

# The brain structure, immunometabolic and genetic mechanisms underlying the association between lifestyle and depression

Received: 21 December 2022

Accepted: 4 August 2023

Published online: 11 September 2023

 Check for updates

Yujie Zhao<sup>1,2,3,12</sup>, Liu Yang<sup>1,12</sup>, Barbara J. Sahakian <sup>1,4,5</sup>, Christelle Langley <sup>4,5</sup>, Wei Zhang<sup>1,2,3</sup>, Kevin Kuo<sup>1</sup>, Zeyu Li<sup>1,2,3</sup>, Yihan Gan<sup>1</sup>, Yuzhu Li<sup>1,2,3</sup>, Yang Zhao<sup>6</sup>, Jintai Yu <sup>1</sup>✉, Jianfeng Feng <sup>1,2,3,7,8,9,10</sup>✉ & Wei Cheng <sup>1,2,3,7,11</sup>✉

Lifestyle factors have been acknowledged to be modifiable targets that can be used to counter the increasing prevalence of depression. This study aims to investigate combining an extensive range of lifestyle factors, including alcohol consumption, diet, physical activity, sleep, smoking, sedentary behavior and social connection, that contribute to depression, and examine the underlying neurobiological mechanisms. Over nine years of follow-up, a multivariate Cox model was utilized on 287,282 participants from UK Biobank to demonstrate the protective roles of seven lifestyle factors and combined lifestyle score on depression. Combining genetic risk and lifestyle category in 197,344 participants, we found that a healthy lifestyle decreased the risk of depression across a population with varied genetic risk. Mendelian randomization confirmed the causal relationship between lifestyle and depression. A broad range of brain regions and peripheral biomarkers were related to lifestyle, including the pallidum, the precentral cortex, triglycerides and C-reactive protein. Structural equation modeling on 18,244 participants revealed underlying neurobiological mechanisms involving lifestyle, brain structure, immunometabolic function, genetics and depression. Together, our findings suggest that adherence to a healthy lifestyle could aid in the prevention of depression.

The rising prevalence of major depressive disorder (MDD) is imposing a substantial burden on public health worldwide<sup>1–3</sup>. The factors affecting the onset of depression are complicated and involve both behavioral and biological factors. Therefore, to prevent the onset of depression,

comprehensive strategies are needed to explore the risk factors and mechanisms so as to explain all aspects of the pathophysiology of depression. A recent meta-review assessed the evidence regarding how multiple lifestyle factors impact the risk and treatment outcomes

<sup>1</sup>Institute of Science and Technology for Brain-Inspired Intelligence, and Department of Neurology and Institute of Neurology, Huashan Hospital, Fudan University, Shanghai, China. <sup>2</sup>Key Laboratory of Computational Neuroscience and Brain Inspired Intelligence (Fudan University), Ministry of Education, Shanghai, China. <sup>3</sup>MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. <sup>4</sup>Department of Psychiatry, University of Cambridge, Cambridge, UK. <sup>5</sup>Behavioral and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK. <sup>6</sup>The George Institute for Global Health, University of New South Wales, Sydney, New South Wales, Australia. <sup>7</sup>Fudan ISTBI—ZJNU Algorithm Centre for Brain-inspired Intelligence, Zhejiang Normal University, Jinhua, China. <sup>8</sup>Zhangjiang Fudan International Innovation Center, Shanghai, China. <sup>9</sup>Department of Computer Science, University of Warwick, Coventry, UK. <sup>10</sup>School of Data Science, Fudan University, Shanghai, China. <sup>11</sup>Shanghai Medical College and Zhongshan Hospital Immunotherapy Technology Transfer Center, Shanghai, China. <sup>12</sup>These authors contributed equally: Yujie Zhao, Liu Yang. ✉e-mail: [jintai\\_yu@fudan.edu.cn](mailto:jintai_yu@fudan.edu.cn); [jianfeng64@gmail.com](mailto:jianfeng64@gmail.com); [wcheng@fudan.edu.cn](mailto:wcheng@fudan.edu.cn)

across a range of mental health disorders, including depression, and discussed the potentially shared neurobiological pathways<sup>4</sup>. In contrast to previous studies, which focused on investigating a specific aspect of lifestyle in the prevention of depression and its associated genetic and neurobiological mechanisms<sup>5–10</sup>, we incorporate a wide range of lifestyle factors, including alcohol consumption, smoking habits, physical activity, diet, circadian habits, sedentary behavior and social connection. Research has revealed that, when exercising, muscles secrete myokines, which contribute to the regulation of hippocampal function, improving mood symptoms<sup>11</sup>. It has also been found that overeating and a sedentary lifestyle may increase the risk of depression by suppressing adaptive cellular stress responses<sup>12</sup>. Circadian disruption is associated with an increased risk of depression resulting from insufficient sleep and the disturbance of melatonin<sup>13</sup>. Smoking and alcohol dependence are associated with lesions in brain circuits<sup>14,15</sup>, and reduced social connection is known to impact metabolic and brain health status, which is associated with depression<sup>16</sup>. By integrating a number of lifestyle factors into a comprehensive score and systematically revealing its relationship with incident depression and the underlying mechanisms, we aim to provide an effective strategy for reducing the risk of depression.

Compelling evidence has indicated that the genetic architecture of psychiatric risk is complex and is dominated by multiple contributing factors<sup>17</sup>, so we also aim to investigate the combined effect of stratified polygenic-risk score (PRS) and lifestyle conferring risk for depression. To understand the neurobiological mechanisms by which genetic variation increases risk, the previous literature considered the brain as the intermediate phenotype affected by risk gene variants to further elucidate the mechanistic aspects of brain function implicated in psychiatric disease<sup>17</sup>. Genetic studies have also revealed the contributions of genetic variation in innate immune mechanisms for diseases<sup>18</sup>. Therefore, we further comprehensively investigate the intermediate neurobiological processes influenced by genetic variation underlying the association between lifestyle and depression.

We hypothesized that adherence to a healthy lifestyle would contribute to reducing the risk of depression across a population with different genetic profiles. We also hypothesized that there would be multiple shared neurobiological mechanisms modulated by genetic variants underlying the association of lifestyle with depression. Therefore, our study aims to examine the composite lifestyle risk for depression and its underlying neurobiological mechanisms. We made use of the UK Biobank (UKB), a large prospective cohort, with information on behavioral, brain imaging, biochemistry and genetic measures. The present study has three objectives. The first is to investigate the composite lifestyle impact for depression risk and the combining effect of polygenic and lifestyle conferring risk for depression. Second, we aim to perform a correlation analysis to estimate the relationships of lifestyle with brain structural imaging, blood cells, biochemistry and metabolic markers to unravel how lifestyle factors regulate neurobiological processes. Because brain morphology alteration and disruption of immunometabolic systems have been linked to depression in previous studies, our third aim is to construct a structural equation model (SEM) to elucidate the relationships between lifestyle, PRS, brain structure, immunometabolic function and depression.

## Results

### Population characteristics

We utilized UKB data from a total of 502,409 participants to perform survival analysis and correlation analysis and to construct a structural equation model. For survival analyses we retrieved 287,282 individuals (mean age 57.52 years, 50.70% female) who had participated in the assessment of lifestyle factors at baseline (2006–2010) and had linkage to clinical diagnosis records. The numbers of participants falling into healthy or unhealthy lifestyles for each factor are listed in Table 1. The mean score for the seven lifestyles is 4.75 (s.d. = 1.36). Among the

participants, 1.25% were categorized as following an unfavorable lifestyle (scores in the range 0–1), 38.90% followed an intermediate lifestyle (scores of 2–4), and 59.85% followed a favorable lifestyle (scores of 5–7). Of the 287,282 participants, 12,916 had an onset of depression during a median follow-up of 9.01 years. The demographic characteristics of the participants and the distribution of lifestyle factors are provided in Table 1. The PRSs for depression from 197,344 participants were normally distributed (Supplementary Fig. 5) and categorized into three levels, low (25.09%), intermediate (49.97%) and high (24.94%). Brain structural imaging data were collected in 2014 and contained T1-weighted structural magnetic resonance imaging (MRI) results from 32,839 participants, which were used in the correlation analysis. Blood chemistry and cell count indicators from ~480,000 participants and NMR metabolic biomarkers from ~120,000 participants were collected at baseline. The sample size of each biomarker used in the correlation analyses is presented in Supplementary Tables 18 and 19. We also used data from 448,849 participants who had completed a Patient Health Questionnaire-4 (PHQ-4, collected from 2006–2010) score. Combining the data of lifestyle score, PRS, immunometabolic markers, brain structural imaging and depression, the final analysis included 18,244 participants for estimation of the structural equation model. Extended Data Fig. 1 shows the research guidelines for the study. Supplementary Fig. 1 and Supplementary Table 4 present the sample size utilized in each analysis.

### Survival analyses on the association between lifestyle and depression

Multivariable Cox proportional hazard regression models were used to examine the association between lifestyle and depression. In terms of lifestyle factors, each showed a significant association with depression independently. For each factor, the healthy category showed a lower risk of depression as compared with the unhealthy category. Specifically, the results showed that moderate alcohol consumption decreased the risk of depression by 11% (hazard ratio (HR) of 0.89 (95% CI, 0.85–0.92)), healthy diet decreased the risk of depression by 6% (0.94, 0.90–0.97), regular physical activity decreased the risk of depression by 14% (0.86, 0.83–0.90), never smoking decreased the risk of depression by 20% (0.80, 0.78–0.83), healthy sleep decreased the risk of depression by 22% (0.78, 0.75–0.81), low-to-moderate sedentary behavior decreased the risk of depression by 13% (0.87, 0.84–0.90), and frequent social connection decreased the risk of depression by 18% (0.82, 0.78–0.86). For the lifestyle class, compared with the unfavorable lifestyle, intermediate and favorable lifestyle were associated with 41% (0.59, 0.53–0.65) and 57% (0.43, 0.38–0.47) lower risk of depression, respectively. When all factors were combined, the HRs for depression for participants with a lifestyle score of 1, 2, 3, 4, 5, 6 and 7 compared to those with a score of 0 were 0.83 (95% CI, 0.62–1.11), 0.61 (0.46–0.80), 0.53 (0.40–0.70), 0.45 (0.34–0.59), 0.39 (0.30–0.52), 0.33 (0.25–0.43) and 0.28 (0.21–0.37), respectively (Fig. 1). When treated as quantitative variable, each one-point increment in healthy lifestyle score was associated with an HR of 0.85 (95% CI, 0.84–0.86) (Supplementary Table 5).

### Combining effect of genetic risk and lifestyle on depression

Risk of depression was reduced monotonically across PRS classes, and the HR values for depression in participants with intermediate and low PRS compared to those with high PRS were 0.87 (95% CI, 0.83–0.91) and 0.75 (95% CI, 0.71–0.80) (Supplementary Table 8). We also found a monotonically decreasing trend of depression risk across increasingly favorable lifestyle class when controlling for the genetic risk group (Supplementary Table 10). When combining genetic risk and lifestyle classes, with participants with high genetic risk and unfavorable lifestyle as the reference group, participants with low genetic risk and favorable lifestyle showed the greatest reduction in depression risk, (HR, 0.36 (95% CI, 0.28–0.46); Fig. 2). However, there was no significant interaction between PRS and lifestyle score associated with risk of

**Table 1 | Baseline characteristics of participants**

Characteristics	Number of participants (%) <sup>a</sup>		
	Depression (n=12,916)	No depression (n=274,366)	Total (n=287,282)
Age, mean (s.d.)	56.57 (8.21)	57.57 (7.87)	57.52 (7.89)
Sex			
Female	7,624 (59.03%)	138,023 (50.31%)	145,647 (50.70%)
Male	5,292 (40.97%)	136,343 (49.69%)	141,635 (49.30%)
BMI, mean (s.d.) <sup>b</sup>	28.16 (5.22)	27.43 (4.62)	27.46 (4.65)
SES, mean (s.d.) <sup>c</sup>	-0.79 (3.27)	-1.46 (2.98)	-1.43 (3.00)
Education level <sup>d</sup>			
College degree	3,334 (25.81%)	89,243 (32.53%)	89,243 (32.53%)
A level	1,350 (10.45%)	30,210 (11.01%)	30,210 (11.01%)
O level	2,837 (21.97%)	58,774 (21.42%)	58,774 (21.42%)
CSE	865 (6.70%)	13,970 (5.09%)	13,970 (5.09%)
NVQ	945 (7.32%)	18,995 (6.92%)	18,995 (6.92%)
Professional qualification	698 (5.40%)	14,875 (5.42%)	14,875 (5.42%)
Others	2,767 (21.42%)	46,292 (16.87%)	46,292 (16.87%)
Missing value	120 (0.93%)	2,007 (0.73%)	2,007 (0.73%)
Lifestyle factor			
Alcohol			
Unhealthy alcohol consumption	4,238 (32.81%)	72,657 (26.48%)	76,895 (26.77%)
Moderate alcohol consumption	8,678 (67.19%)	201,709 (73.52%)	210,387 (73.34%)
Diet			
Unhealthy diet	5,334 (41.30%)	105,165 (38.33%)	110,499 (38.46%)
Healthy diet	7,582 (58.70%)	169,201 (61.67%)	176,783 (61.54%)
Physical activity			
Unhealthy physical activity	3,391 (26.25%)	57,588 (20.99%)	60,979 (21.23%)
Regular physical activity	9,525 (73.75%)	216,778 (79.01%)	226,303 (78.77%)
Smoking			
Current and previous smoking	7,018 (54.34%)	130,515 (47.57%)	137,533 (47.87%)
Never smoking	5,898 (45.66%)	143,851 (52.43%)	149,749 (52.13%)
Sleep			
Unhealthy sleep	4,395 (34.03%)	71,603 (26.10%)	75,998 (26.45%)
Healthy sleep	8,521 (65.97%)	202,763 (73.90%)	211,284 (73.55%)
Sedentary behavior			
High sedentary behavior	7,203 (55.77%)	138,889 (50.62%)	146,092 (50.85%)
Low-to-moderate sedentary behavior	5,713 (44.23%)	135,477 (49.38%)	141,190 (49.15%)
Social connection			
Infrequent social connection	2,165 (16.76%)	35,624 (12.98%)	37,789 (13.15%)
Frequent social connection	10,751 (83.24%)	238,742 (87.02%)	249,493 (86.85%)
Lifestyle class			
Unfavorable [0-1]	366 (2.83%)	3,226 (1.18%)	3,592 (1.25%)
Intermediate [2-4]	6,160 (47.69%)	105,595 (38.49%)	111,755 (38.90%)
Favorable [5-7]	6,390 (49.47%)	165,545 (60.34%)	171,935 (59.85%)
Lifestyle score			
0	53 (0.41%)	358 (0.13%)	411 (0.14%)
1	313 (2.42%)	2,868 (1.05%)	3,181 (1.11%)
2	928 (7.19%)	11,918 (4.34%)	12,846 (4.47%)
3	2,087 (16.16%)	32,587 (11.88%)	34,674 (12.07%)

**Table 1 (continued) | Baseline characteristics of participants**

Characteristics	Number of participants (%) <sup>a</sup>		
	Depression	No depression	Total
	(n=12,916)	(n=274,366)	(n=287,282)
4	3,145 (24.35%)	61,090 (22.27%)	64,235 (22.36%)
5	3,416 (26.45%)	78,254 (28.52%)	81,670 (28.43%)
6	2,240 (17.34%)	62,611 (22.82%)	64,851 (22.57%)
7	734 (5.68%)	24,680 (9.00%)	25,414 (8.85%)
Genetic risk category <sup>e</sup>			
Low	1,863 (21.28%)	47,660 (25.27%)	49,523 (25.09%)
Intermediate	4,353 (49.71%)	94,258 (49.98%)	98,611 (49.97%)
High	2,540 (29.01%)	46,670 (24.75%)	49,210 (24.94%)

<sup>a</sup>Percentages may not sum to 100 because of rounding. <sup>b</sup>BMI, body mass index. <sup>c</sup>SES, socioeconomic status (measured by Townsend deprivation index). <sup>d</sup>Classification of education level is defined according to UKB Field ID 6138. Missing values were imputed using the mean value. <sup>e</sup>Genetic risk category is classified based on PRS as high (highest quintile), intermediate (2–4 quintile) and low (lowest quintile).

depression ( $P_{\text{interaction}} = 0.41$ ), indicating that lifestyle may not be substantially altered by the genetic risk for depression, and lifestyle has a strong protective role across different levels of polygenetic risk population. By further performing comparisons among the eight groups except the reference group using permutation tests, we found that participants with high genetic risk but favorable lifestyle had a lower risk of depression than those with intermediate or low genetic risk but unfavorable lifestyle (intermediate genetic risk,  $P = 3.0 \times 10^{-4}$ ; low genetic risk,  $P = 5.8 \times 10^{-3}$ ) and those with intermediate genetic risk and intermediate lifestyle ( $P = 1.1 \times 10^{-2}$ ) (Supplementary Table 9). Participants with intermediate genetic risk and favorable lifestyle also showed lower risk of depression than those with low genetic risk and adherence to unfavorable and intermediate lifestyles (unfavorable,  $P = 2.9 \times 10^{-3}$ ; intermediate,  $P = 2.1 \times 10^{-2}$ ; Supplementary Table 9). The results also elucidated that favorable lifestyle would mitigate the genetic vulnerabilities of depression in population with different levels of polygenetic risk and safeguard the high-risk population against depression.

### Mendelian randomization evidence for the association between lifestyle and depression

Mendelian randomization (MR) analysis showed a protective causal relationship between lifestyle and depression. By utilizing genetic variants as proxies, the results from the inverse variance weighted (IVW) method found that per one-point increment in lifestyle score was associated with 35% reduced risk of depression (odds ratio (OR), 0.65; 95% CI, 0.55–0.76;  $P = 1.0 \times 10^{-7}$ ; Fig. 3). Model-based sensitivity analyses showed that the estimates were similar in size in the simple median method (0.63, 0.50–0.80;  $P = 1.8 \times 10^{-4}$ ), weighted median method (0.65, 0.51–0.83;  $P = 4.7 \times 10^{-4}$ ) and weighted mode method (0.52, 0.31–0.90;  $P = 2.6 \times 10^{-2}$ ), illustrating the robustness of the causal relationship. Scatter plots of lifestyle and depression risk association for the instruments are presented in Fig. 3a, with colored lines representing the slopes of different regression analyses. MR estimates for the effects of single-nucleotide polymorphisms (SNPs) associated with lifestyle on depression risk are presented in a forest plot in Supplementary Fig. 13. Sensitivity analyses also examined the reverse causality and showed significant results (0.85, 0.78–0.92;  $P = 5.9 \times 10^{-5}$ ) with the IVW method (Supplementary Fig. 14), potentially indicating a bidirectional causal relationship between lifestyle and depression.

### Association of lifestyle with brain volume and peripheral markers

The correlations between lifestyle and brain structures are in line with expectations that higher lifestyle scores are associated with larger brain volumes (for example, in cortical structures such as the superior

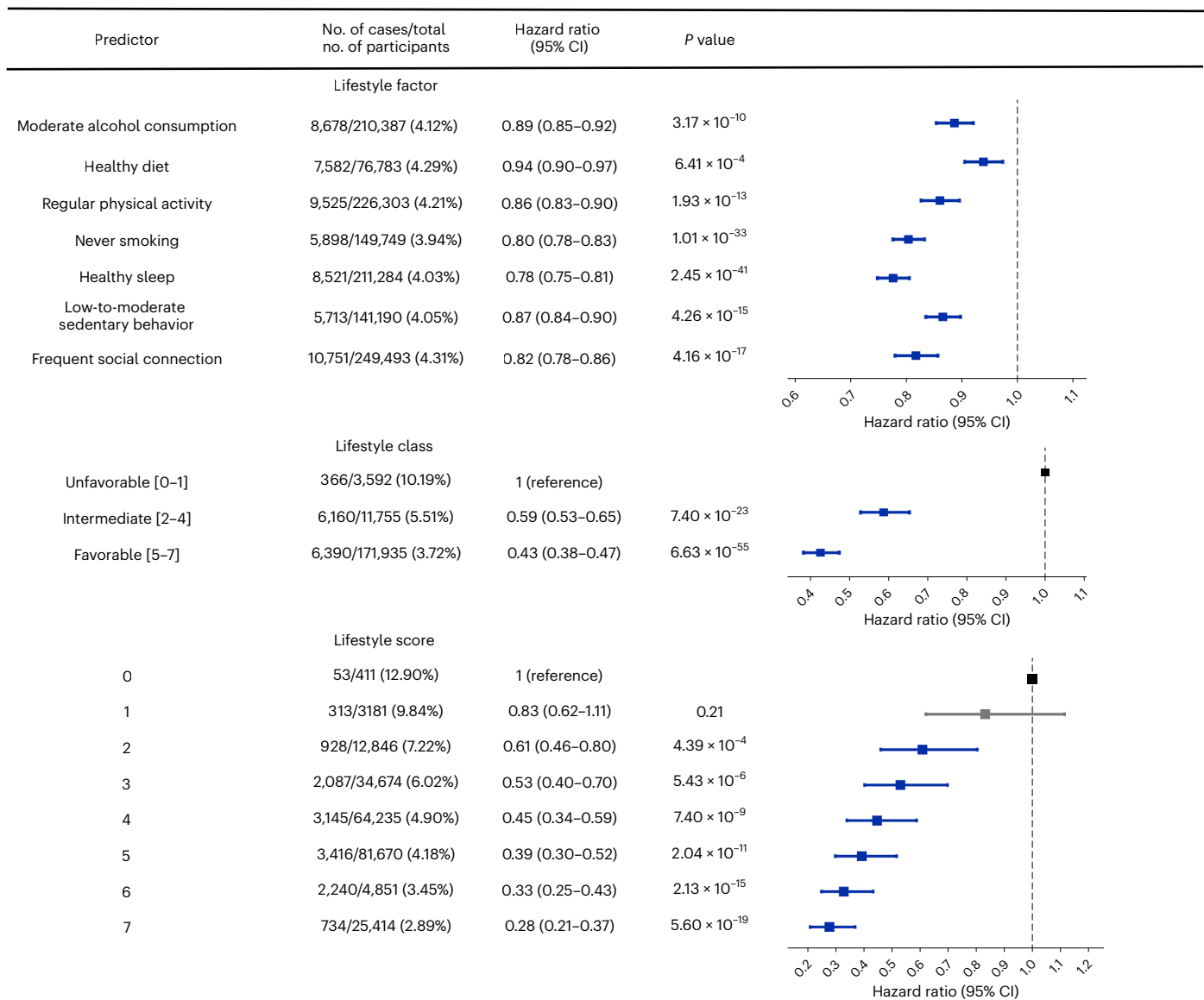
prefrontal cortex, orbitofrontal cortex, precentral cortex and insula, and subcortical structures including the pallidum, thalamus, amygdala and hippocampus; Fig. 4a). The full results for the correlation between each lifestyle factor and brain structure volumes are listed in Supplementary Table 14. We also found that these brain structures are negatively associated with depression symptoms assessed by PHQ-4 (including the precentral, orbitofrontal and middle temporal cortices, the thalamus, hippocampus and putamen; Supplementary Fig. 16 and Supplementary Table 16). We used the lifestyle and depression scores assessed at the neuroimaging visit timepoint (2014+) to calculate the correlation with brain volumes and obtained similar results (Supplementary Figs. 15 and 17 and Supplementary Tables 15 and 17). The spatial correlation between lifestyle and depression-associated brain maps at the neuroimaging visit was  $-0.52$  ( $P = 5.4 \times 10^{-7}$ ; Supplementary Fig. 18), which supports the consistency of the relationship of brain structural volumes and lifestyle and depression.

For the peripheral markers, 48 blood markers and 130 metabolic markers passed the significance threshold after Bonferroni adjustment (Fig. 4b). The  $r$  values for each marker are presented in Fig. 4b. We found that C-reactive protein ( $r = -0.065$ ,  $P < 1.0 \times 10^{-295}$ ) and triglycerides ( $r = -0.075$ ,  $P < 1.0 \times 10^{-295}$ ) were the most significant of the blood biochemistry markers, and neutrophil count ( $r = -0.106$ ,  $P < 1.0 \times 10^{-295}$ ) and leukocyte count ( $r = -0.094$ ,  $P < 1.0 \times 10^{-295}$ ) were the most significant among the blood cells. In terms of metabolic markers, the degree of unsaturation ( $r = 0.153$ ,  $P < 1.0 \times 10^{-295}$ ) and glycoprotein acetyls ( $r = -0.109$ ,  $P = 1.1 \times 10^{-240}$ ) were the most significant markers positively and negatively correlated with lifestyle, respectively. Full results are presented in Supplementary Tables 18 and 19.

### Structure equation model

We employed a structural equation model to analyze the relationship between lifestyle, PRS, brain structure, immunometabolic function and depression in 18,244 participants, further revealing the neurobiological mechanisms underlying the association between lifestyle and depression. First, confirmatory factor analyses were used to examine the latent variables in the structural equation model, including depression, brain structure and immunometabolic function. The left superior frontal cortex in cortical volume and left insula in subcortical volume were the most significant markers to predict the latent variable brain structure (both  $\beta = 0.58$ ;  $P < 0.001$ ). C-reactive protein, triglycerides, glycated hemoglobin (HbA1c) and glucose were the four significant markers predicting the latent variable immunometabolic function ( $\beta = 0.06$ , 0.09, 0.86 and 0.55, respectively;  $P < 0.001$ ). The latent variable depression was represented by the combined four PHQ items ( $\beta = 0.81$ , 0.75, 0.59 and 0.51, respectively;  $P < 0.001$ ). The loading coefficients for





**Fig. 1 | The association of lifestyle factor, class and score with depression risk.**

We utilized participants ( $n = 287,282$ ) from UKB with completed lifestyle factor assessments in the survival analyses. The upper panel presents the association of lifestyle factors with depression. The bottom panel presents the association of a per-one-point increment in healthy lifestyle score with depression. The middle panel presents the association between each class of lifestyle and depression,

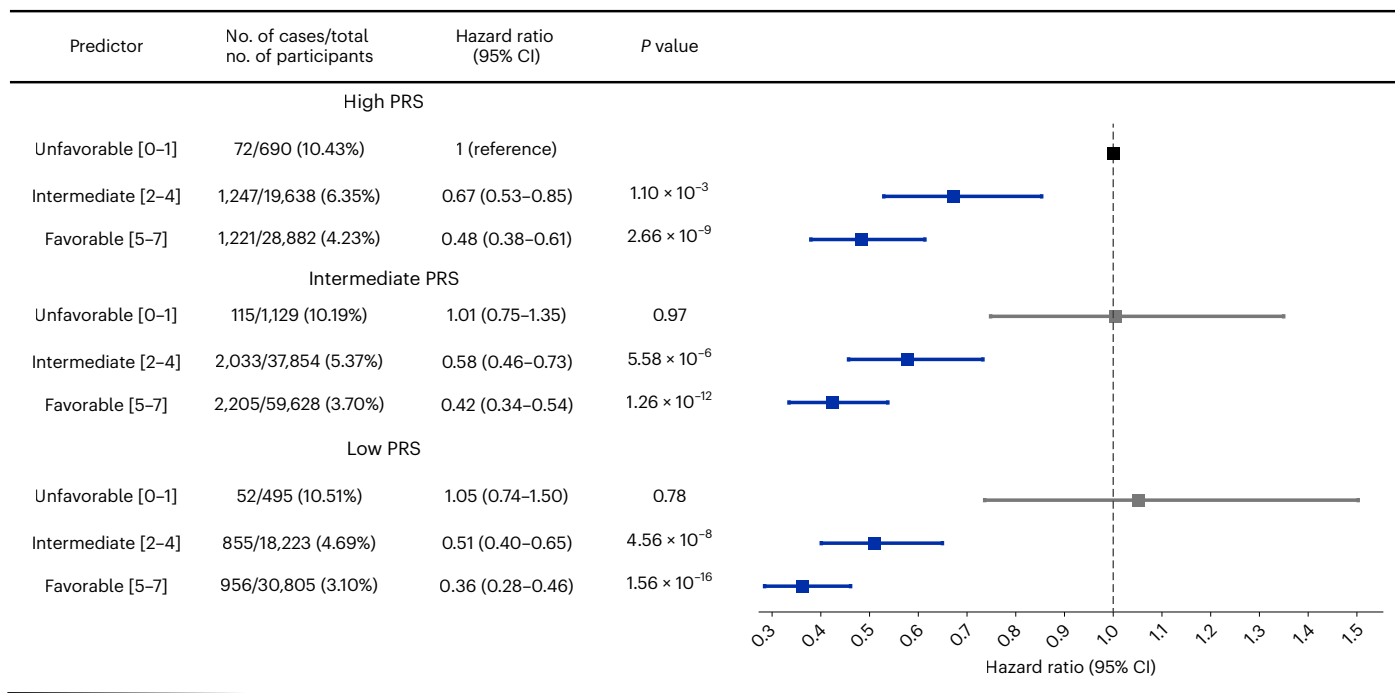
with each lifestyle score reclassified into three classes (favorable scoring 5 to 7, intermediate scoring 2 to 4 and unfavorable scoring 0 to 1). All these models were adjusted for age, sex, Townsend deprivation index, BMI and education level. Two-sided unadjusted association *P* values from multivariate Cox models are given. Data are presented as mean values  $\pm$  s.e.m. The widths of the lines extending from the center points represent 95% CI.

each marker to the corresponding latent variables are shown adjacent to the model panel in Fig. 5.

Next, we input the prepared latent variables into the structural equation model to estimate the path coefficients. We tested all hypothetical paths in the model (Fig. 5). Lifestyle was a significant predictor of depression ( $\beta = -0.157$ ,  $P < 1.0 \times 10^{-20}$ ), immunometabolic function ( $\beta = -0.043$ ,  $P = 4.7 \times 10^{-7}$ ) and brain structure ( $\beta = 0.038$ ,  $P = 2.9 \times 10^{-6}$ ). PRS ( $\beta = 0.036$ ,  $P = 1.5 \times 10^{-5}$ ), brain structure ( $\beta = -0.023$ ,  $P = 1.2 \times 10^{-2}$ ) and immunometabolic function ( $\beta = 0.020$ ,  $P = 3.5 \times 10^{-2}$ ) were also significant predictors of depression. PRS was a significant predictor of lifestyle ( $\beta = -0.022$ ,  $P = 3.5 \times 10^{-3}$ ) and immunometabolic function ( $\beta = 0.018$ ,  $P = 3.4 \times 10^{-2}$ ). All paths represented significant associations and passed the false discovery rate (FDR) correction except for the association between brain structure and PRS, and the association between brain structure and immunometabolic function.

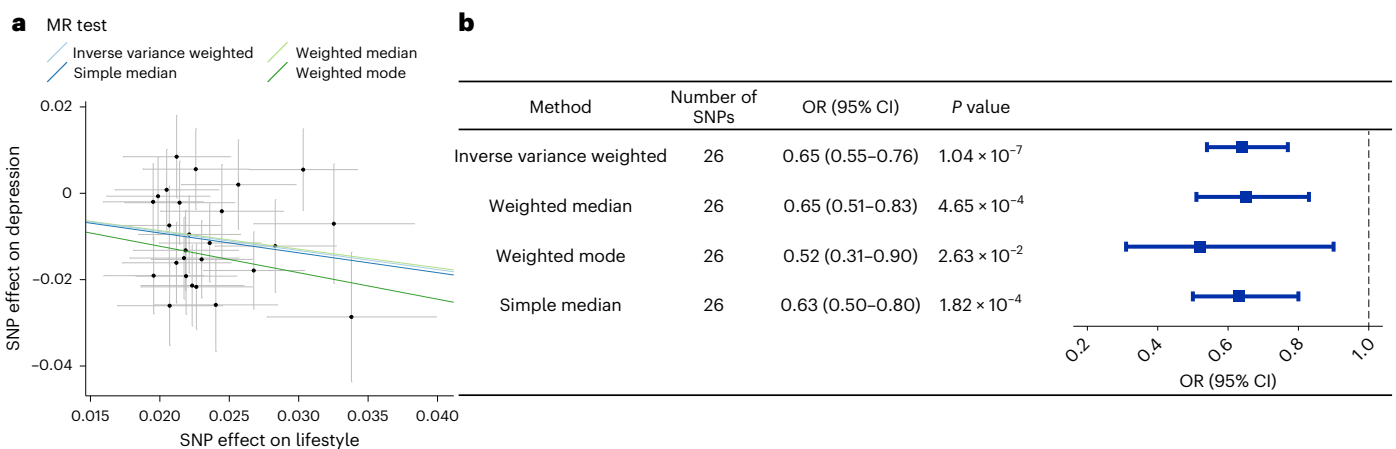
### Sensitivity analyses

The main survival analyses results, including the association of lifestyle factors, classes and scores with depression risk, were also confirmed in the sensitivity analyses across three subtypes of depression, including single depressive episode, recurrent depressive disorder and treatment-resistant depression (TRD). For lifestyle factors, healthy sleep was still the greatest factor in reducing the risk of a single depressive episode (HR, 0.78 (95% CI, 0.75–0.81)) and TRD (HR, 0.64 (95% CI, 0.58–0.71)) (Supplementary Figs. 6 and 8). However, for recurrent depressive disorder, frequent social connection reduced the risk most (HR, 0.61 (95% CI, 0.51–0.72)) (Supplementary Fig. 7). To further explore the risk of depression according to previous and current smoking status, with never smoking as the reference group, we found a significant monotonically increasing risk of depression in the previous–never smoking comparison and current–never smoking



**Fig. 2 | Risk of depression according to genetic risk and lifestyle.** We utilized 197,344 participants with both PRS and lifestyle scores in the analyses. The results were adjusted for age, sex, Townsend deprivation index, BMI and education level. Two-sided unadjusted association *P* values from multivariate Cox models are

given. Individuals with high risk for depression and unfavorable lifestyle were used as the reference category. Data are presented as mean values  $\pm$  s.e.m. The widths of the lines extending from the center points represent 95% CI.



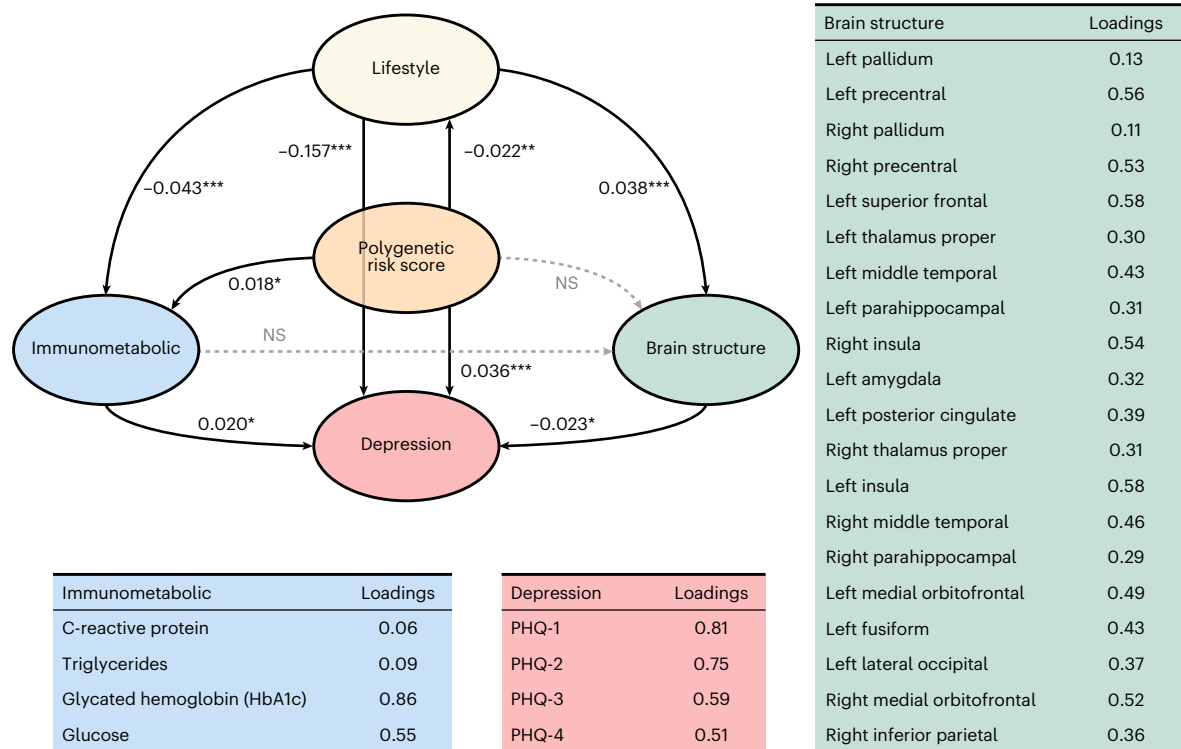
**Fig. 3 | MR plots for the relationship between lifestyle and depression.** **a**, Scatter plot of SNP effects on lifestyle versus depression, with the slope of each line corresponding to the estimated MR effect per method. The data are expressed as raw  $\beta$  values with 95% CI values. **b**, Model-based sensitivity analyses from four methods ( $n = 26$  SNPs). The effect estimates represent the odds ratio

for depression per one-point increment in healthy lifestyle score. Data are presented as mean values  $\pm$  s.e.m. The width of the lines extending from the center point represent the 95% CI. Two-sided unadjusted association *P* values from four models are given.

comparison across the three subtypes (Supplementary Table 6). The consistent monotonical reduction of depression risk across lifestyle classes was also found in each subtype of depression (Supplementary Figs. 6–8). In addition, the favorable lifestyle class with a larger sample size as the reference group also showed similar significant results (Supplementary Table 7). In terms of the combining lifestyle score, each one-point increment in the healthy lifestyle score was associated with an HR of 0.85 (95% CI, 0.84–0.86) for a single depressive episode, 0.82 (95% CI, 0.78–0.87) for recurrent depressive disorder and 0.81 (95% CI, 0.78–0.84) for TRD (Supplementary Table 5).

Additionally, similar patterns of depression risk according to genetic risk category were found in the three subtypes of depression (Supplementary Table 8). For the three subtypes of depression, we also confirmed a monotonical association of decreasing depression risk across increasing lifestyle class when the comparison was restricted to the same genetic risk group (Supplementary Tables 11–13). Similarly, the risk of depression in the three subtypes was not altered by the interaction of genetics and lifestyle, indicating that adherence to a healthy lifestyle would attenuate genetic risk across the different PRS groups.





**Fig. 5 | Structural equation model.** A total of 18,244 participants were utilized in structural equation model analyses. Standardized coefficients are shown. Lifestyle was a significant predictor of depression ( $\beta = -0.157$ ,  $P < 1.0 \times 10^{-20}$ ), immunometabolic function ( $\beta = -0.043$ ,  $P = 4.7 \times 10^{-7}$ ) and brain structure ( $\beta = 0.038$ ,  $P = 2.9 \times 10^{-6}$ ). PRS ( $\beta = 0.036$ ,  $P = 1.5 \times 10^{-3}$ ), brain structure ( $\beta = -0.023$ ,  $P = 1.2 \times 10^{-2}$ ) and immunometabolic function ( $\beta = 0.020$ ,  $P = 3.5 \times 10^{-3}$ ) were also significant predictors of depression. PRS was also a significant predictor of lifestyle ( $\beta = -0.022$ ,  $P = 3.5 \times 10^{-3}$ ) and immunometabolic

function ( $\beta = 0.018$ ,  $P = 3.4 \times 10^{-2}$ ). All paths represent significant associations except for the association between brain structure and PRS, and the association between brain structure and immunometabolic function. Latent variables including brain structure, depression and immunometabolic function were estimated in the model and are shown adjacent to the model panel, respectively. All significant paths could pass the FDR correction ( $\alpha = 0.05$ ). \*Two-sided unadjusted  $P < 0.05$ ; \*\*two-sided unadjusted  $P < 0.01$ ; \*\*\*two-sided unadjusted  $P < 0.001$ .

Previous studies identifying the relationship between single lifestyle factors and depression found that lack of social connection<sup>19</sup>, higher sedentary behavior levels<sup>20</sup>, failure to engage in physical activities<sup>21,22</sup>, smoking<sup>23,24</sup> and an unhealthy diet<sup>25</sup> are associated with an increased risk of depression. Similar to previous studies showing that the relationships between depression and sleep duration<sup>26,27</sup> or alcohol consumption are U-shaped<sup>28</sup>, we also found that light-to-moderate drinking and 7–9-h sleep duration were optimal to reduce the risk of depression. Among all these protective risk factors, having optimal sleep showed the greatest risk reduction for depression, including a single depressive episode and treatment-resistant depression. These findings are consistent with previous evidence showing improvements in depressive symptoms following non-pharmacological insomnia-specific treatment<sup>29</sup>. As for recurrent depressive disorder, frequent social connection was the most protective factor, indicating the benefits of social activity prescription in helping patients better manage their depressive conditions<sup>30</sup>. By incorporating multidimensional factors, our study shows the impact of a broad spectrum of lifestyle factors on depression as well as an approximate comparison between the impacts of these separate factors while controlling for their interaction.

To further determine a generalized effect of incorporated lifestyle index on depression, we combined the seven lifestyle factors and stratified participants into unhealthy, intermediate and favorable lifestyle classes<sup>31,32</sup>. In addition, we also conducted analyses of combined lifestyle and genetic risk classes, aiming to better compare depression risk for individuals with different levels of lifestyles and polygenic risk profiles. Our findings of an overall monotonic association with an increasing genetic risk and an unfavorable lifestyle and

a non-significant gene-by-lifestyle interaction are consistent with a previous study<sup>33</sup> that revealed that adherence to a healthy lifestyle may lower the risk of depression, regardless of the genetic risk. These results indicate that a healthy lifestyle plays a significantly strong protective role in any polygenic risk population. They further highlight the importance of adjusting to a healthy lifestyle to reduce the risk of depression.

The potential mechanism by which lifestyle factors impact depression risk is complex and multifactorial, involving genetic, behavioral and neurobiological aspects. A previous study found that the impacts of unfavorable lifestyle factors on dysregulated pathways are associated with depression, and involve neurotransmitter processes and immuno-inflammatory measures<sup>34</sup>. Another review has also suggested that the contribution of genetic variation to the mechanistic aspects of brain function is implicated in psychiatric disorders<sup>17</sup>. Our study employed a structural equation model to investigate the precise mechanisms by which lifestyle, brain structure, immunometabolic function and genetic risk affect depression. We found that the pathway from lifestyle to immunometabolic function is significant. Previous research has associated life stress exposures with indicators of adverse glycemic measures, deterioration of immune function and the accelerated accumulation of age-associated molecular and cellular damage<sup>35</sup>. Studies have reported that poor physical activity and sleep restriction result in dysregulation of the hypothalamic-pituitary-adrenal (HPA) axis<sup>36,37</sup>. Dietary habits are related to the levels of specific metabolites present in plasma<sup>38,39</sup>. Social status, which may affect stress levels, influences the immune system at multiple levels and alters signaling pathways in response to infection<sup>40</sup>. The influence of immunometabolic factors on



depression is noteworthy and consistent with previous investigations. A previous study reported that immunometabolic dysregulation is also critical in the pathophysiology of depression<sup>41</sup>. Increased inflammation and disrupted energy-regulating neuroendocrine signaling (for example, leptin and insulin) in depressed patients have also been found<sup>42</sup>, suggesting that a dysfunctional molecular condition in the neuroimmune system is potentially the cause of the exacerbation of depression symptoms.

Altered brain morphology is another plausible mechanism underlying the association of lifestyle with depression. Specifically in our results, higher lifestyle scores are correlated with a larger volume of the orbitofrontal cortex and the medial prefrontal cortex, which might suggest improved cognitive control and emotion regulation<sup>43,44</sup>. It is noted that the posterior lateral and medial orbitofrontal cortex and medial prefrontal cortex have shown gray-matter volumetric reductions and functional impairment in patients with mood disorders<sup>44</sup>. For specific aspects of lifestyle, previous research has demonstrated that amygdala neurons play a pivotal role in mediating social behaviors<sup>45</sup> and that the acquisition of knowledge through social networks is associated with activation of the lateral prefrontal cortex and medial prefrontal cortex<sup>46</sup>. Consumption of an unhealthy diet and addiction to smoking may lead to encephalopathy and focal brain damage, which are associated with depression symptoms<sup>24,47</sup>. Adequate sleep also promotes synaptic plasticity mechanisms in the hippocampus that optimize emotional responses to future behavioral stressors<sup>48</sup>. Lack of physical activity and excessive consumption of alcohol may lead to oxidative stress, neuroadaptive changes and differences in brain structure and function<sup>49,50</sup>, which are associated with mental health disorders. The findings described above support our results regarding the lifestyle-related brain structures and validate the modulatory role of brain function in the association between lifestyle and depression. However, in the context of the SEM framework, the association between the brain and immunometabolic function was not significant, although the brain-immune interaction has previously been suggested as a biological basis for major depressive disorder<sup>51,52</sup>. We supposed that the two strongly significant paths between brain measures and immunometabolic function with depression may attenuate the effect of brain-immune interaction in the comprehensive mathematical model. Further studies could examine an experimental and clinical approach to precisely test the molecular and cellular mechanisms of the brain-immune interaction associated with depression.

In addition, we further considered the bidirectional and multifactorial relationship between lifestyle and depression. The reversed MR result revealed the bidirectional causal effect of depression on lifestyle. We also added two possible SEMs to respectively explain the reversed directionality of depression on lifestyle and the mediation mechanism of immunometabolic function and brain structure mechanisms underlying the path from depression to lifestyle. The findings of the two models are aligned with previous evidence indicating that major depressive disorder is linked to an unhealthy lifestyle, characterized by insomnia or hypersomnia, psychomotor agitation or retardation, and fluctuations in appetite<sup>53</sup>. Additionally, individuals with depression were also accompanied by a change of motivation to socialize and decreased social participation in life<sup>54</sup>. Depressed patients also exhibit signs of immune activation, including elevated levels of proinflammatory cytokines, C-reactive protein and cortisol, which may interact with the HPA axis and sympathetic nervous system, leading to a disruption of circadian rhythms<sup>55</sup>, changes in appetite<sup>56</sup>, motor abnormalities<sup>57</sup> and stress<sup>58</sup>. Another study utilizing functional MRI revealed that depression-related increases in appetite are associated with hyperactivation of the mesocorticolimbic reward circuitry, and depression-related appetite loss is linked to hypoactivation of insular regions<sup>59</sup>.

Furthermore, we also found that the impact of lifestyle on depression is notably higher than any other significant relationship in the

SEM, with an effect size ranging from three- to tenfold. The effect sizes of our results are similar to those of previous studies<sup>33,60</sup> and this could be attributed to a variety of reasons. First, in the SEM, we noted that the impact of comprehensive lifestyle on depression is notably higher due to the combined effect of multiple mechanisms. Previous studies have identified that single lifestyle factors may influence the prevalence of depression through diverse genetic and biological pathways, as seen in the case of sleep. Sleep not only impacts the HPA axis, hippocampal oscillations and sympathetic nervous system<sup>36</sup>, but also changes the levels of proinflammatory cytokines, C-reactive protein and cortisol levels in the peripheral immune system<sup>55</sup>. Furthermore, immunometabolic markers in peripheral systems may be modulated by a negative feedback mechanism to ensure the stability of the environment in the body and the maintenance of immune homeostasis<sup>61</sup>. As for brain structural imaging, the association between brain structure or function and complex cognitive or mental health phenotypes may be limited by small sample sizes, which may potentially have been too small to capture reproducible brain-behavioral phenotype associations<sup>62</sup>. Therefore, there may be a mild effect of immunometabolic and brain markers on depression symptoms. It is also worth noting that previous studies have also shown that the HR of lifestyle on depression is more significant than those of genetic factors<sup>33</sup>. Importantly, environmental factors, such as lifestyle, can be modified, in contrast to genetic factors.

There are three main strengths of the present study. First, it utilizes the large UKB cohort and investigates the association of multidimensional lifestyle factors with depression and the underlying neurobiological mechanisms involving brain structural imaging and biochemical, metabolic and genetic data. The use of large sample sizes guarantees the precise estimation of effect sizes. We conducted an exploratory analysis to identify broad phenotypes of brain structures and peripheral systems associated with lifestyle. Because the dataset is longitudinal, we also considered the time difference for behavioral and neuroimaging data collection and added additional analyses to examine the consistency over time. Second, we define a healthy lifestyle based on validated national guidelines and previous studies, which guarantees a standard lifestyle criterion for populations. In addition, our findings regarding the combined effect of genetic and lifestyle factors on depression highlights that adherence to a favorable lifestyle is beneficial for populations with different genetic vulnerabilities. Although our findings address the risk of depression and how it can be mitigated by a healthy lifestyle, it is likely that maintaining a healthy lifestyle would be beneficial for those with depression. Finally, we performed structural equation modeling to specify multifactorial interplayed processes and potential mechanisms underlying this association, including lifestyle, genetics, brain structure, immunometabolic function and depression.

Our study has several limitations. First, lifestyle was assessed by subjective self-report questionnaires, which might be prone to measurement error. In future, objective measures, such as accelerometer-based ones, could be used. However, due to time limitations and the availability of technology, these are often not feasible for use in large-scale populations. In this study, the use of actigraphy data would result in a relatively smaller sample size and an inadequate follow-up time, which would limit our statistical model. As for outcomes, we used the date of depression diagnosis, which is a typical procedure for Cox model estimation, but does not take into account the time lapsing between the first onset of an episode and being diagnosed. Second, our results should be further validated in an independently ascertained population. Our study is limited due to the selection bias of the UKB cohort, as the participants are relatively healthier than the general population. There are limited data on ethnic minorities. We also acknowledge that the large number of variables collected by the UKB can make it difficult to identify and control for all potential confounding factors, which may have biased our observational results. Future research utilizing longitudinal brain imaging and peripheral markers

would provide a comprehensive illustration of the changing association pattern during the aging process. Third, we have acknowledged the incongruity of the stratification methods employed in lifestyle and PRS categorization according to previously published studies. Although the categories were not established based on a uniformly validated standard, we characterized better comparisons for the risk of depression between participants distributed in different lifestyle and genetic risk levels. The outcomes of categorized and continuous lifestyle scores were generally consistent, with an increase in lifestyle score being associated with a lower risk of depression, which confirms the robustness of our results. Finally, we used the PHQ-4 score as an assessment of depression symptoms to estimate the SEM, instead of the depression diagnosis, considering that the diagnosis of depression is a binary variable, and the dates were distributed over a longer period ranging from 2006 to 2021. We additionally employed an independent *t*-test on the PHQ-4 score between the diagnosed-depression group and the non-depression group and found a significant difference ( $t = 52.589, P < 2.2 \times 10^{-16}$ ), so we suggest that the continuous PHQ-4 score better reflects the depressive status. The limited number of participants with a diagnosis of depression would result in a smaller sample size (310 with onset depression in a total of 13,324 participants with completed multimodal data used in the SEM), which would limit the accuracy of the model estimation. Future research could utilize clinical diagnostic records and larger sample sizes of patients with a diagnosis of depression.

In conclusion, the present study has confirmed a causal protective relationship between multiple dimensional lifestyle factors and the risk of depression. Adherence to a healthy lifestyle would aid in the prevention of depression across a population with different genetic risk profiles. A structural equation model was used to reveal the underlying neurobiological mechanisms involving lifestyle, brain structures, immunometabolic function, genetics and depression. Our results highlight the importance of promoting a comprehensive healthy lifestyle for depression intervention.

## Methods

### Study population

The UK Biobank (UKB) is a prospective cohort that recruited more than 500,000 participants aged 37–73 years who attended one of 22 assessment centers between 2006 and 2010<sup>63</sup>. The UKB has received approval from the National Information Governance Board for Health and Social Care and the National Health Service North West Centre for Research Ethics Committee (ref. 11/NW/0382). All participants provided informed consent through electronic signature. The data utilized in the analyses contains demographic and behavioral assessments, depression diagnoses, brain structural imaging, blood biochemistry markers, blood cell counts and NMR metabolic markers.

### Lifestyle factors and healthy lifestyle score

We constructed a lifestyle score for 394,053 participants based on seven factors—smoking, physical activity, alcohol consumption, diet, sleep duration, sedentary behavior and social connection—that were assessed at baseline using a touchscreen questionnaire<sup>32</sup>. Full details for each factor, corresponding questionnaires and national guidelines are listed in Supplementary Table 3. Participants scored one point for the healthy category of each factor based on national recommendations. According to the quit smoking guidelines from the National Health Service (NHS) and National Institute for Health and Care Excellence (NICE) 2015 quality standard on ‘Smoking: reducing and preventing tobacco use’, never smoking was classified as the healthy category<sup>31,64</sup>, and current and previous smoking were considered as the unhealthy category. Physical activity was assessed using the International Physical Activity Questionnaire-Short Form (IPAQ-SF)<sup>65</sup>. Regular physical activity was defined as meeting the American Heart Association recommendations and World Health Organization guidelines (at least

150 min of moderate activity or 75 min of vigorous activity per week (or an equivalent combination) or engaging in moderate physical activity at least five days a week or vigorous activity once a week (more than 10 min)<sup>66,67</sup>. Due to the U-shaped relationship between alcohol consumption and depression risk<sup>28</sup>, moderate alcohol consumption was defined as never drink or 0–14 g per day for women and 0–28 g per day for men, with the maximum limit reflecting US Dietary Guidelines for Americans (DGA)<sup>68,69</sup>. Following the DGA<sup>70–72</sup> and recommendations on dietary priorities for cardiometabolic health<sup>68</sup>, healthy diet was based on an appropriate consumption of at least four of seven food groups including fruits, vegetables, fish, processed meats, unprocessed red meats, whole grains and refined grains, with specific cutoff criteria as listed in Supplementary Table 3. In accordance with the NHS, the American Academy of Sleep Medicine (AASM) and the Sleep Research Society (SRS) recommendation<sup>73</sup>, sleep duration of 7–9 h was considered healthy. According to the WHO guidelines<sup>66</sup> and previously published literature<sup>32</sup>, screen-based sedentary behavior was estimated according to the Global Physical Activity Questionnaire (GPAQ)<sup>74</sup> as the sum of self-reported hours spent watching TV and using a computer (not including using a computer at work) during a typical day, with 0–4 h reclassified as low-to-moderate and healthy. Social connection was assessed according to the social isolation index<sup>75</sup>. The index sums up the following three items: number of people in the household (one point was given for living alone), frequency of friend/family visits (one point was given for answering about once a month, once every few months, never or almost never, or no friends or family outside household), and participation in leisure/social activity (one point was given for answering none)<sup>75</sup>. Individuals were defined as least isolated when scoring 0, moderately isolated when scoring 1 and most isolated when scoring 2 or 3<sup>75</sup>. Participants who were least and moderately isolated were defined as healthy with frequent social connection<sup>32,75</sup>. The lifestyle score ranged from 0 to 7, with a higher score indicating adherence to a healthier lifestyle<sup>31</sup>. The lifestyle score was subsequently categorized as favorable (5 to 7 lifestyle factors), intermediate (2 to 4) and unfavorable (0 to 1) for further analyses<sup>31,32</sup>.

### Depression diagnoses

Depression diagnoses were ascertained using hospital inpatient records from Hospital Episode Statistics for England, Scottish Morbidity Record data for Scotland and Patient Episode Database for Wales. Additional cases were detected through linkage to death register data from the National Health Service Digital, National Health Service Central Register and National Records of Scotland. Diagnoses were recorded according to the International Classification of Diseases (ICD-10) system codes, including depressive episode (F32.0–9) and recurrent depressive disorder (F33.0–9). Depression diagnoses were also retrieved from primary care data using read codes (version 2 (Read v2) and version 3 (CTV3 or Read v3)), which are presented in Supplementary Table 1. We integrated the data from hospital inpatient, primary care and death register sources, while those from self-report only were excluded. The date of diagnosis was defined as the earliest date recorded.

### Structural MRI data

Quality-controlled T1-weighted neuroimaging data ( $n = 39,932$ ) were measured at the UKB assessment center. The scanner was a standard Siemens Skyra 3T system with a standard Siemens 32-channel RF receive head coil. The details of image acquisition are provided on the UKB website (<http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=2367>). T1 images were processed with FreeSurfer. Surface templates were used to extract imaging-derived phenotypes, referred to as atlas regions’ surface volume<sup>76</sup>. Subcortical regions were extracted via FreeSurfer’s aseg tool<sup>77</sup>. FreeSurfer aparc (category ID 192) and the aseg (category ID 190) atlas corresponding to 68 cortical regions<sup>76</sup> and 14 subcortical regions<sup>77</sup> were used in this study. We used the Qoala-T approach to check the FreeSurfer outputs and excluded those that failed to pass

quality control from the FreeSurfer imaging-derived phenotypes. By merging data with the lifestyle score, we used 32,839 brain structural imaging data for correlation analyses.

## Peripheral markers

Blood biochemistry (category ID 17518) and blood count (category ID 100081) data were taken from ~480,000 participants at the baseline assessment (2006–2010). The biomarker assay quality procedures are provided in an open-source document ([https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/biomarker\\_issues.pdf](https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/biomarker_issues.pdf)). Beckman Coulter LH750 instruments were used to analyze the blood count samples, which were collected in 4 ml of ethylenediaminetetraacetic acid (EDTA) vacutainers. More information about the hematology analysis is provided at <https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/haematology.pdf>. Thirty blood biochemistry markers and 31 blood cell counts were available from these two UKB categories. The study used 59 markers from 61, excluding nucleated red cell count and percentage because the count or percentage of the majority (>90%) was 0. We then categorized the blood biochemistry markers as ‘liver function’, ‘renal function’, ‘endocrine’, ‘immunometabolic’ or ‘bone and joint’ according to previous literature<sup>78</sup>. Blood count was categorized as ‘white blood cell’, ‘red blood cell’ or ‘platelet’. More details are provided in Supplementary Table 18.

Metabolic markers were measured in randomly selected EDTA plasma samples using an NMR-based metabolic biomarker profiling platform, which included ~120,000 UKB participants at the baseline assessment. The process for assessing the metabolic markers is available at <https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=220>. The NMR metabolomics (category ID 220) included 249 metabolic markers (168 directly measured and 81 ratios of these). We used 168 directly measured metabolic markers and categorized metabolic markers as ‘amino acids’, ‘apolipoproteins’, ‘lipoprotein particle sizes’, ‘lipoprotein particle concentrations’, ‘fatty acids’, ‘triglycerides’, ‘phospholipids’, ‘cholesteryl esters’, ‘free cholesterol’, ‘cholesterol’, ‘other lipids’, ‘total lipids’, ‘ketone bodies’, ‘glycolysis-related metabolites’, ‘fluid balance’ and ‘inflammation’. More details are available in Supplementary Table 19.

## PRS for depression

The computation of PRSs for individuals required two distinct datasets. The first dataset, consisting of individual-level genotype data, was obtained from all 502,409 participants in UKB v3 imputation. Detailed genotyping and quality control procedures are available in a previous publication<sup>79</sup>. We excluded SNPs with call rates of <95%, minor allele frequency of <0.1%, deviation from the Hardy–Weinberg equilibrium with  $P < 1 \times 10^{-10}$  and selected subjects with British ancestry that had no more than ten putative third-degree relatives in the kinship table. Finally, we obtained 8,239,652 SNPs from 337,151 participants.

The second dataset was a meta-analysis of large-scale genome-wide association (GWAS) results for depression<sup>80</sup>, which included 9,874,289 total SNPs from 480,359 individuals (135,458 cases and 344,901 controls). After excluding data from the 23andMe (75,607 cases and 231,747 controls, restricted by privacy) and UKB (14,260 cases and 15,480 controls) to avoid sample overlap with the UKB genotype data, this collective GWAS encompassed 6,435,918 variants from 143,265 individuals (45,591 cases and 97,674 controls).

We utilized PRSice-2 (<http://www.prsice.info>) to calculate the PRS for depression. First, SNPs were clumped with a cutoff of  $r^2 = 0.1$  in a 250-kb window<sup>81</sup>. Second, the  $P$  thresholds were set at  $P < 0.0005$ ,  $P < 0.001$ ,  $P < 0.005$ ,  $P < 0.01$ ,  $P < 0.05$ ,  $P < 0.1$ ,  $P < 0.5$  and  $P < 1$  (ref. 82), and we finally used PRS ( $P < 0.05$ ) for further analyses. The PRS was then categorized into low (lowest quintile), intermediate (quintiles 2 to 4) and high (highest quintile) risk levels<sup>31,32</sup>. We also used the original GWAS, excluding samples from 23andMe to calculate the PRS, which included 5,401,556 variants and 173,005 individuals (59,851 cases and 113,154 controls), and found a high correlation between the two PRSs ( $r = 0.512$ ,  $P < 1.0 \times 10^{-295}$ ).

## Covariates

The data for covariates were collected at baseline. All models were adjusted for age at baseline (field ID 21022), sex (field ID 31), Townsend deprivation index (field ID 189, referring to an area-based measure of socioeconomic deprivation), BMI (field ID 21001) and education level (field ID 6138). Association analyses involving brain structural imaging data were additionally adjusted neuroimaging scanning sites and estimated total intracranial volume. Association analyses involving genetic data additionally corrected for the top 20 ancestry principal components. The detailed covariates utilized in each statistical analysis are presented in Supplementary Fig. 2.

## Statistical analyses

**Cox proportional hazard regression model.** Cox proportional hazard regression models were utilized to examine the association of lifestyle categories, genetic risk categories and the combination of genetic and lifestyle categories (nine categories using low genetic risk and favorable lifestyle as reference) with time to depression. The missing data of covariates were imputed by their mean value. The proportion of missing data is presented in Supplementary Table 2. HRs and corresponding 95% CIs were calculated in complete-case analyses. Participants were considered at risk for depression from baseline (2006–2010) and followed up until the date of first diagnosis, death, loss to follow-up or last date with available information (April 2021), whichever came first. The proportional hazards assumption was assessed using the Schoenfeld residuals method<sup>83</sup> and satisfied.

**Association of lifestyle scores with brain volume and peripheral markers.** To identify the mediating role of brain structural imaging and peripheral markers underlying the association between lifestyle and depression, we conducted an exploratory Pearson correlation analysis of lifestyle at baseline with these neurobiological phenotypes. FDR correction was conducted for multiple comparisons (Supplementary Table 14). Considering the time difference from baseline (2006–2010) to imaging visit timepoint (2014+), we also calculated correlation analyses between brain volumes and lifestyle scores assessed at the imaging visit (Supplementary Table 15). Depression scores at the two timepoints were also correlated with brain volumes (Supplementary Tables 16 and 17). Both lifestyle and depression score at the two collected timepoints were similar and highly correlated (lifestyle,  $r = 0.85$ ,  $P = 1.2 \times 10^{-23}$ ; depression,  $r = 0.85$ ,  $P = 3.2 \times 10^{-24}$ ). Therefore, lifestyle and depression did not change to a great extent. We further calculated the spatial correlation to compare the similarity of lifestyle and depression-related brain maps at baseline and imaging visit (Supplementary Fig. 18).

The associations of lifestyle at baseline with peripheral markers were also tested, with Bonferroni correction conducted for multiple comparisons (Supplementary Tables 18 and 19).

**Mendelian randomization.** We further investigated the possible causal relationship between lifestyle and depression using bidirectional two-sample MR analysis, which employs genetic variants as the instrumental variables to estimate the corresponding causal effect size.

For lifestyle, we performed GWAS on lifestyle scores. The genotype data were the same as used in the PRS calculation (including 8,239,652 SNPs from 337,151 participants). We used PLINK 2.0 (<https://www.cog-genomics.org/plink/2.0/>) for the calculation, adjusting for age, sex and the first 20 ancestry principal components. For depression, we also used the meta-analysis of depression GWAS<sup>80</sup> and excluded the participants from UKB and 23andMe to avoid sample overlap, which was consistent with the PRS calculation.

Two-sample MR analyses used four methods—IVW, weighted median, simple median and weighted mode—implemented in the R package TwoSampleMR (<https://mrcieu.github.io/TwoSampleMR/>). The genetic instruments were selected with  $P < 1 \times 10^{-7}$  and we removed correlated SNPs ( $r^2 > 0.01$ , kb < 1,000) to avoid linkage-disequilibrium.



Cochran's Q statistics was used to assess the heterogeneity of IVW, and radial MR was used to remove the SNPs contributing the most heterogeneity. After removing two SNPs with significant heterogeneity ( $rs4930349$ ,  $P_{IVW,Q} = 3.4 \times 10^{-2}$ ;  $rs66994942$ ,  $P_{IVW,Q} = 4.2 \times 10^{-3}$ ), we performed the heterogeneity test again by using Q statistics and the Mendelian Randomization-Pleiotropy Residual Sum and Outlier (MR-PRESSO) test. No further heterogeneity was observed ( $P_{IVW,Q} = 0.26$ ). The MR-Egger method was used and found no sign of pleiotropy ( $P_{\text{pleiotropy}} = 0.52$ ). In the main MR analysis, we utilized 26 SNPs and prioritized the models using the fixed-effect IVW method, because no pleiotropy and heterogeneity effects were observed. For sensitivity analyses, we performed the simple median method, weighted median and weighted mode methods to ensure the robustness of the IVW result. We also assessed the reverse causality by testing the effect of depression genetic liability (using the same  $P$  value and clumping threshold) on lifestyle similarly. Finally, we created scatter and forest plots to visually inspect the data for a strong influence by single variants (Fig. 4 and Supplementary Figs. 13 and 14).

**Structural equation model.** A structural equation model was estimated for participants with lifestyle scores, depression scores, brain structural imaging, PRS and immunometabolic markers (implemented in the R package lavaan 0.8). Three latent variables were estimated using confirmatory factor analysis. The latent variable representing depression was estimated using four PHQ-4 items. The latent variable of immunometabolic function was calculated from the first four significant markers correlated with lifestyle score. Finally, the latent variable for brain structures was derived from the first 20 cortical and subcortical brain volumes significantly correlated with lifestyle score. These three latent variables were investigated to determine the directional dependencies with PRS and lifestyle via path modeling. Before inputting into the model, we normalized the PRS, lifestyle score and each sub-score of the latent variables, respectively, to maintain the same scale. FDR corrections were conducted to adjust the  $P$  values accounting for multiple comparisons of all paths.

**Sensitivity analyses.** In the sensitivity analyses, we further tested the association of lifestyle with risk of three depression subtypes: single depressive episode, recurrent depressive disorder and TRD. The ICD-10 defines 'depressive episode' as F32.0-9, indicating a singular occurrence, and 'recurrent depressive disorder' as F33.0-9. The data were also retrieved from primary care by utilizing Read codes (Supplementary Table 1). According to the treatment records in the primary care prescriptions table (gp\_scripts, UKB Field ID 42039), antidepressant use was gathered utilizing British National Formulary codes 0403010 (tricyclic and related antidepressant drugs), 0403020 (monoamine oxidase inhibitors), 0403030 (selective serotonin reuptake inhibitors) and 0403040 (other antidepressant drugs). TRD was defined by at least two switches between antidepressant drugs, each prescribed for at least six weeks<sup>84</sup>, as well as having partially distinct genetic<sup>85</sup> and clinical-demographic characteristics<sup>86</sup> compared with non-TRD. As such, TRD should be considered as a unique subtype of depression<sup>87-89</sup>. Non-response after the second medication switch was confirmed for subjects considered treatment-resistant.

Due to the limited sample size scoring 0 and 1 in recurrent depressive disorder and TRD, we combining 0 and 1 as the reference group (Supplementary Figs. 7 and 8). Considering the difference between previous smokers and current smokers, we further calculated the risk of depression according to different smoking statuses with never smoking as reference (Supplementary Table 6). Regarding the small reference group issue, we also used favorable lifestyle class with a larger sample size as reference (Supplementary Table 7). In addition, we calculated depression risk according to genetic and lifestyle across three subtypes of depression (Supplementary Figs. 9-11).

Considering the change of depression symptoms during the follow-up period from baseline to imaging visit, we selected two groups based on the PHQ-4 classifying criterion<sup>90</sup>. One group was severely depressed at baseline (PHQ-4  $\geq 9$ ) and recovered at imaging visit (PHQ-4  $\geq 6$ ); this was labeled the recovered group ( $n = 538$ ). The other group was suffering persistent symptoms over time (PHQ-4  $\geq 9$ , whether at baseline or imaging visit); this was labeled the persistent group ( $n = 809$ ). To investigate whether the change in depression symptoms would impact the association between depression and brain volumes, we utilized an independent-sample  $t$ -test to examine the difference in brain volumes between the two groups. We found there were only two significant brain volumes (left middle temporal and right insula), both with  $P = 0.04$ , but neither survived the FDR correction (Supplementary Table 20). We also considered whether antidepressant treatment would impact the association between depression and brain volumes in these two groups. In the recovered group, we further classified the treated-recovered group ( $n = 132$ ) and not-treated-recovered group ( $n = 451$ ) based on antidepressant use. We found there were only two significant brain volumes (right cuneus and right posterior cingulate), with  $P = 0.03$  and  $0.01$ , but neither survived the FDR correction (Supplementary Table 21). In the persistent group, we also classified the treated-persistent group ( $n = 251$ ) and the not-treated-persistent group ( $n = 558$ ) based on antidepressant use. We found that there was only one significant brain volume (right accumbens area), with  $P = 0.03$ , but it did not survive FDR correction (Supplementary Table 22). As such, we suggest that the time difference between baseline and imaging visit did not substantially influence the correlation between depression and brain structure.

Finally, we estimated two possible SEMs accounting for multiple directionalities among the latent variables. In the first model, we changed the direction from depression to lifestyle and kept the other paths' direction unchanged to test the reversed effect (Supplementary Fig. 19). In the second model, we further examined the mediation mechanism of immunometabolic function and brain structure underlying the path from depression to lifestyle, by changing the directions between depression and immunometabolic function, depression and brain structure, immunometabolic function and lifestyle, and brain structure and lifestyle based on possible model 1 (Supplementary Fig. 20). Both of these possible models were fitted under the same procedures as the original model and provided an improved interpretation of our results.

## Reporting summary

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

## Data availability

Access to individual-level UK Biobank data (phenotypic, neuroimaging and genotype) is available to bona fide researchers through application to the UK Biobank website (<https://www.ukbiobank.ac.uk>). Additional information about registration for access to the data is available at <http://www.ukbiobank.ac.uk/register-apply/>. Use of UK Biobank data was performed under application no. 19542. Summary statistics from previous GWAS of depression that were used in this study are publicly available through the Psychiatric Genomics Consortium (Psychiatric Genomics Consortium, PGC, unc.edu) and can be downloaded from <https://pgc.unc.edu/for-researchers/download-results/> (doi 10.6084/m9.figshare.21655784; PubMed 29700475; file daner\_pgc\_mdd\_meta\_w2\_no23andMe\_rmUKBB.gz)

## Code availability

The code used for survival analyses is an adaptation of the R package survival (<https://github.com/therneau/survival>) and has been made available through the GitHub repository: [https://github.com/yjzhao1004/lifestyle\\_depression](https://github.com/yjzhao1004/lifestyle_depression). GWAS analyses were conducted



by using PLINK 2.0 (<https://www.cog-genomics.org/plink/2.0/>). We utilized PRSice-2 (<https://choishingwan.github.io/PRSice/>) to conduct PRS analysis. Two sample MR analyses was performed using the R package TwoSampleMR (<https://mrcieu.github.io/TwoSampleMR/>) and RadialMR (<https://github.com/WSpiller/RadialMR/>). SEM was constructed using the R package lavaan 0.8 (<https://lavaan.ugent.be>). All analyses that were run in R were run in R v.4.2.1 (<https://www.r-project.org>).

## References

- Winter, N. R. et al. Quantifying deviations of brain structure and function in major depressive disorder across neuroimaging modalities. *JAMA Psychiat* **79**, 879–888 (2022).
- Friedrich, M. J. Depression is the leading cause of disability around the world. *JAMA* **317**, 1517 (2017).
- Collins, P. Y. et al. Grand challenges in global mental health. *Nature* **475**, 27–30 (2011).
- Firth, J. et al. A meta-review of ‘lifestyle psychiatry’: the role of exercise, smoking, diet and sleep in the prevention and treatment of mental disorders. *World Psychiatry* **19**, 360–380 (2020).
- Moore, T. M., Hauser, R. M., Depp, C. A. & Eyster, L. T. Rethinking the impact of genetic risk for depression: an analysis of population-subgroup differences in a large electronic health record dataset. *Front. Psychiatry* **8**, 46 (2017).
- Peyrot, W. J. et al. Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5,765 subjects from the psychiatric genomics consortium. *Biol. Psychiatry* **84**, 138–147 (2018).
- Jacka, F. N. et al. A randomised controlled trial of dietary improvement for adults with major depression (the ‘SMILES’ trial). *BMC Med.* **15**, 23 (2017).
- Opel, N. et al. Preventing the onset of major depressive disorder: a meta-analytic review of psychological interventions. *Int. J. Methods Psychiatr. Res.* **29**, e1830 (2020).
- Opel, N. et al. Mediation of the influence of childhood maltreatment on depression relapse by cortical structure: a 2-year longitudinal observational study. *Lancet Psychiatry* **6**, 318–326 (2019).
- Sarris, J. et al. Multiple lifestyle factors and depressed mood: a cross-sectional and longitudinal analysis of the UK Biobank (N=84,860). *BMC Med.* **18**, 354 (2020).
- Pedersen, B. K. Physical activity and muscle-brain crosstalk. *Nat. Rev. Endocrinol.* **15**, 383–392 (2019).
- Mattson, M. P. Energy intake and exercise as determinants of brain health and vulnerability to injury and disease. *Cell Metab.* **16**, 706–722 (2012).
- Al-Massadi, O. et al. Multifaceted actions of melanin-concentrating hormone on mammalian energy homeostasis. *Nat. Rev. Endocrinol.* **17**, 745–755 (2021).
- Joutsa, J. et al. Brain lesions disrupting addiction map to a common human brain circuit. *Nat. Med.* **28**, 1249–1255 (2022).
- Mavromatis, L. A. et al. Association between brain structure and alcohol use behaviors in adults: a Mendelian randomization and multiomics study. *JAMA Psychiat* **79**, 869–878 (2022).
- Hari, R. & Kujala, M. V. Brain basis of human social interaction: from concepts to brain imaging. *Physiol. Rev.* **89**, 453–479 (2009).
- Meyer-Lindenberg, A. & Weinberger, D. R. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nat. Rev. Neurosci.* **7**, 818–827 (2006).
- Gregersen, P. K. & Behrens, T. W. Genetics of autoimmune diseases—disorders of immune homeostasis. *Nat. Rev. Genet.* **7**, 917–928 (2006).
- Holt-Lunstad, J. Social connection as a public health issue: the evidence and a systemic framework for prioritizing the ‘social’ in social determinants of health. *Annu. Rev. Public Health* **43**, 193–213 (2022).
- Huang, Y. et al. Sedentary behaviors and risk of depression: a meta-analysis of prospective studies. *Transl. Psychiatry* **10**, 26 (2020).
- Schuch, F. B. et al. Physical activity and incident depression: a meta-analysis of prospective cohort studies. *Am. J. Psychiatry* **175**, 631–648 (2018).
- Schuch, F. et al. Physical activity and sedentary behavior in people with major depressive disorder: a systematic review and meta-analysis. *J. Affect. Disord.* **210**, 139–150 (2017).
- Bjørngaard, J. H. et al. The causal role of smoking in anxiety and depression: a Mendelian randomization analysis of the HUNT study. *Psychol. Med.* **43**, 711–719 (2013).
- Tidey, J. W. & Miller, M. E. Smoking cessation and reduction in people with chronic mental illness. *BMJ* **351**, h4065 (2015).
- Lassale, C. et al. Healthy dietary indices and risk of depressive outcomes: a systematic review and meta-analysis of observational studies. *Mol. Psychiatry* **24**, 965–986 (2019).
- Zhai, L., Zhang, H. & Zhang, D. Sleep duration and depression among adults: a meta-analysis of prospective studies: Research Article: sleep duration and depression. *Depress. Anxiety* **32**, 664–670 (2015).
- Li, Y. et al. The brain structure and genetic mechanisms underlying the nonlinear association between sleep duration, cognition and mental health. *Nat. Aging* **2**, 425–437 (2022).
- Almeida, O. P., Hankey, G. J., Yeap, B. B., Golledge, J. & Flicker, L. The triangular association of ADH1B genetic polymorphism, alcohol consumption and the risk of depression in older men. *Mol. Psychiatry* **19**, 995–1000 (2014).
- Li, L., Wu, C., Gan, Y., Qu, X. & Lu, Z. Insomnia and the risk of depression: a meta-analysis of prospective cohort studies. *BMC Psychiatry* **16**, 375 (2016).
- Drinkwater, C., Wildman, J. & Moffatt, S. Social prescribing. *BMJ* **364**, l1285 (2019).
- Lourida, I. et al. Association of lifestyle and genetic risk with incidence of dementia. *JAMA* **322**, 430–437 (2019).
- Wang, B. et al. Association of combined healthy lifestyle factors with incident dementia in patients with Type 2 Diabetes. *Neurology* **99**, e2336–e2345 (2022).
- Cao, Z. et al. Polygenic risk score, healthy lifestyles and risk of incident depression. *Transl. Psychiatry* **11**, 189 (2021).
- Lopresti, A. L., Hood, S. D. & Drummond, P. D. A review of lifestyle factors that contribute to important pathways associated with major depression: diet, sleep and exercise. *J. Affect. Disord.* **148**, 12–27 (2013).
- Kivimäki, M., Bartolomucci, A. & Kawachi, I. The multiple roles of life stress in metabolic disorders. *Nat. Rev. Endocrinol.* **19**, 10–27 (2023).
- Meerlo, P., Koehl, M., Van Der Borght, K. & Turek, F. W. Sleep restriction alters the hypothalamic-pituitary-adrenal response to stress: sleep restriction and HPA axis reactivity. *J. Neuroendocrinol.* **14**, 397–402 (2002).
- Martikainen, S. et al. Higher levels of physical activity are associated with lower hypothalamic-pituitary-adrenocortical axis reactivity to psychosocial stress in children. *J. Clin. Endocrinol. Metab.* **98**, E619–E627 (2013).
- Bar, N. et al. A reference map of potential determinants for the human serum metabolome. *Nature* **588**, 135–140 (2020).
- Chen, L. et al. Influence of the microbiome, diet and genetics on inter-individual variation in the human plasma metabolome. *Nat. Med.* **28**, 2333–2343 (2022).
- Snyder-Mackler, N. et al. Social status alters immune regulation and response to infection in macaques. *Science* **354**, 1041–1045 (2016).

41. Penninx, B. W., Milaneschi, Y., Lamers, F. & Vogelzangs, N. Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile. *BMC Med.* **11**, 129 (2013).
42. Milaneschi, Y., Lamers, F., Berk, M. & Penninx, B. W. J. H. Depression heterogeneity and its biological underpinnings: toward immunometabolic depression. *Biol. Psychiatry* **88**, 369–380 (2020).
43. MacFall, J. R., Payne, M. E., Provenzale, J. E. & Krishnan, K. R. R. Medial orbital frontal lesions in late-onset depression. *Biol. Psychiatry* **49**, 803–806 (2001).
44. Drevets, W. C. Orbitofrontal cortex function and structure in depression. *Ann. N.Y. Acad. Sci.* **1121**, 499–527 (2007).
45. Wei, J.-A. et al. Amygdala neural ensemble mediates mouse social investigation behaviors. *Natl. Sci. Rev.* **10**, nwac179 (2023).
46. Jiang, Y., Mi, Q. & Zhu, L. Neurocomputational mechanism of real-time distributed learning on social networks. *Nat. Neurosci.* **26**, 506–516 (2023).
47. Hadjihambi, A. et al. Partial MCT1 invalidation protects against diet-induced non-alcoholic fatty liver disease and the associated brain dysfunction. *J. Hepatol.* **78**, 180–190 (2023).
48. Aime, M. et al. Paradoxical somatodendritic decoupling supports cortical plasticity during REM sleep. *Science* **376**, 724–730 (2022).
49. Gozal, D., Nair, D. & Goldbart, A. D. Physical activity attenuates intermittent hypoxia-induced spatial learning deficits and oxidative stress. *Am. J. Respir. Crit. Care Med.* **182**, 104–112 (2010).
50. MacKillop, J. et al. Hazardous drinking and alcohol use disorders. *Nat. Rev. Dis. Primers* **8**, 80 (2022).
51. Branchi, I. et al. Brain-immune crosstalk in the treatment of major depressive disorder. *Eur. Neuropsychopharmacol.* **45**, 89–107 (2021).
52. Dantzer, R., O'Connor, J. C., Freund, G. G., Johnson, R. W. & Kelley, K. W. From inflammation to sickness and depression: when the immune system subjugates the brain. *Nat. Rev. Neurosci.* **9**, 46–56 (2008).
53. Baxter, L. C. Appetite changes in depression. *Am. J. Psychiatry* **173**, 317–318 (2016).
54. Segrin, C. Social skills deficits associated with depression. *Clin. Psychol. Rev.* **20**, 379–403 (2000).
55. Keller, J. et al. Cortisol circadian rhythm alterations in psychotic major depression. *Biol. Psychiatry* **60**, 275–281 (2006).
56. Andréasson, A., Arborelius, L., Erlanson-Albertsson, C. & Lekander, M. A putative role for cytokines in the impaired appetite in depression. *Brain Behav. Immun.* **21**, 147–152 (2007).
57. Damme, K., Park, J. S., Walther, S., Shankman, S. & Mittal, V. Depression and motor abnormalities across development, symptom dimensions and familial risk. *Biol. Psychiatry* **89**, S297–S298 (2021).
58. Plackett, B. The vicious cycle of depression and obesity. *Nature* **608**, S42–S43 (2022).
59. Simmons, W. K. et al. Depression-related increases and decreases in appetite: dissociable patterns of aberrant activity in reward and interoceptive neurocircuitry. *Am. J. Psychiatry* **173**, 418–428 (2016).
60. Bittner, N. et al. Combining lifestyle risks to disentangle brain structure and functional connectivity differences in older adults. *Nat. Commun.* **10**, 621 (2019).
61. Mueller, D. L. Tuning the immune system: competing positive and negative feedback loops. *Nat. Immunol.* **4**, 210–211 (2003).
62. Marek, S. et al. Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660 (2022).
63. Palmer, L. J. UK Biobank: bank on it. *Lancet* **369**, 1980–1982 (2007).
64. *Smoking: Reducing and Preventing Tobacco Use* (NICE, 2015); <https://www.nice.org.uk/guidance/qs82>
65. Lee, P. H., Macfarlane, D. J., Lam, T. & Stewart, S. M. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. *Int. J. Behav. Nutr. Phys. Act.* **8**, 115 (2011).
66. *WHO Guidelines on Physical Activity and Sedentary Behaviour* (World Health Organization, 2020).
67. Lloyd-Jones, D. M. et al. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic impact goal through 2020 and beyond. *Circulation* **121**, 586–613 (2010).
68. Mozaffarian, D. Dietary and policy priorities for cardiovascular disease, diabetes and obesity: a comprehensive review. *Circulation* **133**, 187–225 (2016).
69. Phillips, J. A. Dietary Guidelines for Americans, 2020–2025. *Workplace Health Saf.* **69**, 395–395 (2021).
70. Taylor, E. N., Fung, T. T. & Curhan, G. C. DASH-style diet associates with reduced risk for kidney stones. *J. Am. Soc. Nephrol.* **20**, 2253–2259 (2009).
71. Fung, T. T. et al. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Arch. Intern. Med.* **168**, 713–720 (2008).
72. Rai, S. K. et al. The Dietary Approaches to Stop Hypertension (DASH) diet, western diet, and risk of gout in men: prospective cohort study. *BMJ* **357**, j1794 (2017).
73. Watson, N. F. et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. *Sleep* **38**, 843–844 (2015).
74. Herrmann, S. D., Heumann, K. J., Der Ananian, C. A. & Ainsworth, B. E. Validity and reliability of the Global Physical Activity Questionnaire (GPAQ). *Meas. Phys. Educ. Exerc. Sci.* **17**, 221–235 (2013).
75. Smith, R. W. et al. Social isolation and risk of heart disease and stroke: analysis of two large UK prospective studies. *Lancet Public Health* **6**, e232–e239 (2021).
76. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).
77. Fischl, B. et al. Whole brain segmentation. *Neuron* **33**, 341–355 (2002).
78. Wainberg, M. et al. Clinical laboratory tests and five-year incidence of major depressive disorder: a prospective cohort study of 433,890 participants from the UK Biobank. *Transl. Psychiatry* **11**, 380 (2021).
79. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
80. Wray et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
81. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
82. Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium A genome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. *Nat. Commun.* **11**, 2301 (2020).
83. Schoenfeld, D. Partial residuals for the proportional hazards regression model. *Biometrika* **69**, 239–241 (1982).
84. Gaynes, B. N. et al. Defining treatment-resistant depression. *Depress. Anxiety* **37**, 134–145 (2020).
85. Fabbri, C. et al. Genetic and clinical characteristics of treatment-resistant depression using primary care records in two UK cohorts. *Mol. Psychiatry* **26**, 3363–3373 (2021).
86. Fava, M. Diagnosis and definition of treatment-resistant depression. *Biol. Psychiatry* **53**, 649–659 (2003).

87. Fagiolini, A. & Kupfer, D. J. Is treatment-resistant depression a unique subtype of depression? *Biol. Psychiatry* **53**, 640–648 (2003).
88. Malhi, G. S. & Byrow, Y. Is treatment-resistant depression a useful concept? *Evid. Based Ment. Health* **19**, 1–3 (2016).
89. Yang, L. et al. Depression, depression treatments, and risk of incident dementia: a prospective cohort study of 354,313 participants. *Biol. Psychiatry* **93**, 802–809 (2023).
90. Stanhope, J. Patient Health Questionnaire-4. *Occup. Med.* **66**, 760–761 (2016).

## Acknowledgements

This study used the UK Biobank Resource under application no. 19542. We thank all participants and researchers from the UK Biobank. The study was supported by grants from the National Natural Science Foundation of China (nos. 82071997 and 82071201), Science and Technology Innovation 2030 Major Projects (2022ZD0211600), the National Key R&D Program of China (nos. 2018YFC1312900 and 2019YFA0709502), Shanghai Municipal Science and Technology Major Project (no. 2018SHZDZX01), Shanghai Rising-Star Program (21QA1408700), 111 Project (B18015) and Zhangjiang Lab, Tianqiao and Chrissy Chen Institute, the State Key Laboratory of Neurobiology and Frontiers Center for Brain Science of the Ministry of Education, Shanghai Center for Brain Science and Brain-Inspired Technology, Fudan University.

## Author contributions

W.C., J.F. and J.Y. proposed the study. Yujie Zhao and L.Y. analyzed the data. W.C., J.F., J.Y., Yang Zhao and B.J.S. contributed to interpretation of results. Yujie Zhao and L.Y. drafted the manuscript. B.J.S., C.L., Yang Zhao and W.C. edited the manuscript. Yujie Zhao, L.Y., J.Y., W.Z., K.K., Z.L., Y.G., Y.L., Yang Zhao and W.C. contributed to visualization. All authors considered how to analyze the data and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Ethics statement

The UK Biobank has ethical approval by the National Research Ethics Service (ref. 11/NW/0382).

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s44220-023-00120-1>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44220-023-00120-1>.

**Correspondence and requests for materials** should be addressed to Jintai Yu, Jianfeng Feng or Wei Cheng.

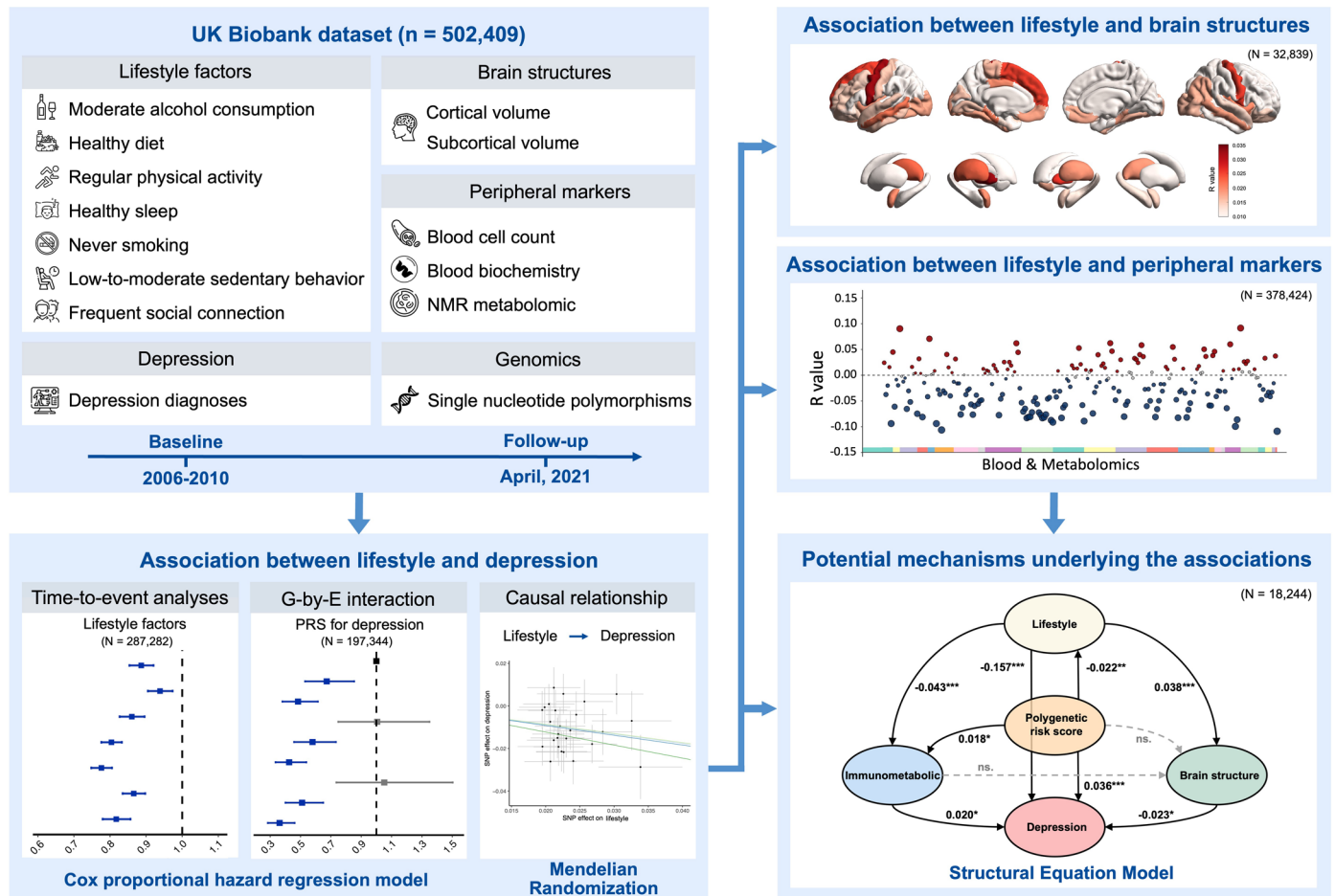
**Peer review information** *Nature Mental Health* thanks Mark Kvarda, Femke Lamers, Oleg Medvedev and Tiina Paunio for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2023



**Extended Data Fig. 1 | Guideline of the study.** Top left, UK Biobank data used in the study including lifestyle factors, depression diagnoses, brain structural imaging, peripheral markers and genomics. Bottom left, cox model specifying the depression risk according to lifestyle and genetics and mendelian randomization specifying the causal relationship between lifestyle and depression. Top right,

association between lifestyle and brain imaging, peripheral markers. Bottom right, structural equation model specifying the potential mechanisms underlying the associations between lifestyle, PRS, brain structure, immunometabolic function and depression.



## 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  
Freesurfer v6.0 was used to process the imaging data.  
PLINK 2.0 and PRSice version 2 ([www.PRSice.info](http://www.PRSice.info)) were used to perform genome-wide association analysis and calculate the polygenic risk score respectively.  
R version 4.2.1 package:  
survival 3.4-0 was used to perform Cox proportional hazard regression model,  
TwoSampleMR 0.5.6 and RadialMR 1.0 was used to perform Mendelian Randomization analysis,  
lavaan 0.8 was used to perform structural equation model.  
Scripts used to perform the analyses are available at [https://github.com/yjzhao1004/lifestyle\\_depression](https://github.com/yjzhao1004/lifestyle_depression)

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](#)

Access to individual-level UK Biobank data, both phenotypic, neuroimaging and genotype, is available to bona fide researchers through application on the UK Biobank website (<https://www.ukbiobank.ac.uk>). Additional information about registration for access to the data is available at <http://www.ukbiobank.ac.uk/register-apply/>. Use of UK Biobank data was performed under application number 19542. Summary statistics from previous GWAS of depression which were used in this study are publicly available through the Psychiatric Genomics Consortium (PGC – Psychiatric Genomics Consortium ([unc.edu](http://unc.edu))) and can be downloaded from <https://pgc.unc.edu/for-researchers/download-results/>.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Both male and female subjects from UK biobank study were included in the study. Summary statistics on sex distributions were reported in Table 1. All statistic models were adjusted for sex.

Population characteristics

We utilized a total of 502,409 participants from UK Biobank cohort to perform survival analysis, correlation analysis, and structural equation model. For survival analyses, we retrieved 287,282 individuals (mean age 57.52, with 50.70% female) who participated in the assessment of lifestyle factors at baseline (2006-2010) and had linkage to clinical diagnosis records. The number of participants in healthy or unhealthy lifestyle in each factor was listed in Table 1. The mean score of 7 lifestyle was 4.75 (s.d., 1.36), of which 1.25% were categorized as following an unfavorable lifestyle (scores ranging from 0-1), 38.90% followed an intermediate lifestyle (scores ranging from 2-4), and 59.85% followed a favorable lifestyle (scores ranging from 5-7). Of the 287,282 participants, 12,916 of them had an onset of depression during a median follow-up of 9.01 years. Demographic characteristics of the participants and distribution of lifestyle factors were provided in Table 1. The polygenic risk scores for depression from 197,344 participants (mean age 57.81, with 49.82% female) were normally distributed (see Supplementary Fig. 5) and categorized into three levels, low risk (25.09%), intermediate risk (49.97%), high risk (24.94%). Brain structural imaging data were collected in 2014, containing T1-weighted structural MRI from 32,839 participants used in the correlation analysis. Blood chemistry and cell count indicators from about 480,000 participants and NMR metabolic biomarkers from about 120,000 participants were collected at baseline. The sample size of each biomarker in correlation analyses were presented in Supplementary table 18 and 19. We additionally used data from a total of 448,849 participants with completed Patient Health Questionnaire-4 (2006-2010) score. Combining the data of lifestyle score, PRS, immunometabolic markers, brain structural imaging and depression, the final analysis included 18,244 participants (mean age 55.20, with 48.79% female) to estimate the structural equation model. Fig. 1 showed the research guideline of the study. Supplementary Fig.1 and Supplementary table 4 presented the sample size utilized in each analysis.

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. Previous investigation showed UK biobank subject to a healthy sample bias.

Ethics oversight

The UKB has received approval from the National Information Governance Board for Health and Social Care and the National Health Service North West Centre for Research Ethics Committee (Ref: 11/NW/0382). All participants provided informed consent through electronic signature at the baseline assessment.

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](https://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

No statistical methods were used to predetermine sample sizes and all currently available sample in the UK Biobank were used including 394,053 subjects with lifestyle factor data, 287,282 subjects with depression diagnosis record, 32,839 subjects with imaging data, about 480,000 subjects with blood chemistry data, about 120,000 subjects with metabolic markers data and 337,151 subjects with genotype data. The sample sizes were provided in the method section.

Data exclusions	For cox proportional hazard regression model, we excluded patients with missing lifestyle factor data, baseline depression, unavailable baseline data, missing follow-up data and those with self-report depression. Flow chart of the sample size selection in survival analyses and structural equation model were presented in Supplementary Fig.1. For genome-wide association analyses on lifestyle score, we excluded single-nucleotide polymorphisms (SNPs) with call rates < 95%, minor allele frequency < 0.1%, deviation from the Hardy–Weinberg equilibrium with $p < 1E-10$ and selected subjects that were estimated to have recent British ancestry and have no more than ten putative third-degree relatives in the kinship table, consistent with the previous study.
Replication	All available data were used to maximize statistical power of the analysis therefore we did not repeat the analysis.
Randomization	All models were adjusted for age at baseline, sex, Townsend deprivation index, body mass index (BMI) and education levels. Association analyses involving neuroimaging data were additionally adjusted neuroimaging scanning sites and estimated total intracranial volume. Association analyses involving genetic data, PRS components (first 20 principal components of ancestry) were further added as covariates. Full details of covariates utilized in the statistical analyses see Supplementary Fig.2.
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

### Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging

## Magnetic resonance imaging

### Experimental design

Design type	Structural MRI
Design specifications	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. T1-weighted structural image was acquired using straight sagittal orientation for 5 minutes.
Behavioral performance measures	Lifestyle assessments were first administered via touchscreen interface in the UK Biobank assessment center at the baseline visit (2006-2010). Seven aspects of lifestyle factor assessments including alcohol consumption, diet, physical activity, sleep duration, smoke, sedentary behavior and social connection were utilized to calculate lifestyle score in the current study (n = 394,053). Measurement of depressive symptoms via 4-item Patient Health Questionnaire-4 (PHQ-4) was first assessed in the UK Biobank assessment center (2006-2010, n=448,849).

### Acquisition

Imaging type(s)	T1-weighted structural imaging
Field strength	3T
Sequence & imaging parameters	The EPI-based acquisitions utilize simultaneous multi-slice (multiband) acceleration. Biobank uses pulse sequences and reconstruction code from the Center for Magnetic Resonance Research (CMRR), University of Minnesota <a href="https://www.cmrr.umn.edu/multiband">https://www.cmrr.umn.edu/multiband</a> . The resolution is 1x1x1 mm and 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: <a href="https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf">https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf</a>
Area of acquisition	Whole brain
Diffusion MRI	<input type="checkbox"/> Used <input checked="" type="checkbox"/> Not used

## Preprocessing

Preprocessing software	T1 images were processed with FreeSurfer; surface templates were used to extract imaging-derived phenotypes referred to as atlas regions' surface volume <sup>76</sup> . Subcortical regions were extracted via FreeSurfer's aseg tool. FreeSurfer aparc (Category ID 192) and aseg (Category ID 190) atlas corresponding to 68 cortical regions <sup>76</sup> and 14 subcortical regions were used in this study. We used the Qoala-T approach to check FreeSurfer outputs and excluded those failed to pass quality control from the FreeSurfer imaging-derived phenotypes.
Normalization	see above
Normalization template	fsaverage
Noise and artifact removal	see above
Volume censoring	see above

## Statistical modeling & inference

Model type and settings	1) Cox proportional hazard regression model was used to examine the association of lifestyle with risk of depression. 2) Mendelian randomization (MR) analysis was used to examine the causal relationship between lifestyle and depression. 3) Structural equation model was used to analyze the association between lifestyle, polygenic risk score, brain structure, immunometabolic function and depression.
Effect(s) tested	For cox proportional hazard regression model, MR analysis and structural equation model, Wald tests were utilized to derive the two-sided p value. Additionally, for survival analysis, the proportional hazards assumption was assessed using the Schoenfeld residuals method and satisfied. For MR analysis, we used inverse variance weighted method, simple median method, weighted median method and weighted mode method to estimate the causal effect.
Specify type of analysis:	<input checked="" type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input type="checkbox"/> Both
Statistic type for inference (See <a href="#">Eklund et al. 2016</a> )	voxel-wise association, voxel-wise Bonferroni correction
Correction	Bonferroni

## Models & analysis

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input checked="" type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input checked="" type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis