modifiable factors and their joint effect on dementia risk in the UK Biobank. *Nat. Hum. Behav.* **7**, 1185–1195 (2023).
53. Choi, K. W. et al. An exposure-wide and Mendelian randomization approach to identifying modifiable factors for the prevention of depression. *Am. J. Psychiatry* **177**, 944–954 (2020).
54. Nassan, M. et al. Genetic evidence for a potential causal relationship between insomnia symptoms and suicidal behavior: a Mendelian randomization study. *Neuropsychopharmacology* **47**, 1672–1679 (2022).
55. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
56. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
57. Ke, G. et al. Lightgbm: a highly efficient gradient boosting decision tree. *Adv. Neural Inform. Process. Syst.* **30**, 3149–3157 (2017).
58. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inform. Process. Syst.* **30**, 4768–4777 (2017).

Acknowledgements

This study used the UK Biobank Resource under application number 19542. We thank all participants and researchers from the UK Biobank. We are particularly grateful to N. Mullins, who helped conduct the new GWAS meta-analysis of SA. This work was partly supported by a grant from the National Key Research and Development Program of China (no. 2019YFA0709502 to J.F.), a grant from Shanghai Municipal Science and Technology Major Project (no. 2018SHZDZX01 to J.F.), a grant from ZJ Lab, Shanghai Centre for Brain Science and Brain-Inspired Technology and a grant from the 111 Project (no. B18015 to J.F.). This work was partly supported by the Shanghai Rising-Star Program (no. 21QA1408700 to W.C.). This work was supported by a grant from Science and Technology Innovation 2030-Brain Science and Brain-Inspired Intelligence Project (no. 2022ZD0212800 to Y.J.). This work was partly supported by a grant from China Postdoctoral Science Foundation (no. 2022M710804 to B.Z.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

B.Z., W.C. and J.F. conceived and designed the experiment. B.Z. did the analyses with support from J.Yo., J.K., Y.L., R.Z., W.Z., H.W. and C.S. B.Z. drafted the paper with contributions from W.C., E.T.R. and X.W. and comments from all other authors. Y.J., S.X. and C.X. contributed to the visualization of data. X.W., W.C., J.Yu. and J.F. contributed to the interpretation of results. B.Z., W.C. and J.F. had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. All authors read and approved the final paper.

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/s41562-024-01903-x.

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

Peer review information *Nature Human Behaviour* thanks Sylvane Desrivières, Miguel Renteria and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

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]Institute of Science and Technology for Brain-Inspired Intelligence and Department of Neurology, Huashan Hospital, Fudan University, Shanghai, China. [2]Key Laboratory of Computational Neuroscience and Brain Inspired Intelligence (Fudan University), Ministry of Education, Shanghai, China. [3]Oxford Centre for Computational Neuroscience, Oxford, UK. [4]Department of Computer Science, University of Warwick, Coventry, UK. [5]Medical Psychological Centre, The Second Xiangya Hospital, Central South University, Changsha, China. [6]Medical Psychological Institute, Central South University, Changsha, China. [7]China National Clinical Research Centre on Mental Disorders (Xiangya), Changsha, China. [8]Department of Neurology, Qingdao Municipal Hospital, Qingdao University, Qingdao, China. [9]MOE Frontiers Centre for Brain Science, Fudan University, Shanghai, China. [10]Fudan ISTBI-ZJNU Algorithm Centre for Brain-inspired Intelligence, Zhejiang Normal University, Jinhua, China. ✉e-mail: wcheng@fudan.edu.cn; jianfeng64@gmail.com

nature portfolio

Corresponding author(s): Jianfeng Feng
Wei Cheng

Last updated by author(s): Apr 20, 2024

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	No software was involved in data collection (data used is all directly available from UK Biobank, as described in detail in the paper).
Data analysis	R version 4.2.3 packages: PHESANT package (v1.1) was used to perform the phenome-wide association study (PheWAS); TwoSampleMR (v0.5.6) was used to perform Mendelian Randomization analysis; R-implemented base 'glm' function to perform logistic regression analysis; lavaan (v0.6-16) was used to perform Mediation analysis; PRS-CSx tool (v1.1.0) based on Python 3.9 was used to estimate PRS score; PLINK 2.0 was used to calculate PRS and perform GWAS analysis; Python 3.9 with LightGBM library (v3.3.2) was used to develop the machine learning models; The primary code used in this study has been made publicly accessible through the GitHub repository (https://github.com/beimagic/Suicide_Risk_factors).

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 data used in the present study are available from UK Biobank 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 UKB-based summary statistics for the GWAS of risk factors can be obtained from the MRC IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/). GWAS summary data for protein variables can be downloaded from the UK Biobank Pharma Proteomics Project (https://www.synapse.org/#!Synapse:syn51364943/). GWAS summary data for suicide attempts can be applied via the PGC SUI Data Access Portal (https://pgc.unc.edu/for-researchers/data-access-committee/data-access-portal/). European ancestry reference data from the 1000 Genomes Project can be found via https://github.com/getian107/PRScsx?tab=readme-ov-file.

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 Supplementary Table 2. All statistic models were adjusted for sex.
Reporting on race, ethnicity, or other socially relevant groupings	All 334,706 UK Biobank participants included in the calculation of PRS and PheWAS are white ethnicity group. Among the validation dataset comprising 3,558 suicide attempt cases and 149,976 controls from the UK Biobank, over 95.68% of participants in both groups are white ethnicity group. Therefore, our statistical analyses did not incorporate race as a covariate.
Population characteristics	Of the 334,706 individuals included in the calculation of PRS for suicide attempts and the subsequent PheWAS, 53.59% were female, and the mean (s.d.) age was 56.91 (7.99) years. Among the dataset comprising 3,558 SA cases and 149,976 controls, 56.28% were female, and the mean (s.d.) age was 56.03 (7.72) years. Further demographic characteristics are shown in Supplementary Table 2. All statistical analyses were adjusted for sex and age. For comprehensive statistical insights, Table 1 presents details of the sample sizes used for each risk factor category in the PheWAS and logistic regression analyses.
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 from the UK Biobank. After rigorous quality control of genotype data, 334,706 unrelated individuals of British ancestry were included in the calculation of PRS for suicide attempts and subsequent PheWAS. Additionally, a total of 3,558 suicide attempters and 149,976 non-suicide attempters were included based on self-reported suicide behavior. Table 1 lists the sample sizes used in the PheWAS and logistic regression analyses for each risk factor category.
Data exclusions	For the PRS calculation, we excluded single-nucleotide polymorphisms (SNPs) with call rates < 95%, minor allele frequency < 0.5%, and deviation from Hardy-Weinberg equilibrium with P < 1 x 10-6. Participants with less than 5% missing rates, not outliers in heterozygosity, had no sex chromosome aneuploidy, of British ancestry, and had no more than 10 putative third-degree relatives in the kinship table were selected to calculate the PRS for suicide attempts. Note that, to minimize sample overlap between the PheWAS and the case-control analyses,

	individuals with suicide attempts (n = 2,432) were excluded from the subsequent PheWAS calculation, resulting in a total of 334,706 samples in identifying risk factors associated with genetic predisposition to suicide attempts.
Replication	All available data were used to maximize the statistical power of the analysis; therefore, we did not repeat the analysis.
Randomization	All statistical models were adjusted for age, sex, and assessment centers in the current study. Association analyses involving genetic data were also adjusted for the first ten genetic principal components and genotyping array. Association analyses involving gray matter volumes were also adjusted for total intracranial volume .
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 image 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 imaging
Field strength	3T
Sequence & imaging parameters	The EPI-based acquisitions utilize simultaneous multi-slice (multiband) acceleration. UK Biobank uses pulse sequences and reconstruction code from the Center for Magnetic Resonance Research (CMRR), University of Minnesota https://www.cmrr.umn.edu/multiband. The resolution is 1x1x1 mm and the field of view is 208x256x256 matrix. Straight sagittal orientation is used. TR and TE are 2000ms and 2.01ms respectively. The flip angle is 8 deg. Detailed sequence

and imaging parameters are openly available here: https://biobank.ndph.ox.ac.uk/showcase/showcase/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 T1-weighted structural imaging data were preprocessed with the Statistical Parametric Mapping software version 12 (https://www.fil.ion.ucl.ac.uk/spm/) using the CAT12 toolbox (https://neuro-jena.github.io/cat/) with default settings. The preprocessing involved high-dimensional spatial normalization with an integrated Dartel template in Montreal Neurological Institute (MNI) space, followed by nonlinear modulations and correction for each individual's head size. Following these procedures, gray matter images (voxel size: 1.5×1.5×1.5 mm3) were obtained for all participants. The AAL2 atlas with 94 cortical brain regions was used to extract structural imaging-derived phenotypes referred to as regional gray matter volumes.

The diffusion imaging data is corrected for eddy currents and head motion, and has outlier slices (individual slices in the 4D data) corrected,using the Eddy tool (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/EDDY). GDC is then applied, resulting in the 4D output filed. This is then fed into two complementary analyses, one based on tract-skeleton processing, and the other based on a richer modelling of within-voxel tract structure, followed by probabilistic tractography analysis (BEDPOSTx/PROBTRACKx). Maps for fractional anisotropy (FA) and mean diffusivity (MD) were generated, and FA maps were used to generate tract masks, using probabilistic tractography analysis by AutoPtx package from FSL. Twenty-seven tracts were generated (12 bilateral and 3 unilateral tracts). Weighted mean FA and MD were then calculated for each tract. Neurite Orientation Dispersion and Density Imaging measures were also generated as supplementary measures, which include intra-cellular volume fraction (ICVF), isotropic or free water volume fraction (ISOVF), and orientation dispersion (OD).

Normalization: The T1-weighted images and diffusion images were nonlinearly normalized to MNI152 space by CAT12 and FSL-based warping, respectively.

Normalization template: The T1-weighted images were normalized to the Dartel template in MNI space and the diffusion images were normalized to the FMRIB58_FA MNI template.

Noise and artifact removal: The diffusion data was corrected for eddy currents and head motion, and has outlier-slices (individual slices in the 4D data) correction, using the Eddy tool in FSL.

Volume censoring: No volume censoring was applied for the T1-weighted images and diffusion images.

Statistical modeling & inference

Model type and settings: (1) PheWAS: The PHESANT package in R was used to test the PheWAS associations. In this study, 2,291 risk factors involving 2,576 associations (including multicategorical variables) were tested. Sex, age, the first 10 genetic principal components, genotyping array, assessment centres, scanner positions on the x, y, and z axes were covariates for all neuroimaging phenotypes, and total intracranial volume was included as a additional covariate for regional gray matter volumes.
(2) Logistic regression analysis: Logistic regression analyses were performed using the R-implemented 'glm' function on all neuroimaging phenotypes for SA cases and controls to quantify significant associations within the case-control framework. These analyses were adjusted for sex, age, and assessment centers, with total intracranial volume as an additional covariate for regional gray matter volumes.

Effect(s) tested: Standardized regression coefficients (β) were estimated as effect sizes. Two-sided statistical tests were applied in all analyses.

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

Statistic type for inference (See Eklund et al. 2016): Voxel-wise association

Correction: Bonferroni correction

Models & analysis

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