Plasma proteomics reveal heterogeneous subtypes of depression linked to inflammation and aging

Authors: Zhengxu Lian[1,2#], Lena Palaniyappan[3,4,5,6#] , Zhaowen Liu[7#], Nanyu Kuang[8], Yu Liu[1,2],Yuchao Jiang[9], Barbara J Sahakian[1,10,11], Trevor W. Robbins[1,10,11], Jianfeng Feng[1,2,12,13,14,15], Benjamin Becker[16,17] , Wei Cheng[1,2], Xinran Wu[18*], Jie Zhang[1,2*], Ed Bullmore[10]

1 Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai 200433, PR China

2 Key Laboratory of Computational Neuroscience and Brain Inspired Intelligence, Fudan University, Ministry of Education, PR China

3 Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, QC, Canada

4 Department of Psychiatry, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

5 Robarts Research Institute, University of Western Ontario, London, ON, Canada

6 Department of Medical Biophysica, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

7 School of Computer Science of Northwestern Polytechnical University, Xi'an, Shanxi, PR China

8 Neuroimaging Research Branch, National Institute on Drug Abuse, National Institutes of Health, Baltimore, Maryland, USA.

9 School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen, PR China

10 Department of Psychiatry, University of Cambridge, Cambridge, UK.

11 Department of Psychology and Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK.

12 Shanghai Center for Mathematical Sciences, Shanghai, 200433, PR China

13 Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK

14 Collaborative Innovation Center for Brain Science, Fudan University, Shanghai, 200433, PR China

15 Fudan ISTBI—ZJNU Algorithm Centre for Brain-inspired Intelligence, Zhejiang Normal University, Jinhua, PR China

16 The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, PR China

17 Department of Psychology, The University of Hong Kong, Hong Kong, PR China

18 School of Psychology, Shanghai Jiao Tong University, Shanghai, PR China

# These authors contributed equally to the work.

* Correspondence should be addressed to:

Jie Zhang, zhangjie80@fudan.edu.cn, Tel: (+86)15026570528, Institute of Science and Technology for Brain Inspired Intelligence, Fudan University, Shanghai, PR China

Xinran Wu, wxr1173199759@gmail.com, School of Psychology, Shanghai Jiao Tong University, Shanghai, PR China

Abstract：

The clinical heterogeneity of depression has defied biological classification, limiting personalized treatment. Previous neuroimaging- or symptom-based subtyping of depression failed to clarify the underlying pathoetiology, while plasma proteins which integrates signals from multiple organ systems, offers a promising way to define biologically grounded subtypes. Using plasma proteomics from 2,127 incident depression cases in a cohort of 53000 individuals, we identified three biologically distinct subtypes differing in inflammation, aging, and metabolic profiles. The most prevalent subtype (‘inflammation/ageing’) was characterized by aging-related inflammation, poorest prognosis with hippocampal atrophy and highest suicide risk, mediated by age-related amygdalar atrophy; this subtype had highest anhedonia burden. A distinct ‘inflammation/energy dysregulation’ group had metabolic pathway enrichment with high inflammation and lifestyle risk factors (smoking) but no ageing trend, predominantly physical/psychomotor symptoms and decreased thalamic volume. In contrast, the ‘inflammation-resilient’ group had the lowest inflammatory proteomic

loading, lowest depression severity, more resilient lifestyle and increased hippocampal volume. These proteomic signatures, detectable years before symptom onset, enable risk stratification and suggest subtype-specific targeted physical and lifestyle interventions.

Introduction：

Depression is one of the leading causes of disability worldwide, characterized by high prevalence, frequent relapse, and diverse clinical manifestations.[1] The substantial heterogeneity of Major Depressive Disorder (MDD) in both clinical phenotypes and biological mechanisms severely impedes the identification of reliable diagnostic biomarkers and the development of targeted therapeutic strategies.[2,3]

Prior depression subtyping, whether symptom-based[4], treatment-response-based[5,6], or neuroimaging-based[7,8], has failed to illuminate underlying pathophysiology. Brain-centric models alone cannot capture depression's systemic, whole-body nature, particularly immune and metabolic dysregulation.

The two promising systemic leads for depression are inflammatory and metabolic pathways associated with depression; the former is often linked to immune system[9], and the latter to endocrine disorders (e.g., diabetes) and cardiovascular risk.[10] There has been early success for mechanism-based stratifications by incorporating these pathways [11] (e.g., molecular subtypes based on the differential expression of oxidative

stress patterns[12]) . Plasma proteins, as functional endpoints of gene-environment interactions, integrate immune, neural, and metabolic signals, offering a promising window for decisive progress.

Prior proteomic studies have linked individual plasma proteins to depression risk[13] at a generic level, without addressing the disorder's hallmark heterogeneity. Recognizing MDD as a systemic pathology involving immune activation and metabolic imbalance is necessary for developing accurate, mechanism-based stratifications. To date, attempts to address heterogeneity via proteomics has bene hampered by small sample sizes, cross-sectional designs and pre-selected protein panels. For example, using 171 proteins to distinguish atypical (n=128) from melancholic depression (n=231) and healthy controls (n=414)[14]; comparing bipolar disorder (n=6), major depressive disorder (n=6), persistent depressive disorder (n=3), and healthy controls (n=3)[15].

Overcoming these limitations necessitates large samples with an age range reflective of incident depression to capture 'process' variables across the lifespan, alongside a longitudinal design to determine directionality and comprehensive data-driven determination of proteome-based subtypes.

To address these gaps, we leveraged longitudinal plasma proteomics in 53,026 middle-aged/older adults to identify latent subtypes of depression, characterize their clinical trajectories, and uncover mechanistic pathways amenable to early intervention. As illustrated in Fig. 1, based on data-driven selection of 215 proteins associated with

2,127 incident-cases of depression (Fig. 1b), we identified three latent proteomic dimensions and corresponding patient subgroups. We systemetically examined how these proteomic dimensions relate to brain structure, molecular function, organ health, lifestyle, and prognosis trajectory, and undertook survival analysis to study subgroups. Pathway enrichment and mediation analyses uncovered potential mechanistic pathways that can be targeted for interventions.

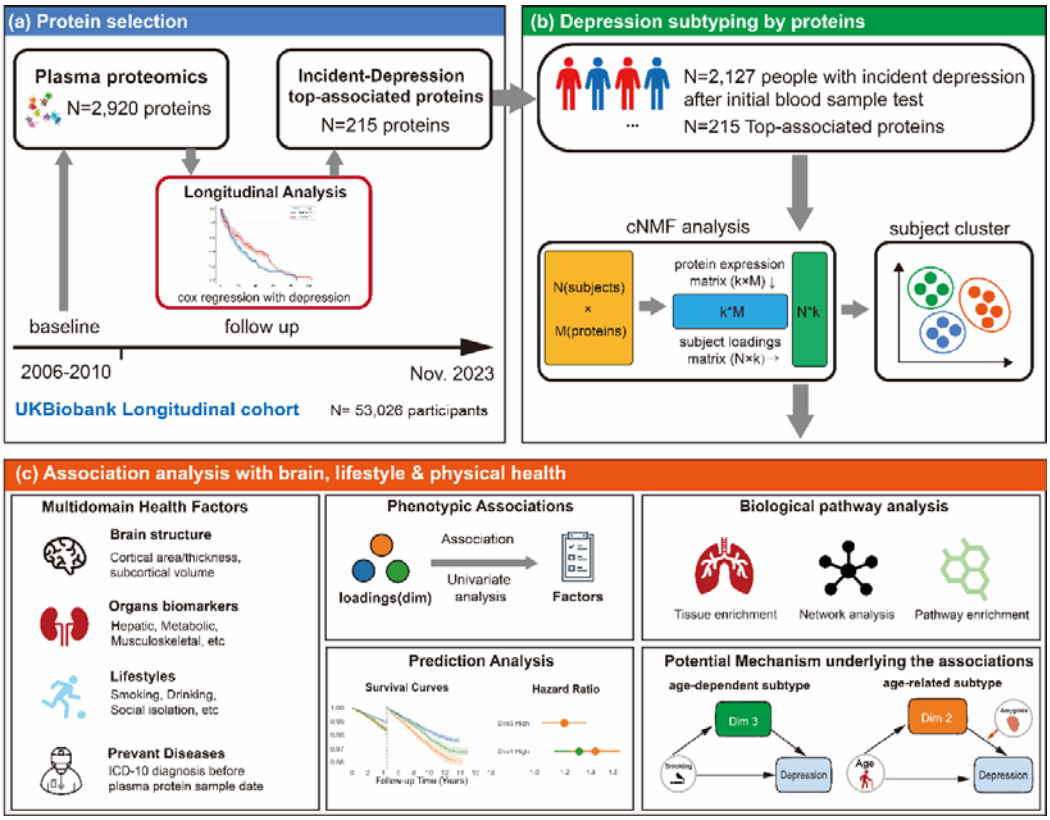

Figure 1 | Schematic overview of the study design.

(a) Protein selection workflow. Using the UK Biobank longitudinal cohort (N=53,026), we screened 2,920 baseline plasma proteins using Cox proportional hazards regression models and identified 215 proteins significantly associated with the risk of incident depression.

(b) Depression subtyping by selected proteins. Among 2,127 individuals with incident depression, consensus Non-negative Matrix Factorization (cNMF) was applied to the expression matrix of the 215 selected proteins. This analysis extracted latent molecular dimensions (co-expressed protein patterns). Individuals were subsequently classified into distinct depression subtypes based on their maximal loading on these dimensions.

(c) Association analysis with brain, lifestyle & physical health. We anlyzed the dimensions/subtypes through four analyses: (i) Biological pathway analysis: we performed tissue enrichment, network analysis, and pathway enrichment on proteins with high weights in each dimension to decode their biological functions. (ii) Phenotypic Associations: we correlated continuous dimension loadings with multidomain health factors, including brain structure, organ biomarkers, lifestyles, and prevalent diseases. (iii) Prediction Analysis: we evaluated the clinical prognostic value by generating survival curves and calculating hazard ratios for the identified subtypes. (iv) Potential Mechanism: we conducted mediation analyses using dimension loadings to explore how these molecular dimensions mediate the pathways between upstream risk factors (e.g., aging, smoking) and depression onset.

Results

Our goal was to capture the latent molecular dimensions from baseline plasma proteins that were associated with incident depression, and identify mechanistically distinct subtypes. To this end, we first applied consensus non-negative matrix factorization (cNMF)[16] to 215 significantly associated proteins (at baseline, selected via Cox

regression) in 2,127 individuals who developed depression afterwards. This yielded three latent proteomic dimensions that define three corresponding patients subtypes. Each proteinomic dimension offers a dual readout (see Methods and cNMF analysis in Fig. 1b): at protein level, it is defined by the protein expression matrix, where high-contributing proteins carry important information about the underlying biological/physiological function of each dimension; at the individual level, it yields a quantitative 'dimension score' derived from the subject loadings matrix, i.e., a linear combination of the protein expression driven primarily by the top-contributing proteins of that dimension. For a given individual, the highest dimension score (among all 3 dimentions) defines the subtype to which that individual is categorized, consistent with prior subtyping work using CNMF[17] . We determined the optimal number of dimensions based on comprehensive stability analysis/error analysis (inter-model reproducibility analysis, consensus stability assessment, and stability-error trade-off evaluation), which identified a three-component solution (k=3) as the most robust model (detailed in methods and supplementary eMethods and eFigure 2). We then projected these latent dimensions onto all participants (N=53,026) with baseline proteomics (see Methods) to facilitate broader association analyses. These dimension scores were leveraged to examine relationships with depressive symptoms (PHQ-9), brain structures, organ and lifestyle phenotypes, and to derive longitudinal survival and mechanistic mediation models. Unless otherwise specified, this full cohort constituted the basis for all subsequent analyses involving phenotypic associations, mediation models, and survival outcomes.

We visualized subjects with incident depression(N=2,127) within a coordinate space defined by systemic inflammation (proxied by log-transformed CRP level) and age of individuals, see Fig. 2a, where depression patients are color-coded by their subtype. This revealed a clear physiological separation among 3 subtypes. Specifically, the Dimension 1–dominant subtype(Subtype 1) and Dimension 3–dominant subtype (Subtype 3) were age-independent (right insect of Fig.2a), distinguished by low versus high inflammation levels, respectively. In sharp contrast, the Dimension 2–dominant subtype (Subtype 2) was characterized by an age-related high-inflammation profile. This age-dependency was quantitatively corroborated by generalized additive model (GAM) fits, where Dimension-2 scores demonstrated a significant increase with age, whereas Dimensions 1/3 showed no such age-related trends. Top-contributing proteins (top 33% by weight) showed minimal overlap across dimensions (Fig. 2b), confirming distinct molecular architectures. Longitudinal symptom profiles diverged markedly: Subtype 2 patients exhibited higher anhedonia and suicidal ideation, while Subtype 3 presented with fatigue and sleep disturbances (Fig. 2c). This dissociation between traditionally co-occurring symptoms suggests distinct mechanistic pathways.

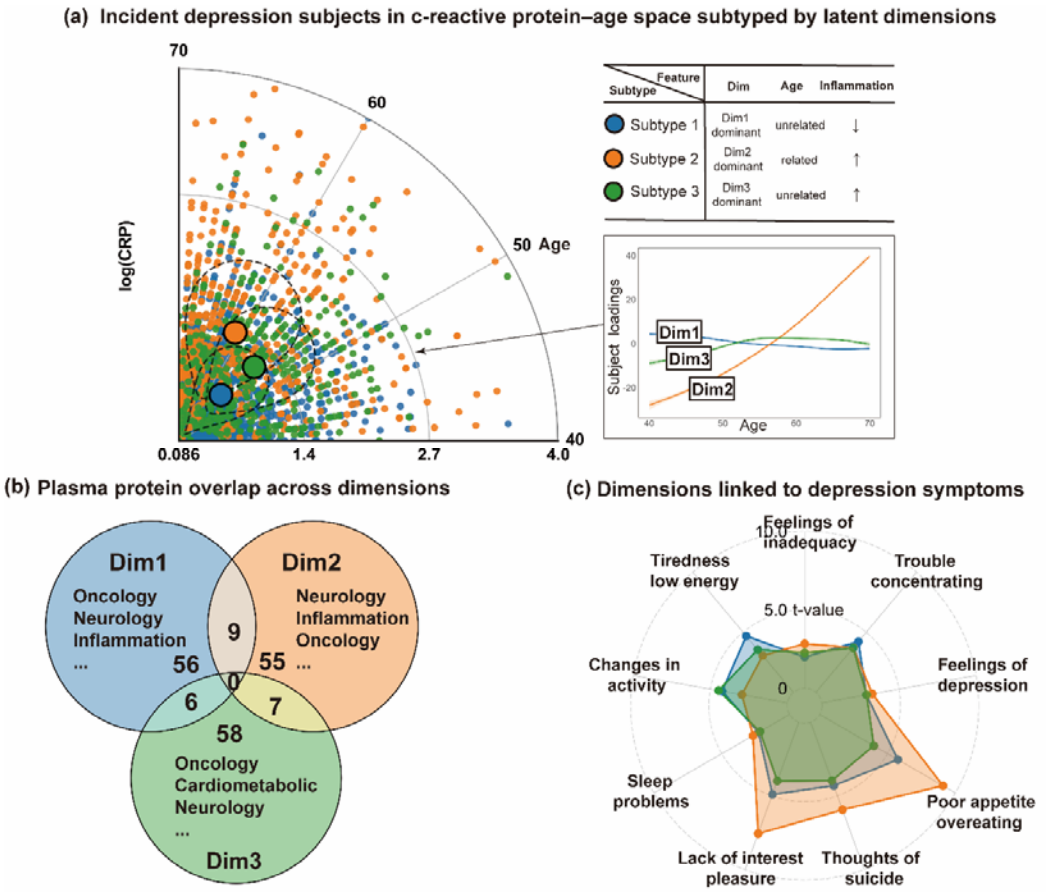

Figure 2 | Latent plasma protein dimensions identified by cNMF and their phenotypic characteristics.

(a) Application of cNMF to 215 plasma proteins significantly associated with incident depression revealed three latent dimensions. Individuals were assigned to a subtype based on their dominant dimension loading. Subjects with incident depression (N=2,127) are shown in a two-dimensional space defined by inflammation level (C-reactive protein) and age. The inset panel displays generalized additive model fits of continuous dimension scores in individuals with incident depression across age, demonstrating a significant age-related increase in Dimension-2 score, whereas Dimensions 1 and 3 showed no age-related trends. (b) Venn diagram of

top-contributing proteins in each dimension, illustrating high molecular specificity with minimal overlap; only a small set of proteins was shared across dimensions. Protein categories of the top three functional groups are labeled.

(c) Radar plots illustrates the strength of association (absolute t-values) between PHQ-9 symptoms and continuous dimension scores in the full cohort (N = 53,026). Dimension-2 score (the defining feature of the corresponding subtype) was specifically associated with anhedonia, appetite loss, and elevated suicidal ideation. In contrast, Dimensions 1 and 3 scores tracked more strongly with somatic symptoms, including reduced energy and psychomotor changes.

We next examined the neuroanatomical correlates of each proteomic dimension to identify the core neuroimaging features of each subtype. We analyzed their associations with subcortical volumes, cortical surface area and cortical thickness, respectively. For subcortical structures, higher Dimension-1 score showed significant positive associations with bilateral hippocampal volume, globus pallidus, ventral diencephalon, and bilateral thalamus, suggesting a potential neuroprotective structural profile. In contrast, Dimension-2 score tracked with reduced volumes in the hippocampus, globus pallidus, caudate, and superior frontal gyrus, reflecting a prototypical age-related atrophy pattern; notably, the caudate is a key region involved in reward processing. Dimension-3 score exhibited significant negative associations with bilateral thalamus, right hippocampus, and ventral diencephalon, possibly indicating structural damage linked to cumulative health risks. Regarding cortical morphology, Dimension-1 score

was associated with increased surface area of the left middle temporal gyrus, whereas Dimension-2 score tracked with reduced surface area in the left superior frontal gyrus. Dimension-3 score was linked to cortical thinning in the left superior frontal gyrus, potentially reflecting degeneration in regions relevant to perceptual processing and cognitive control.

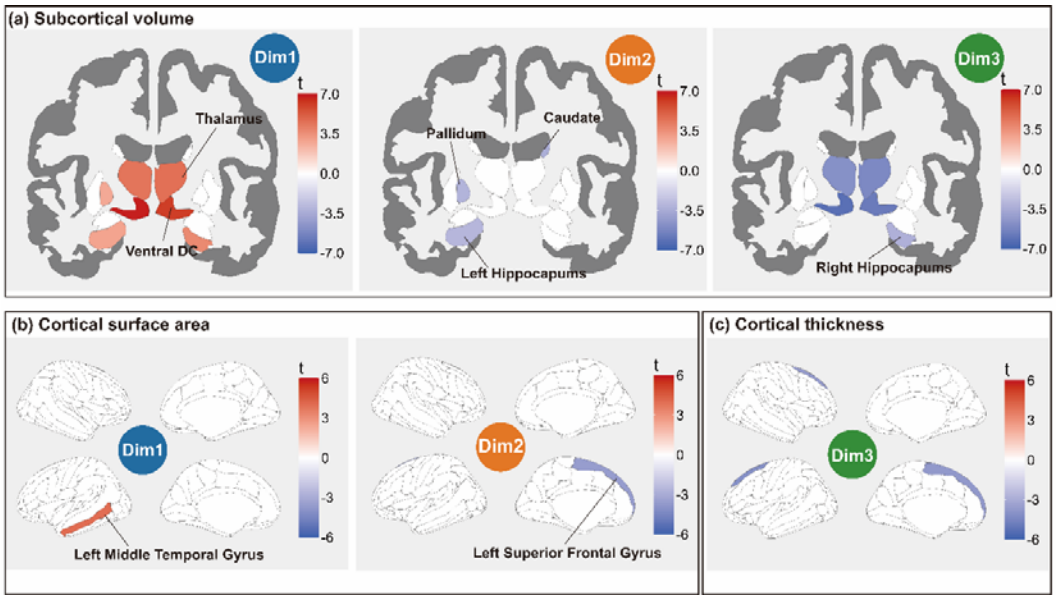

Figure 3 | Whole-brain associations between proteinomic dimension scores and brain structure.

(a) *t*-value maps showing associations between each protein dimension score and subcortical volumes. Red indicates positive associations and blue indicates negative associations; only regions surviving FDR correction are displayed. (b) *t*-value maps showing associations with cortical surface area, with FDR correction. (c) *t*-value maps showing associations with cortical thickness. Only regions surviving FDR correction

are displayed.

Molecular profiling revealed mechanistically distinct pathways (Fig. 4a). Subtypes 1 and 2 proteins enriched for immune signaling, while Subtype 3 proteins clustered in metabolic processes. Despite different WGCNA modules, Subtypes 1 and 3 showed tight protein-protein interactions (Supplementary eFigure 1), suggesting counter-regulatory roles in shared inflammatory pathways—consistent with their opposite inflammation profiles. Although the proteins specific to Dimension 1 and Dimension 3 fell into different WGCNA modules, the PPI analysis showed they are closely connected (Supplementary eFigure 1). This indicates that these two dimensions likely operate as counter-forces within shared biological processes. This aligns with the low inflamation level for Dimension-1 dominant subtype(Subtype 1) vs. high inflamation level for Dimension-3 dominant subtype (Subtype 3). See also the directionally opposite correlation patterns between Dimensions 1 and 3 and physiological health in following parts.

To trace the physiological origins of each subtype, we examined the tissue-specific expression patterns of their top-contributing proteins. This organ-level enrichment analysis highlighted a clear divergence in biological sources across 3 subtypes (Fig. 4c). The protein associated with Dimension 1–dominant subtype was predominantly expressed in tissues such as the colon, lung, and coronary arteries. In contrast, the core proteins defining the Dimension 2–dominant subtype showed stronger

expression in skeletal muscle, lung, and visceral adipose tissue. Distinctly, the Dimension 3–dominant subtype's profile was enriched in the renal medulla, pancreas, and esophageal mucosa. These different organ-expression profiles suggest that each subtype is driven by specific, tissue-originating pathways that influence depression risk.

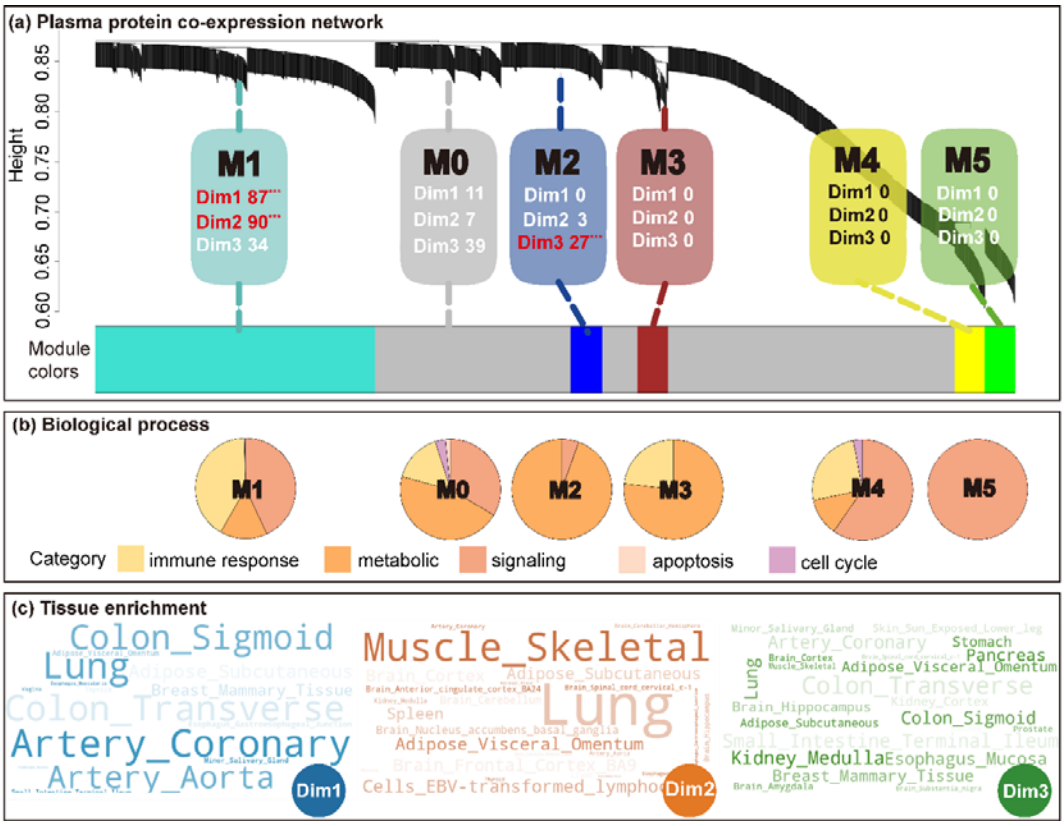

Figure 4 | Functional pathways and organ-level expression profiles of protein dimension scores.

(a) Weighted gene co-expression network analysis (WGCNA) performed on the top-contributing proteins of each proteomic imension. The number following each dimension label indicates the percentage of the core proteins from that dimension that was assigned to the corresponding module. Red text marked with *** denote

significant enrichment ($P < 0.001$, hypergeometric test). (b) Pie chart showing the distribution of biological processes enriched within each module. (c) Tissue enrichment analysis (based on GTEx data) for the top-contributing proteins.

We next sought to characterize the observable lifestyle profiles and organ-level health risks associated with each subtype. We first analyzed the associations between continuous dimension scores and phsiological markers across seven major organ systems, as well as lifestyle behaviors (Fig. 5a–f). In the resulting heatmaps, phenotypic categories are ranked by their mean absolute effect size ($|\beta|$), with top-ranking categories highlighting the specific physiological and lifestyle profiles that characterize each dimension-dominant subtype.

Physiological profiles diverged sharply. Subtype 1 showed resilience across hepatic, immune, and respiratory systems (higher FEV1, lower CRP). Subtype 3 exhibited the inverse pattern with hepatic dysfunction, elevated inflammation, and metabolic dysregulation. Subtype 2 uniquely tracked with aging-related decline: impaired kidney function (elevated cystatin C), sarcopenia (reduced peak expiratory flow), and chronic inflammation (Fig. 5a-c).

We also delineated the specific behavioral profiles for each subtype. Smoking emerged as the dominant behavioral correlate across all subtypes, but with distinct co-occurring patterns. Subtype 1 patients maintained protective behaviors: social engagement, physical activity (greater walking speed), and healthy diet (fresh fruit consumption). Subtype 2 showed physical inactivity, poor diet and social withdrawal. Subtype 3

demonstrated the most adverse lifestyle profile: insomnia, sedentary behavior, and heavy alcohol use (Fig. 5d-f; The ranked color bar categories are sorted by effect size).

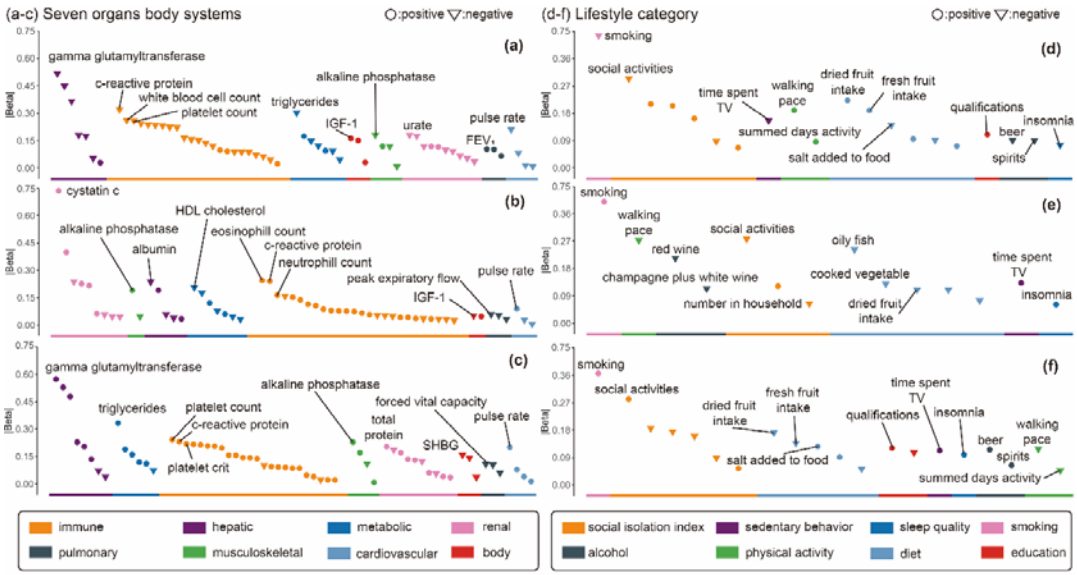

Figure 5 | Association patterns of plasma protein dimensions with organ system and lifestyle phenotypes.

(a–c) Associations of continuous Dimensions 1–3 scores with phenotypes across seven major organ systems. The order of categories along the color bar reflects the mean |β| values of FDR-corrected traits within each category, arranged in descending order from left to right. positions towards the left indicate stronger associations with the corresponding dimension. Only traits surviving FDR correction are displayed, with circles denoting positive associations and inverted triangles denoting negative associations.

(d–f) Associations of continuous Dimensions 1–3 scores with lifestyle indicators. Plot elements and layout are consistent with panels (a–c).

To assess the differences in the long-term prognostic trajectories of these three subtypes, we performed Kaplan–Meier survival analyses using all participants (with baseline plasma proteomic data) stratified by the three dimension-dominant subtypes (Fig. 6a left). To mitigate potential bias due to the strong correlation between Dimension 2 and age, we applied a landmark analysis with a 4.5-year cutoff, separately estimating risks in the early (0–4.5 years) and late (>4.5 years) follow-up periods. After 4.5 years, individuals in Subtype 1 exhibited significantly lower depression risk compared with those in Subtype 2 and 3, while the Dimension 2–dominant subtype showed higher risk than the Dimension 3–dominant subtype (Fig. 6a right). These findings suggest that with advancing age, the inflammation- and aging-related pathways represented by Dimension 2 may confer stronger vulnerability to depression than the risk-laden Dimension 3.

Given that Dimensions 1 and 3 occupy opposite ends of the age-independent inflammation axis, with Dimension 3 conferring unfavorable long-term prognosis, we next probed its mechanistic link to the onset of depression. Smoking's effect on depression risk was mediated by Subtype 3 proteomic signature (indirect effect β=0.27, 95% CI [0.06, 0.54]; Fig. 6b), identifying a modifiable pathway: smoking → metabolic inflammation → depression. This suggests smoking cessation interventions may reduce depression risk by lowering metabolic inflammatory burden. In contrast, Dimension 2 followed a distinct mechanistic route. Aging increased depression risk through Subtype 2 (aging → inflammation), but only in individuals with smaller amygdala volumes (moderation index = -0.011, 95% CI [-0.022, -0.002]; Fig. 6c). This identifies amygdala

volume as a neural vulnerability factor: in aging individuals with amygdala atrophy, inflammation more readily translates to suicidality..

Together, these findings delineate two distinct mechanistic pathways: a neuroanatomically moderated aging pathway through Dimension 2, and a lifestyle-mediated route through Dimension 3.

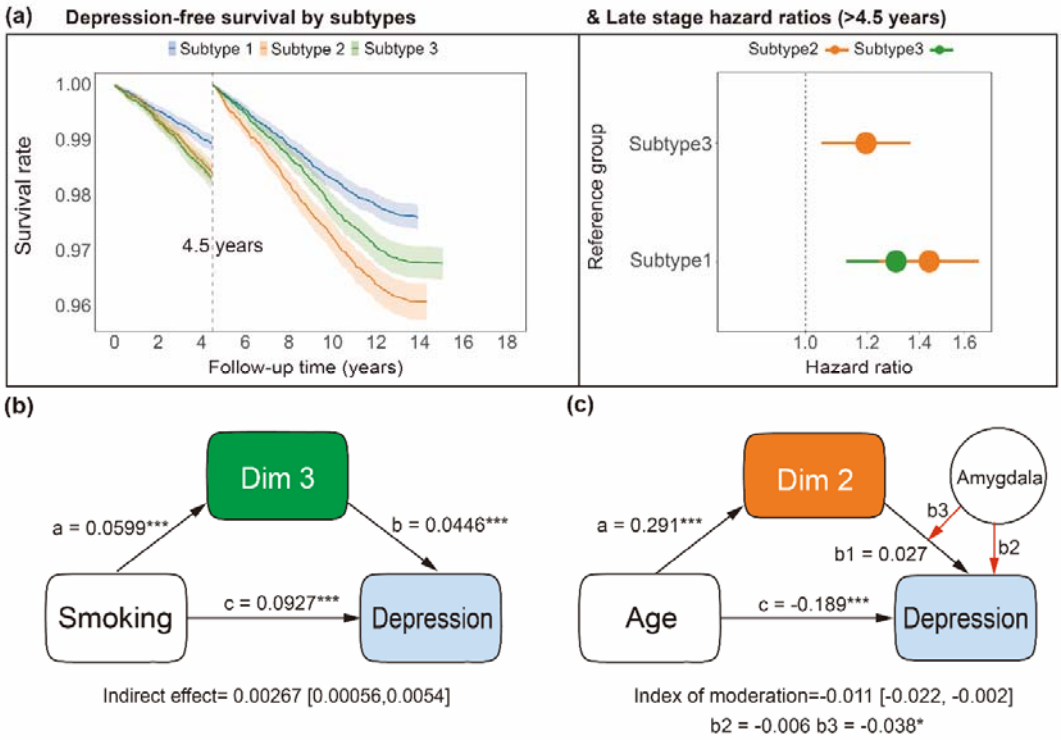

Figure 6 | Survival prediction and mechanistic models of protein dimension score in depression risk.

(a) Left panel: Kaplan–Meier survival curves stratified by subtype in the full cohort (shaded areas: 95% CIs), with a 4.5-year landmark cutoff distinguishing early versus late trends. Right panel: Hazard ratios (HRs) derived from Cox proportional hazards models after 4.5 years. As Subtype 1 exhibited the lowest risk, it served as the reference

group. Accordingly, the plot displays specific pairwise comparisons of Subtypes 2 and 3 against Subtype 1, as well as the contrast between Subtypes 2 and 3. Dots represent point estimates and horizontal lines indicate 95% CIs.

(b) Mediation model of smoking → Dimension 3 → depression, demonstrating that continuous Dimension-3 score partially mediates the pathway from smoking to depression risk. Standardized regression coefficients are shown on the paths; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

(c) Moderated mediation model of age → Dimension 2 → depression, with the pathway moderated by amygdala volume. In the model, b□ denotes the main effect of continuous Dimension-2 score on depression, b□ denotes the main effect of amygdala volume on depression, and b□ denotes the interaction effect of Dimension 2 × amygdala volume. Standardized regression coefficients are shown on the paths; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Discussion:

In this study, we demonstrate that plasma proteomics can stratify depression into biologically distinct subtypes years before symptom onset. Unlike prior neuroimaging or symptom-based approaches, these proteomic subtypes map to mechanistically separable pathways: aging-related inflammation (Subtype 2), metabolic inflammation (Subtype 3), and a resilient, low-inflammation profile (Subtype 1), each with distinct brain signatures, clinical courses, and intervention targets. This represents a shift from

descriptive to mechanistic depression classification. Major conceptual advance of this work is the observation that inflammation-related mechanisms is not restricted to limited subgroup of MDD; instead, inflammatory pathways play a major pathoplastic infuence on most incident cases of depression between the ages of 40 to 70.

While inflammation[13,18–22] and biological agin[23] are established depression risk factors, we reveal their convergence as the critical driver of worst outcomes. Subtype 2 uniquely couples high inflammation with aging, showing highest suicide risk and poorest prognosis, exceeding even the high-inflammation Subtype 3. This suggests aging amplifies inflammatory vulnerability through specific mechanisms (e.g., amygdala atrophy), rather than additive effects.

To elucidate how these divergent inflammatory and aging-related trajectories translate into observable clinical and neurobiological phenotypes, we integrated PHQ-9 symptom profiles, neuroimaging data, and molecular pathway analyses for each subtype. A key clinical finding: anhedonia and fatigue, traditionally viewed as inseparable[24], dissociated into distinct subtypes. Subtype 2's anhedonia mapped to reward circuit atrophy (caudate, pallidum) [25–27], while Subtype 3's fatigue aligned with immunometabolic dysfunction. This has treatment implications: dopaminergic or behavioral activation for Subtype 2's reward deficit versus anti-inflammatory or metabolic interventions for Subtype 3's energy dysregulation. This clinical profile aligns with established models of immunometabolic depression, where systemic

inflammation disrupts energy homeostasis to induce fatigue as a core symptom[28]. Substantiating this result, we observed that Dimension 3 is enriched for proteins involved in energy metabolism and immune signaling (Fig. 4), defining a specific 'immuno-metabolic' pathway.

To translate these neurobiological insights into actionable clinical strategies, we traced the upstream drivers of each subtype across lifestyle behaviors and organ system health factors. The Dimension 1–dominant (age-independent, low inflammation) subtype was linked to healthier lifestyle behaviors, larger subcortical volumes, and protective physiological profiles, suggesting greater resilience that may favor non-pharmacological approaches such as cognitive therapies[29]. In contrast, the Dimension 3–dominant subtype was distinguished by a prominent clustering of adverse lifestyle behaviors. Smoking, social isolation, and poor diet were its top correlates—and mediation analysis confirmed that Dimension 3 partially mediated the pathway between smoking and depression, underscoring the importance of lifestyle modification. Consistent with its immuno-metabolic nature, Dimension 3 also mapped strongly to hepatic, metabolic, and immune systems.

By comparison, the Dimension 2–dominant subtype displayed stronger ties to aging-related risk factors, including dysfunction in renal and musculoskeletal systems. These phenotype associations, together with evidence for bidirectional links between renal dysfunction and depression[30,31], highlight avenues for systemic health interventions. Finally, elucidating the brain structural vulnerability flagged in the

previous section, we found that the effect of age on depression, mediated by Dimension 2, was significantly moderated by amygdala volume. Given that this Dimension was tightly linked to suicidal ideation, and that amygdala volume has been repeatedly implicated in suicidality in depression[32–34], this pathway suggests novel targets for preventing suicidality in late-life depression through protein-based and circuit-specific interventions.

Taken together, our study provides the first longitudinal evidence from a large-scale cohort that plasma proteomics can delineate biologically meaningful and clinical relavant subtypes of depression. Unlike prior work that conceptualized inflammation as a single continuum, our findings reveal distinct subtypes separated along the axes of age and inflammation, with protein signals detectable years before onset. Integrative evidence across symptoms, molecular pathways, brain structure, lifestyle, and organ systems demonstrates that the Dimension 2– and Dimension 3–dominant subtypes reflect separable but clinically meaningful mechanisms: the former marked by aging-related emotional vulnerability and suicidality, and the latter by energy–inflammation pathways linked to fatigue. The interaction between Dimension 2 and amygdala volume further highlights a novel neural mechanism linking biological aging to suicidality in late-life depression. Overall, our findings establish a coherent evidence chain from plasma proteins to symptoms, neurobiology, and lifestyle, deepening our understanding of depression heterogeneity and providing new opportunities for early identification and precision intervention.

We acknowledge several limitations. Our UK Biobank sample comprises middle-aged/older adults of European ancestry, requiring validation in younger and more diverse populations. While we integrated proteomics with imaging and lifestyle data, multi-omics approaches (transcriptomics, metabolomics) would refine mechanistic understanding. Finally, experimental validation of subtype-specific interventions is needed to confirm therapeutic implications.

In conclusion, by revealing mechanistically distinct depression subtypes through plasma proteomics, we enable biology-based risk stratification years before symptom onset. Subtype-specific intervention strategies targeting aging pathways and amygdala vulnerability in Subtype 2, metabolic dysfunction in Subtype 3, and leveraging resilience factors in Subtype 1 offer a roadmap toward more system-informed management of MDD. The convergence of inflammation and aging as the primary driver of suicide risk identifies an urgent target for preventive intervention in aging populations.

Methods

Participants

This study utilized data from the UK Biobank cohort. Ethical approval was obtained from the Human Biology Research Ethics Committee at the University of Cambridge (Cambridge, UK). All participants provided informed consent (https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=200). The application ID for this study is 19542.

The plasma protein data were sourced from the UK Biobank Pharma Proteomics Project consortium. Blood samples were collected in 9 mL EDTA vacutainers and processed into 850 μL aliquots containing EDTA plasma, buffy coat, and red blood cells. Plasma samples were stored at −80 °C before being transported on dry ice to Olink Analysis Service in Sweden. From April 2021 to January 2022, a proximity extension assay combined with Next-Generation Sequencing was employed to quantify 2,920 distinct proteins in parallel(5). After rigorous quality control procedures (details available at biobank.ndph.ox.ac.uk/ukb/ukb/docs/PPP_Phase_1_QC_dataset_companion_doc.pdf), proteins were assessed across four panels encompassing cardiometabolic, inflammatory, neurological, and oncological proteins. Additional details regarding sample selection, as well as Olink assay processing and quality control, have been documented in previous studies(6,7). Ultimately, 2,920 unique proteins were included in this study. Plasma protein data were obtained from Category 1839.

We selected 53,026 participants who had undergone plasma protein sampling between 2006 and 2010 (Instance 0). Incident depression was defined as ICD-10 diagnoses of depressive episode (F32) or recurrent depressive disorder (F33), derived from field IDs 130894 and 130897, occurring after plasma sampling. To minimize confounding, we excluded individuals with other major psychiatric disorders[35].This yielded a final sample of 2,127 participants with incident depression.

Identification of depression-associated plasma proteins

Following previous work[36] , we employed Cox proportional hazards models to identify plasma proteins associated with incident depression, ensuring that depression diagnoses occurred after baseline plasma collection. All regressions adjusted for age, sex, ethnicity, Townsend deprivation index, body mass index (BMI), smoking status, fasting time, season of blood collection (summer/autumn: June–November vs. winter/spring: December–May), and blood age (interval from blood collection to protein assay). To ensure robustness, we selected the most significant proteins with $p < 1 \times 10^{-3}$ (top 10% of significant hits) for downstream clustering analysis, yielding 215 proteins.

Consensus non-negative matrix factorization (cNMF) and subtyping

To derive latent dimensions from the 215 selected proteins, we applied a meta-analysis framework of non-negative matrix factorization (cNMF) [16], which improves robustness

and accuracy compared to alternative decomposition methods such as ICA, LDA, or standard NMF. Briefly, the protein expression matrix (N × M) was decomposed into two components: (i) a usage matrix (N × k), representing each participant's loading on *k* latent components (hereafter referred to as dimensions), and (ii) a weight matrix (k × M), indicating the contribution of each protein to each dimension[37].

The choice of k was guided by two complementary approaches[16,38]. First, we evaluated stability (silhouette score) and reconstruction error (Frobenius norm) across a range of *k* values. [39] Second, we confirmed that the components identified at $k = 3$ remained reproducible across alternative *k* values (Pearson correlation > 0.7), supporting their robustness. Relevant diagnostic plots are provided in the Supplementary eFigure 2.

Following previous work[17], participants in the incident depression group were assigned to subtypes according to their dominant loading. Specifically, participant *i* was classified into subtype *q* if Usage(q, i) was the maximum element of Usage(:, i).

Inflammation–age axis analysis

To characterize phenotypic separation across dimensions, we constructed a two-dimensional space defined by age and inflammation, with the latter indexed by plasma C-reactive protein (CRP) levels. Because CRP values were right-skewed, we applied a log-transformation for visualization (Fig. 2a). Each participant was located in this 2D space according to their age and CRP value.

To further assess nonlinear associations between age and dimension loadings, we fitted generalized additive models (GAMs) for each dimension (Fig. 2a inset). This allowed us to capture age-related trajectories of protein dimension scores beyond linear trends.

Phenotype projection analysis

To extend the analysis beyond the 2,127 incident depression cases, we linearly projected the dimension usage matrix obtained from cNMF onto all 53,026 participants with plasma protein data, as shown in Formula (1). This enabled phenome-wide association analyses of dimension loadings across a much larger cohort.

$$A^{\text{UKB-all}} = X^{\text{UKB-all}} \cdot \left(S^{incident-dep\ group}\right)^{-1}$$

where A denotes the reconstructed usage matrix for all participants, X is the participant-by-protein expression matrix for the full sample, and S is the dimension–protein weight matrix derived from the cNMF analysis of the incident depression subgroup.

We then constructed linear regression models for each phenotype variable across major phenotype domains, with the three plasma protein dimensions as independent variables. All models were adjusted for age, sex, ethnicity, Townsend deprivation index, body mass index (BMI), smoking status, fasting time, season of blood collection

(summer/autumn: June–November vs. winter/spring: December–May), and blood age (interval between blood collection and protein assay). Reported associations were Bonferroni-corrected for multiple testing.

For the depression symptom analyses presented in Fig. 2c, we used the Patient Health Questionnaire-9 (PHQ-9), which was administered after plasma protein sampling, ensuring temporal precedence between baseline protein measures and symptom outcomes.[35]

Subcortical volume data were obtained from UK Biobank Category 190 using the ASEG template. Cortical surface area data were extracted from Category 197 (field IDs 27334–27624) based on the Destrieux atlas, while cortical thickness data were derived from Category 197 (field IDs 27408–27698), also using the Destrieux atlas.

For organ system and lifestyle phenotypes, we adopted phenotype definitions consistent with prior UK Biobank studies[40], A detailed list of included variables is provided in the Supplementary Materials.

For analyses of prior disease associations, we considered all diagnostic labels recorded before plasma protein sampling, excluding depression, as phenotypes. This allowed us to capture the relationship between protein dimensions and pre-existing conditions independent of depression onset.

Weighted gene co-expression network analysis (WGCNA) and pathway enrichment

For each dimension, we identified protein sets by selecting the top 10% of proteins with the highest weights in the cNMF-derived dimension–protein weight matrix. These protein sets were used for the overlap analysis presented in Fig. 2b and served as the input for subsequent functional analyses.

To compare the three protein sets, we employed the WGCNA package in R [41] to construct a co-expression network based on expression similarity across all 2,920 plasma proteins. Modules were identified by clustering proteins with correlated expression patterns.

To explore biological pathways associated with each dimension-specific protein set, we performed functional annotation and tissue enrichment analysis using the gene2func module implemented on the FUMA platform (https://fuma.ctglab.nl/).

In addition, protein–protein interaction (PPI) networks for Dimension 1 and Dimension 3 were constructed using the STRINGdb package in R[42].

Survival prediction analysis

We applied Cox proportional hazards models to evaluate the prospective risk of incident depression across all participants who underwent plasma protein sampling. Subtypes were assigned to all individuals according to the dimension-dominant

classification derived previously, and ICD-10 depression diagnoses (F32, F33) were used as event outcomes. Models were adjusted for age, sex, body fat, and other baseline covariates.

Because Dimension 2 loadings were strongly influenced by age, we adopted a landmark analysis [43], to satisfy the proportional hazards assumption. Specifically, a 4.5-year cutoff was chosen based on the median age differences between Dimension 1/Dimension 3–dominant versus Dimension 2–dominant subtypes in the incident depression group. Separate Cox models were then fitted for the early (0–4.5 years) and late (>4.5 years) follow-up periods, with only depression-free participants retained for the latter.

Mediation and moderation analyses

Mediation analyses were conducted using the lavaan R package[44]. All variables were standardized (z-scored) across participants prior to model fitting to ensure comparability of regression coefficients. Confounding effects of covariates described in the phenotype analyses were removed via nuisance regression before model estimation. Each mediation model was bootstrapped with 5,000 resamples to derive confidence intervals. *P*-values were computed as the proportion of bootstrap samples in which the estimated coefficient was greater (positive effect) or smaller (negative effect) than zero. Model fit was evaluated with three indices—comparative fit index (CFI),

root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR)—as recommended by prior UK Biobank studies[45]. A model was deemed satisfactory if CFI > 0.90, RMSEA < 0.05, or SRMR < 0.08[46].

The moderated mediation model (Fig. 6c) was implemented using the PROCESS macro (Model 14), adapted for R[47]. This allowed us to test whether the mediation pathway from age to depression via Dimension 2 was significantly moderated by amygdala volume.

References

1. Ferrari, A. J. *et al.* Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet* **403**, 2133–2161 (2024).
2. Cuthbert, B. N. & Insel, T. R. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* **11**, 126 (2013).
3. Maj, M. *et al.* The clinical characterization of the adult patient with depression aimed at personalization of management. *World Psychiatry* **19**, 269–293 (2020).
4. Harald, B. & Gordon, P. Meta-review of depressive subtyping models. *J Affect Disord* **139**, 126–140 (2012).

5. Croxtall, J. D. & Scott, L. J. Olanzapine/fluoxetine: a review of its use in patients with treatment-resistant major depressive disorder. *CNS Drugs* **24**, 245–262 (2010).
6. Preskorn, S. H. Subtypes of Major Depressive Disorder Based on Pharmacological Responsiveness. *J Psychiatr Pract* **27**, 448–452 (2021).
7. Tozzi, L. *et al.* Personalized brain circuit scores identify clinically distinct biotypes in depression and anxiety. *Nat Med* **30**, 2076–2087 (2024).
8. Chen, D. *et al.* Neurophysiological stratification of major depressive disorder by distinct trajectories. *Nat. Mental Health* **1**, 863–875 (2023).
9. Bullmore, E. The art of medicine: Inflamed depression. *Lancet* **392**, 1189–1190 (2018).
10. Singh, P. *et al.* Metabolomics in Depression: What We Learn from Preclinical and Clinical Evidences. *Mol Neurobiol* **62**, 718–741 (2025).
11. Berk, M. *et al.* Comorbidity between major depressive disorder and physical diseases: a comprehensive review of epidemiology, mechanisms and management. *World Psychiatry* **22**, 366–387 (2023).
12. Shao, X. *et al.* Novel therapeutic targets for major depressive disorder related to oxidative stress identified by integrative multi-omics and multi-trait study. *Transl Psychiatry* **14**, 443 (2024).
13. Kang, J. *et al.* Plasma proteomics identifies proteins and pathways associated with incident depression in 46,165 adults. *Science Bulletin* https://doi.org/10.1016/j.scib.2024.09.041 (2024) doi:10.1016/j.scib.2024.09.041.

14. Lamers, F. *et al.* Serum proteomic profiles of depressive subtypes. *Transl Psychiatry* **6**, e851–e851 (2016).

15. Han, P., Min, L., Zhu, Y., Li, Z. & Liu, Z. A study on the plasma proteomics of different types of depressive disorders based on label-free data-independent acquisition proteomic technology. *Journal of Affective Disorders* **371**, 91–103 (2025).

16. Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* **8**, e43803 (2019).

17. Kim, H. & Park, H. Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* **23**, 1495–1502 (2007).

18. Jiang, R. *et al.* The brain structure, inflammatory, and genetic mechanisms mediate the association between physical frailty and depression. *Nat Commun* **15**, 4411 (2024).

19. Jiang, R. *et al.* The inflammatory and genetic mechanisms underlying the cumulative effect of co-occurring pain conditions on depression. *Science Advances* **11**, eadt1083 (2025).

20. Berk, M. *et al.* Comorbidity between major depressive disorder and physical diseases: a comprehensive review of epidemiology, mechanisms and management. *World Psychiatry* **22**, 366–387 (2023).

21. Berk, M. *et al.* So depression is an inflammatory disease, but where does the inflammation come from? *BMC Medicine* **11**, 200 (2013).

22. Bullmore, E. Inflamed depression. *The Lancet* **392**, 1189–1190 (2018).

23. Gao, X. *et al.* Accelerated biological aging and risk of depression and anxiety: evidence from 424,299 UK Biobank participants. *Nature Communications* **14**, 2277 (2023).

24. Billones, R. R., Kumar, S. & Saligan, L. N. Disentangling fatigue from anhedonia: a scoping review. *Transl Psychiatry* **10**, 273 (2020).

25. Berridge, K. C. & Kringelbach, M. L. Pleasure systems in the brain. *Neuron* **86**, 646–664 (2015).

26. Krishnan, K. R. R. *et al.* Magnetic Resonance Imaging of the Caudate Nuclei in Depression: Preliminary Observations. *Arch Gen Psychiatry* **49**, 553–557 (1992).

27. Oh, H., Lee, J., Patriquin, M. A., Oldham, J. & Salas, R. Reward Processing in Psychiatric Inpatients With Depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **8**, 731–740 (2023).

28. Milaneschi, Y., Lamers, F., Berk, M. & Penninx, B. W. J. H. Depression Heterogeneity and Its Biological Underpinnings: Toward Immunometabolic Depression. *Biological Psychiatry* **88**, 369–380 (2020).

29. Karrouri, R., Hammani, Z., Benjelloun, R. & Otheman, Y. Major depressive disorder: Validated treatments and future challenges. *World J Clin Cases* **9**, 9350–9367 (2021).

30. Liu, M. *et al.* Bidirectional relations between depression symptoms and chronic kidney disease. *J Affect Disord* **311**, 224–230 (2022).

31. Prevalence of depression in chronic kidney disease: systematic review and meta-analysis of observational studies. *Kidney International* **84**, 179–191 (2013).
32. Wang, L. *et al.* Structural and Functional Abnormities of Amygdala and Prefrontal Cortex in Major Depressive Disorder With Suicide Attempts. *Front. Psychiatry* **10**, 923 (2020).
33. Association between the volume of subregions of the amygdala and major depression with suicidal thoughts and anxiety in a Chinese cohort. *Journal of Affective Disorders* **312**, 39–45 (2022).
34. Zhang, B. *et al.* Identifying behaviour-related and physiological risk factors for suicide attempts in the UK Biobank. *Nat Hum Behav* **8**, 1784–1797 (2024).
35. Cai, N. *et al.* Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics* **52**, 437–447 (2020).
36. Deng, Y.-T. *et al.* Atlas of the plasma proteome in health and disease in 53,026 adults. *Cell* **188**, 253-271.e7 (2025).
37. Stein-O'Brien, G. L. *et al.* Enter the Matrix: Factorization Uncovers Knowledge from Omics. *Trends Genet* **34**, 790–805 (2018).
38. Hwang, W. L. *et al.* Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment. *Nat Genet* **54**, 1178–1191 (2022).
39. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering Signatures of Mutational Processes Operative in Human Cancer. *Cell Reports* **3**, 246–259 (2013).

40. Tian, Y. E. *et al.* Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. *Nat Med* **29**, 1221–1231 (2023).

41. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

42. Szklarczyk, D. *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Research* **51**, D638–D646 (2022).

43. Maeng, M. *et al.* Differential clinical outcomes after 1 year versus 5 years in a randomised comparison of zotarolimus-eluting and sirolimus-eluting coronary stents (the SORT OUT III study): a multicentre, open-label, randomised superiority trial. *Lancet* **383**, 2047–2056 (2014).

44. Rosseel, Y. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* **48**, 1–36 (2012).

45. Tian, Y. E., Cole, J. H., Bullmore, E. T. & Zalesky, A. Brain, lifestyle and environmental pathways linking physical and mental health. *Nat. Mental Health* **2**, 1250–1261 (2024).

46. Byrne, B. M. *Structural Equation Modeling with EQS and EQS-Windows: Basic Concepts, Applications, and Programming*. (Sage Publications, Inc., USA, 1994).

47. Hayes, A. F. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. xvii, 507 (The Guilford Press, New York, NY, US, 2013).