Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Extracellular Matrix Stiffness Transcriptional Signatures in Major Depressive Disorder: Robustness to Cellular Composition, Absence of Diagnostic Specificity, and Blood-Brain Discordance

Yansong Liu
Hunan University of Traditional Chinese Medicine
Ziliang Guo
Hunan University of Traditional Chinese Medicine
Lin Zhu
Hunan University of Traditional Chinese Medicine
Chenglin Diao
Hunan University of Traditional Chinese Medicine
Wanghua Liu
Hunan University of Traditional Chinese Medicine
Junfeng Yan
JunfengYan@hnucm.edu.cn
Hunan University of Traditional Chinese Medicine

Research Article

Keywords: major depressive disorder, extracellular matrix stiffness, circadian rhythm, multi-cohort validation, cellular composition adjustment, cross-modality validation, negative results

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11615260/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Extracellular Matrix Stiffness Transcriptional Signatures in Major Depressive Disorder: Robustness to Cellular Composition, Absence of Diagnostic Specificity, and Blood-Brain Discordance

Yan-Song Liu1[1, 4], Zi-Liang Guo[1, 4], Lin Zhu[3, 4], Cheng-Lin Diao[1, 4], Wang-Hua Liu[1, 2*], Jun-Feng Yan[3, 4*]

1. School of Traditional Chinese Medicine, Hunan University of Chinese Medicine, Changsha 410208, Hunan, China;
2. Provincial Key Laboratory of TCM Diagnostics, Hunan University of Chinese Medicine, Changsha 410208, Hunan, China;
3. School of Informatics, Hunan University of Chinese Medicine, Changsha 410208, Hunan, China;
4. AI TCM Lab Hunan, Changsha 410208, Hunan, China.

Abstract

Major depressive disorder (MDD) lacks a molecular biomarker; extracellular matrix (ECM) remodelling and circadian disruption are candidate mechanisms. We pre-specified and tested three predictions about the ECM stiffness transcriptional programme: that it is altered in MDD, that it distinguishes MDD from bipolarand that it is coupled to circadian output. Discovery was in whole bloodwhole blood (GSE98793, n = 192); validation used a control brain series with recorded time of death (GSE45642) and two post-mortem cohorts (GSE53987, GSE92538). Robustness was assessed by entering estimated cell composition, batch, age, sex and anxiety comorbidity into the design matrix; independent genetic evidence came from FinnGen release 12 (59,333 cases). Blood analysis identified 1,151 differentially expressed genes, and both network strategies converged on matrix and mechanotransduction modules (odds ratio 5.58; permutation P = 0.001). Stiffness scores were elevated in MDD ($P = 7.9 \times 10^{-4}$). After adjustment for cellular composition, the ECM core (P = 0.020) and stiffness (P = 0.046) sets survived, whereas the broad literature-derived and mechanotransduction sets did not. In dorsolateral prefrontal cortex the programme moved oppositely, and blood–brain log-fold changes were negatively correlated. Specificity was not established (P = 0.057; pooled P = 0.27), no 24 h rhythmicity was detected, and common-variant enrichment was absent (all P > 0.18). ECM remodelling is a reproducible transcriptional signature of MDD that survives compositional adjustment, but it is not

---

[Funding Project] National Natural Science Foundation of China (82574912); Natural Science Foundation of Hunan Province (2023JJ30451, 2025JJ50719); Hunan Provincial Health Research Project (20256419).

[Author Profile] Yan-song Liu (1998-), male, Ph.D. Candidate at Hunan University of Chinese Medicine, major in Digital Traditional Chinese Medicine. Email: 3110537614@qq.com.

[Corresponding Author]

Wang-hua Liu (1973-), male, professor, PhD, doctoral supervisor at Hunan University of Chinese Medicine, major in the prevention and treatment of cerebrovascular diseases with Traditional Chinese Medicine. Email: 003439@hnucm.edu.cn.

Jun-Feng Yan (1965-), female, professor, PhD, doctoral supervisor at Hunan University of Chinese Medicine, major in smart Traditional Chinese Medicine Project. Email: JunfengYan@hnucm.edu.cn.

diagnosis-specific at current sample sizes, is not coupled to circadian programmes, and cannot be read from blood as a proxy for the brain.

Keywords: major depressive disorder; extracellular matrix stiffness; circadian rhythm; multi-cohort validation; cellular composition adjustment; cross-modality validation; negative results

Running title: ECM Signatures in MDD: Robustness and Limits

1. Introduction

Major Depressive Disorder (MDD) is a main factor of the global burden of disease and years lived with disability [1]. However, diagnosis still relies on symptom-based clinical interviews and assessment of disease course. The phenotypic heterogeneity and etiological diversity of mood disorders complicate the use of any single diagnostic test as the "gold standard" [2]. Transcriptomic profiling, which supports mechanistic analysis and biomarker screening while enabling large sample sizes, was previously considered a promising technological approach to address this diagnostic limitation. However, more than two decades of research indicate that peripheral blood expression signals at the individual-gene level are reproducible across cohorts but have negligible effect sizes. An integrated analysis of two independent case-control studies showed cross-cohort reproducibility for approximately 90 upregulated innate immunity genes and 75 downregulated adaptive immunity genes [3]; however, most log-fold changes ranged from 0.1 to 0.3, similar in magnitude to variation due to age, sex, medication exposure, blood sampling batches, and leukocyte composition.

These limitations have redirected analytical focus from individual genes to clusters with combined expression changes, replacing single-gene measures with more stable co-expression structures and adding a temporal dimension alongside expression abundance. Disruption of temporal organization is among the most consistent clinical manifestations of MDD, covering advancement or delay of sleep-wake cycles, diurnal mood variation (e.g., worse symptoms in the morning with improvement in the evening), and phase shifts in endocrine rhythms, including cortisol secretion [4,5]. Studies using Time of Death (TOD) as a phase coordinate have identified 24-hour rhythmic variation in hundreds of transcripts across six cortical and limbic brain regions; in patients with MDD, however, this periodicity is attenuated, with shifts in peak timing and disrupted phase relationships among clock genes [6]. These findings provide some of the most direct observations of the human brain's biological clock currently available. However, the precise time of blood collection is rarely documented in clinical practice, and the limited amplitude of core clock gene mRNA expression makes irregular sampling times a source of measurement noise. Therefore, replicating these findings in peripheral blood is challenging, and negative rhythm-analysis results from public peripheral blood datasets cannot exclude rhythmic genes from the disease.

A third line of evidence involves the ECM, which actively regulates cellular behaviour through adhesion receptor signalling, including integrins, and through mechanical properties such as stiffness and deformability [7,8]. In the brain, the ECM forms perineuronal nets (PNNs) around neuronal cell bodies, particularly those of parvalbumin-positive

interneurons, and regulates neuronal excitability and gamma oscillations [9,10]. Chronic stress increases PNN deposition [11,12], whereas antidepressant treatment can restore excitation-inhibition balance and improve behavioural outcomes through matrix metalloproteinase-mediated PNN degradation; PNN density also varies across psychiatric disorders in post-mortem dorsolateral prefrontal cortex tissue [13,14]. These findings establish ECM remodelling as a key component of stress responses and mood regulation. However, most evidence derives from animal models and small-scale post-mortem studies, highlighting the need for transcriptomic analyses that observe both cross-tissue consistency and diagnostic specificity.

The circadian rhythm and ECM axes are mechanistically interdependent, not entirely separate biological systems. Matrix-derived mechanical signals can be transduced *via* the integrin-actin cytoskeleton-MRTF/SRF pathway to modulate the activity of core clock genes [15]. In comparison, the circadian clock regulates the transcription of matrix-associated genes; in matrix-rich tissues (tendons and cartilage) it regulates collagen synthesis and the expression of matrix metalloproteinases (MMPs) and members of the lysyl oxidase (LOX/LOXL) family [16,17], while the structure of the brain's ECM itself shows circadian rhythmicity [18]. Therefore, circadian disruption may contribute to MDD pathogenesis through changes in matrix stiffness; however, to our knowledge, no previous study has combined these two lines of investigation into a single pathway or directly evaluated their coupling in the context of the disorder.

This study pre-specified three predictions about the ECM stiffness transcriptional programme and tested all three: that the programme is altered in MDD; that any alteration distinguishes MDD from bipolar disorder and schizophrenia rather than being shared across diagnoses; and that the programme is coupled to circadian transcriptional output. Testing is sequential. Disease-associated expression change is first established in peripheral blood, the programme is then sought in brain tissue, specificity is then tested by direct comparison between diagnoses rather than by comparing significance statuses, and coupling is evaluated in control brain with documented times of death. Three elements were added because each prediction can fail for reasons unrelated to the underlying biology: adjustment for estimated cellular composition, which can generate or remove an apparently disease-associated signature; direct between-diagnosis contrasts, which carry less power than the corresponding case-control comparisons and are reported as such; and analysis of an external resource differing in modality as well as population, so that a concordant finding across transcriptomic cohorts is not mistaken for independent corroboration. Both positive and negative findings are reported.

2. Research Background and Mechanisms

2.1 Molecular mechanisms of circadian rhythms

Circadian rhythms arise from a set of molecular components termed "clock genes," organized around a transcription-translation feedback loop. Genes of the positive regulatory arm, CLOCK and ARNTL (also known as *BMAL1*), are transcribed into mRNA and translated into proteins; their protein products assemble into a complex that

translocates to the nucleus and binds specific DNA sequences, termed E-boxes, within target gene promoters, thus initiating downstream gene transcription (Figure 1). Among the induced genes, members of the negative arm encode PER and CRY family proteins that accumulate in the cytoplasm, assemble into complexes, and re-enter the nucleus to suppress CLOCK-ARNTL activity. Progressive degradation of PER/CRY proteins releases this inhibition and allows initiation of the next cycle. Repetition of the cycle approximately every 24 h establishes the molecular basis of circadian rhythmicity [19].

In addition to the positive and negative regulatory arms, an auxiliary arm comprising ROR and REV-ERB (NR1D1/2), along with components such as the casein kinases (CSNK1D/E), stabilizes the circadian cycle period. Outputs from this feedback loop comprise clock-controlled genes (CCGs) that regulate metabolism, cell-cycle progression, and hormone secretion; circadian control of physiological functions is mediated through these downstream outputs. The timing of the molecular clock depends not only on mRNA abundance; post-translational modifications, including phosphorylation, nuclear translocation, and degradation of PER/CRY proteins, along with rhythmic variation in protein translation itself, are crucial factors in cycle precision [20,21].

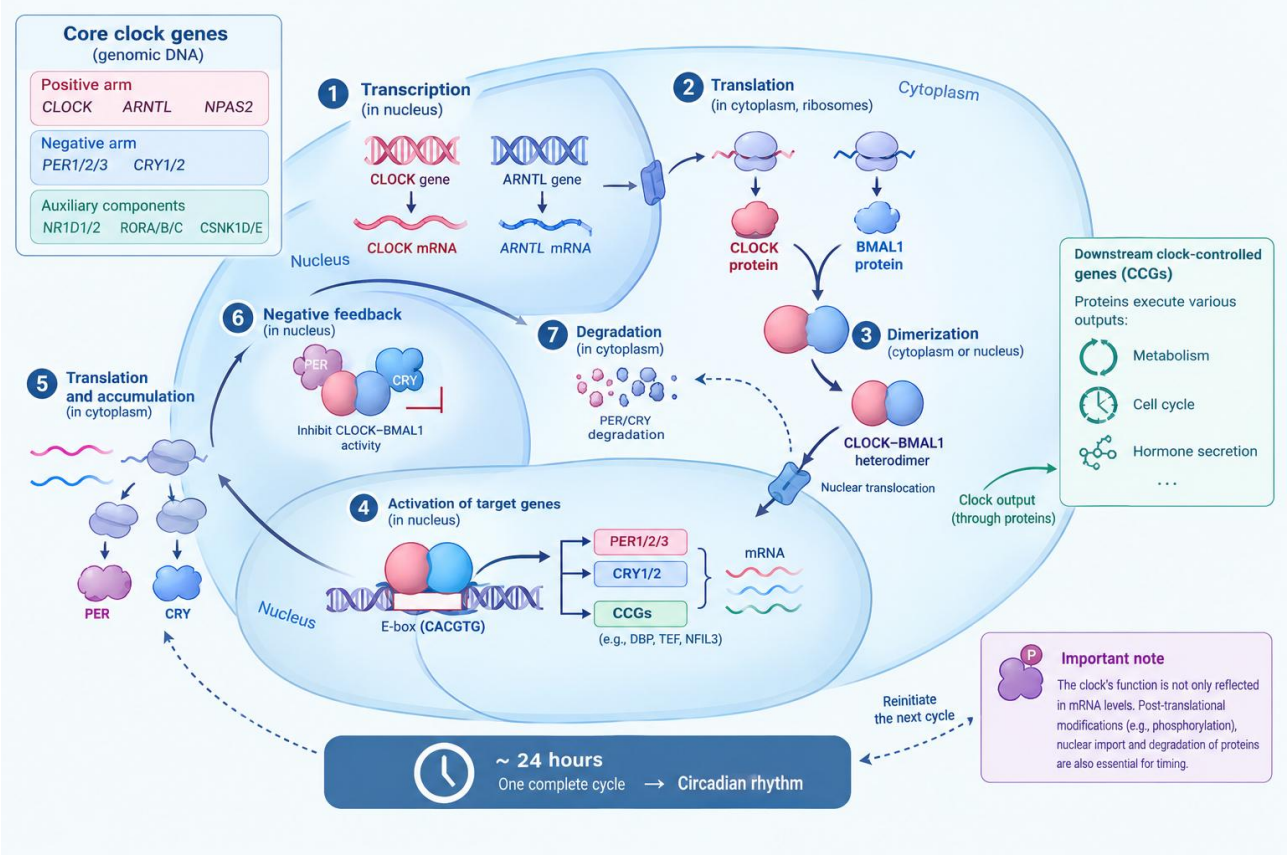


Figure 1. Molecular mechanisms of circadian rhythms.

Clock genes establish an approximately 24-hour cycle through a transcription-translation feedback loop: the positive-arm protein complex (CLOCK and ARNTL) binds E-box elements to initiate transcription of PER/CRY and clock-controlled genes; PER/CRY proteins accumulate and translocate back to the nucleus to inhibit the positive arm,

and protein degradation releases this inhibition, permitting re-initiation of the cycle. ROR/REV-ERB auxiliary arm and CSNK1D/E contribute to cycle stability; clock-controlled genes mediate rhythmic outputs; and post-translational modifications and rhythmic translation itself are crucial factors of temporal precision.

2.2 Mechanotransduction mechanisms of extracellular matrix stiffness

Matrix stiffness denotes the ECM's capacity to resist deformation in a mechanical environment and reflects its molecular composition and modification state. Stromal cells, including fibroblasts and astrocytes, continuously secrete collagen, fibronectin, and laminin to establish a load-bearing scaffold. Enzymes such as lysyl oxidase (LOX/LOXL) and transglutaminase (TGM2) covalently cross-link collagen fibres, while tissue inhibitors of metalloproteinases (TIMPs) suppress matrix metalloproteinase (MMP)-mediated degradation of the matrix. The balance among synthesis, cross-linking, and degradation determines whether the matrix becomes stiffer or softer; stiffness is therefore a dynamic property that undergoes continuous remodelling through cellular activity [7,8].

Mechanotransduction is the process by which cells detect stiffness-related signals and convert them into cellular responses. Integrins embedded in the cell membrane act as primary mechanosensors: increasing matrix stiffness enhances integrin-mediated transmission of tension into the cell, promoting expansion and maturation of focal adhesions, where scaffold proteins such as FAK, VCL, and TLN1 accumulate, followed by increased RHOA-ROCK-mediated cellular contractility; simultaneously, changes in phosphorylation state promote nuclear translocation of the cytoplasmic transcriptional co-activators YAP and TAZ, which associate with TEAD transcription factors to induce expression of genes including *CCN2, ANKRD1*, *collagen*, and *LOX*. The resulting gene products further enhance matrix deposition and cross-linking, thus establishing a positive feedback loop (Figure 2).

In the brain, this system is mediated by a specialized structural component: PNNs surrounding fast-spiking interneurons. Their formation is regulated by neuronal activity, and they constrain synaptic plasticity. The integrin-actin cytoskeleton-MRTF/SRF pathway also couples the matrix's mechanical state to the circadian system, enabling transmission of mechanical signals to the molecular clock and modifying its period [15], while the circadian clock directly controls the expression of multiple matrix genes at the transcriptional level. Shared pathways for molecular crosstalk support simultaneous evaluation of both axes within the same transcriptomic dataset.

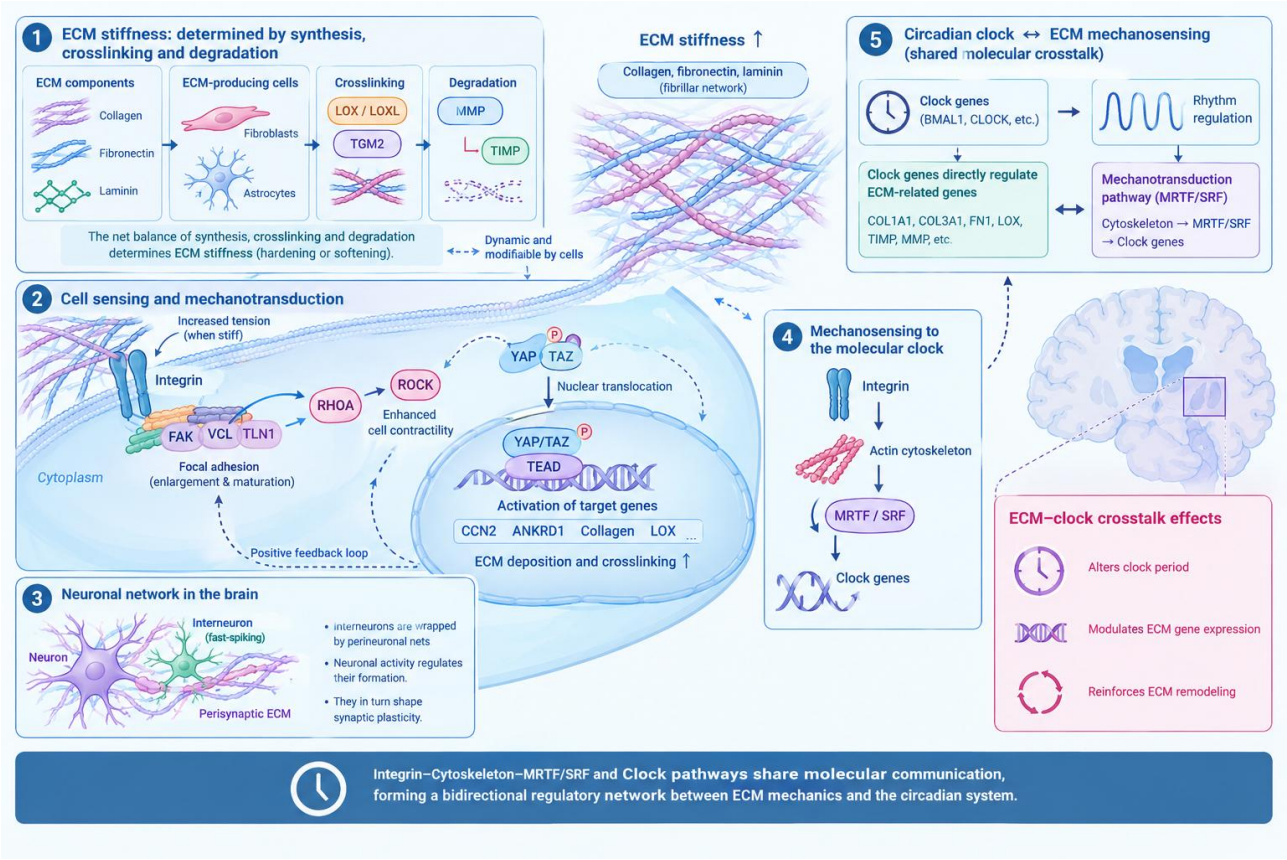


Figure 2. Formation of ECM stiffness and mechanotransduction.

Stromal-cell-secreted structural components stiffen through LOX/LOXL2- and TGM2-mediated cross-linking, along with TIMP-mediated inhibition of degradation. Cells detect tension changes through integrin-focal adhesion complexes and regulate gene expression via RHOA-ROCK and YAP/TAZ-TEAD signalling; the resulting gene products further enhance matrix deposition, creating a positive feedback loop. Mechanisms including the MRTF/SRF pathway link this signalling system to circadian regulation.

3. Study design

The study tests three pre-specified predictions rather than a single directional hypothesis: that the ECM stiffness transcriptional programme is altered in MDD, that any alteration is specific to MDD, and that the programme is coupled to circadian transcriptional output. To evaluate them the analysis proceeds sequentially: identify disease-associated matrix and mechanotransduction signals in peripheral blood; seek the same programme in brain tissue; test specificity by direct comparison between diagnoses within one cohort, one platform and one brain region; and evaluate coupling in control brain samples with documented times of death. Four publicly available transcriptomic datasets are used, each addressing a distinct analytical stage, together with one external genome-wide association resource that differs in modality as well as in population (Figure 3).

Differential expression analyses identify candidate genes. Two co-expression network approaches shift the analytical unit from individual genes to clusters of genes with coordinated expression change: a “seed network”

incorporating pre-selected genes supports targeted localization, whereas an unbiased genome-wide network provides independent validation. Module-gene set overlap tests assign biological context to these modules. Sample-level scoring and a pre-specified exploratory path analysis evaluate associations at the individual level, while estimating cellular composition excludes confounding by differences in cell proportions. Cosinor analysis uses time of death as the phase coordinate and compares three psychiatric diagnoses to address rhythm coupling and diagnostic specificity. Four further elements complete the chain of inference. Gene-set testing with covariates entered directly into the design matrix judges robustness at the level of the set rather than of a single summary score. Direct between-diagnosis contrasts, pooled across cohorts by inverse-variance weighting, test specificity instead of inferring it from differing significance statuses. Single-gene attribution and leave-one-gene-out sensitivity trace a surviving effect to its contributing members. Gene-set analysis of external genome-wide association summary statistics supplies evidence that shares neither modality nor population with the transcriptomic cohorts. Conclusions are supported by unbiased genome-wide network validation, independent-cohort replication, and sensitivity analyses addressing covariates, batch effects, and cellular composition; both positive and negative findings were presented.

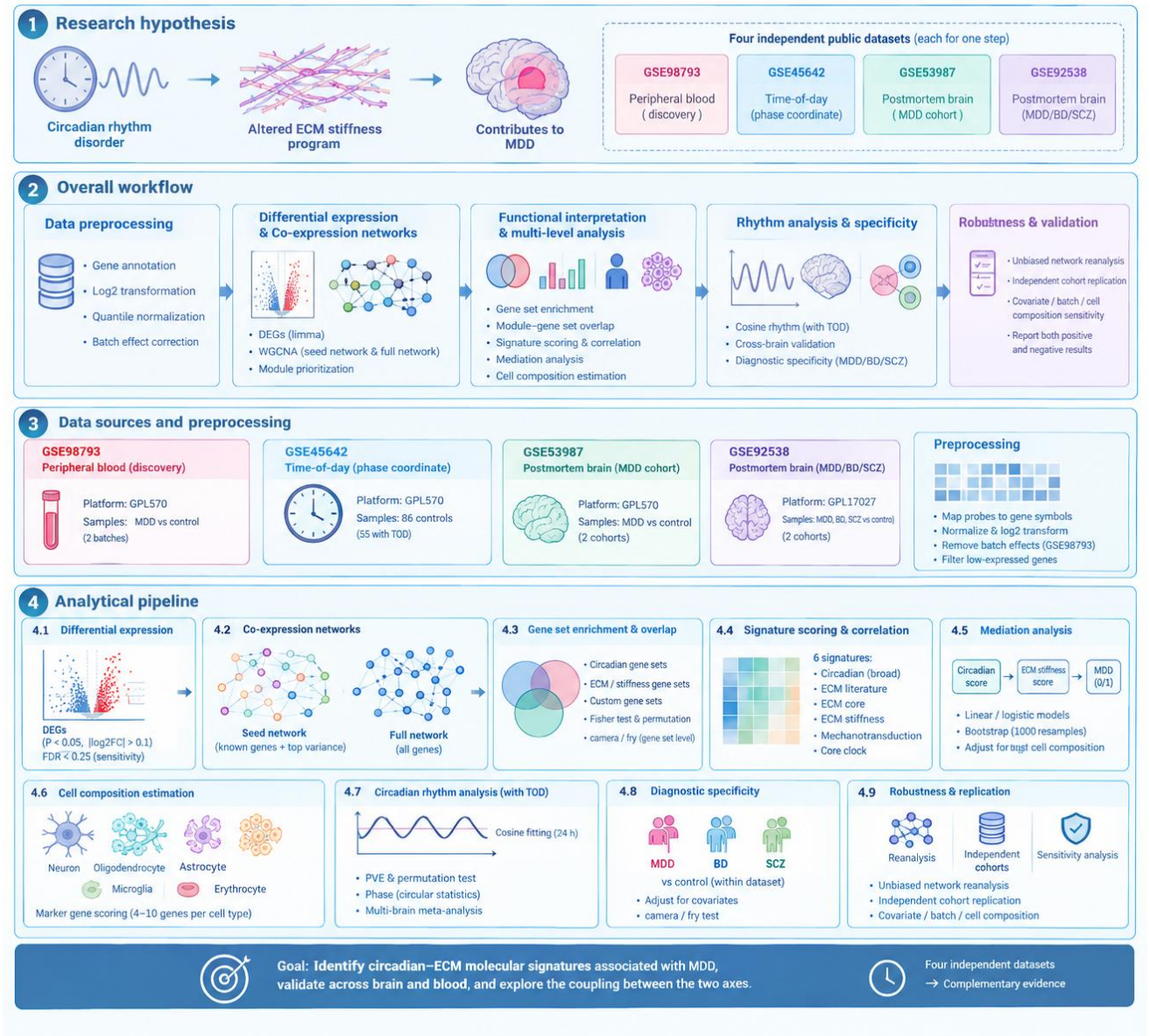


Figure 3. Study design and analysis workflow.

4. Materials and Methods

4.1 Data sources and study design

The study used four publicly available human transcriptomic datasets from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/, accessed on 30 September 2026) [22]; details of tissue origin, analytical platforms, and sample composition are summarised in Table 1. GSE98793 [3] provided discovery evidence from peripheral blood, while GSE45642 [6] provided phase-coordinate sources. GSE53987 [23,24] and GSE92538 [25] comprised two independent post-mortem brain case-control cohorts; GSE92538 also included patients with bipolar disorder (BD) and schizophrenia (SCZ), enabling evaluation of diagnostic specificity. All data were publicly accessible, de-identified, and collected with ethical approval in the original studies; accordingly, no new ethical review was required for this study. Because concordance among transcriptomic cohorts does not constitute independent corroboration, evidence

from a second biological modality was also sought: depression genome-wide association summary statistics from FinnGen release 12 (F5_DEPRESSIO, 59,333 cases).

Table 1. Dataset characteristics and roles in the analysis

Dataset	Tissue	Platform	Sample composition
GSE98793	Peripheral whole blood	GPL570	192 subjects (MDD: 128, Controls: 64)
GSE45642	Post-mortem brain (6 regions)	GPL17027	55 controls (324 arrays); TOD recorded for all
GSE53987	Post-mortem brain (3 regions)	GPL570	105 subjects (MDD: 50, Controls: 55)
GSE92538	Post-mortem brain (BA9/46)	GPL17027	235 subjects (MDD: 47, BD: 42, SCZ: 27, Controls: 119)

BA, Brodmann area; BD, bipolar disorder; SCZ, schizophrenia; MDD, major depressive disorder. All six brain regions represented in GSE45642 are the dorsolateral prefrontal cortex, anterior cingulate cortex, nucleus accumbens, hippocampus, cerebellum, and amygdala; the three regions represented in GSE53987 are the prefrontal cortex, hippocampus, and striatum. GPL570 shows the Affymetrix U133 Plus 2.0 array, whereas GPL17027 denotes the Affymetrix U133A array used with custom Entrez Gene chip definition files.

The original GSE45642 study enrolled 86 control subjects, of whom 55 had documented time of death (TOD) data that provided the coordinates for this analysis [6].

4.2 Data preprocessing

Expression matrices were annotated with gene symbols using platform-specific procedures. For GSE98793 and GSE53987, generated on the GPL570 platform (standard Affymetrix probe sets), platform annotations were used to map probe identifiers to gene symbols, and expression values were log2-transformed and then quantile-normalised [26]. For GSE45642 and GSE92538, generated on the GPL17027 platform (U133A array with custom Entrez chip definition files), probe names directly represented Entrez Gene IDs; these identifiers were converted to gene symbols using the *Homo_sapiens.gene_info* file from the National Center for Biotechnology Information (NCBI), with identical filtering and normalisation procedures used.

GSE98793 comprised two technical batches, each with 96 samples. A linear model implemented with `limma::removeBatchEffect` removed batch effects while preserving group effects. After correction, all absolute correlation coefficients between module eigengenes and the batch variable were below 0.016.

4.3 Differential expression analysis

Differential expression in peripheral blood was evaluated using the linear modelling and empirical Bayes framework implemented in the limma package [27]. Significance criteria were defined as $P < 0.05$ and $|\log_2 FC| > 0.1$; genes satisfying these thresholds were designated DEGs for the primary analysis. The Benjamini-Hochberg procedure [28] also controlled the false discovery rate (FDR) at < 0.25, and genes identified using this criterion were included in sensitivity analysis.

4.4 Weighted gene co-expression network analysis

Weighted Gene Co-expression Network Analysis (WGCNA) was performed using the R package of the same name [29,30]. Correlation coefficients were transformed using a soft-thresholding power function to approximate a scale-free network topology; the power exponent was selected to yield a scale-free topology fit index greater than 0.8. A

topological overlap matrix (TOM) was computed, followed by hierarchical clustering based on the topological overlap distance. Modules were identified by dynamic tree cutting with a minimum module size of 60 genes, and similar modules were merged using a module eigengene correlation threshold of 0.25. Each module's first principal component, termed the module eigengene, summarised its expression trajectory and was evaluated for association with disease status using Pearson correlation analysis.

Co-expression networks were generated for two distinct gene sets. The "seed gene network" used a pre-selection strategy that included peripheral blood DEGs, manually curated gene sets associated with circadian rhythms and matrix/mechanical properties, and the top 6,000 genes ranked by inter-sample variance across the complete expression matrix. Genes selected solely based on high variance were retained to prevent network modules from being dominated by the pre-selected genes. A "whole-genome network" adding all genes was also constructed to provide independent validation of the seed gene network.

4.5 Gene set construction and enrichment testing

Circadian rhythm gene sets were assembled from the union of eight circadian-related pathway sets cataloged in the Molecular Signatures Database (MSigDB v2026.2.Hs, gene annotation mapped to Ensembl 116; https://www.gsea-msigdb.org/gsea/msigdb, accessed on 20 October 2026) [31]: three from Gene Ontology [32,33], covering positive, negative, and general regulation of circadian rhythms, and five from KEGG [34], PID [35], Reactome [36], WikiPathways [37], and BioCarta [38]; all of these pathway entries are distributed through the MSigDB curated pathway collection [39]. The circadian sets were supplemented with manually curated sets of core clock and clock-output genes [19]. The principal matrix stiffness gene sets originated from a study of matrix stiffness in hepatocellular carcinoma [40]. They were supplemented with sets representing ECM structural components, pro-fibrotic and high-stiffness-associated genes, low-stiffness-associated genes, and mechanotransduction genes. The four manually curated matrix/mechanics sets were classified by function as structural components (collagens, elastin, other matrix scaffold genes, and remodelling enzyme genes); pro-stiffness components (genes involved in deposition, cross-linking, and anti-degradation); low-stiffness components (degradation enzyme genes); and mechanotransduction components (YAP/TAZ pathway and focal adhesion genes). The complete list of genes is provided in Tables S1 and File S2 and 2. The six manually curated sets supplied seed genes for co-expression network analysis, while all gene sets were incorporated into module enrichment and signature scoring; gene symbols were normalised and standardised against aliases across all sets. Platform limitations and the background gene set determined the detectable size of each set within a dataset; therefore, the set sizes reported in tables differ, and statistical testing used the total number of detectable genes in the corresponding dataset as the background.

Overlap between gene sets and co-expression modules was evaluated using Fisher's exact test. We also generated a null distribution of overlap counts through 1,000 iterations of gene label permutation, each randomly sampling from the complete gene pool to match the target set size, to estimate empirical *P values* characterising relationships between

the gene sets. Gene-set testing undertaken for formal statistical inference used rotation-based approaches that preserve the internal co-expression structure of the sets, specifically limma::camera [41] and fry.

4.6 Signature scoring and between-group comparisons

For each gene set, expression values were row-normalised across all samples, and the mean expression of all detectable genes in the set was used to define the signature score for each sample. Six gene sets were evaluated: the circadian pathway set (broad definition), matrix stiffness literature set, ECM core set, ECM stiffness set, mechanotransduction set, and core clock set. Between-group differences in scores were examined using the Wilcoxon rank-sum test, and associations among scores were quantified using Spearman rank correlation. Throughout, these scores index the transcriptional abundance of genes annotated to each programme; they are transcriptional proxies and do not measure tissue mechanical properties, which bulk transcriptomic data cannot resolve.

4.7 Exploratory path analysis

A pre-specified exploratory path analysis decomposed the association between the circadian score and disease status through the stiffness score, specifying the circadian pathway set (broad definition) score as the independent variable, the matrix stiffness literature set score as the intermediate variable, and disease status (MDD = 1, control = 0) as the outcome. Sequentially fitted models comprised the total effect model (outcome vs. independent variable), the path a model (intermediate variable vs. independent variable), and the path b/direct effect model (outcome vs. independent variable and intermediate variable); the indirect effect was calculated as the product of the path a and path b coefficients. Given the binary outcome, a linear probability model was used as the baseline specification, which allows direct decomposition of coefficients into path coefficients. The indirect-effect sampling distribution was estimated through 1,000 iterations of case resampling to derive a 95% CI; intervals excluding zero were considered statistically significant [42]. Robustness was also tested by adding cell composition scores as covariates, switching to a logistic regression outcome model, and using batch-corrected scores. Because the independent variable, the intermediate variable and the outcome were measured at a single time point in a case-control sample, and because the outcome was the sampling variable, this decomposition estimates only how the circadian-stiffness covariance relates to case status; it cannot identify a directed causal pathway. The terms mediator and indirect effect are therefore used descriptively, and the results are reported as exploratory and hypothesis-generating.

4.8 Cell composition estimation

A marker-gene scoring method [43] was used, selecting 4 to 10 canonical marker genes per cell type and using their mean expression as an indicator of that cell type's relative abundance. Six categories of marker scores were defined for brain tissue: neurons, oligodendrocytes, astrocytes, microglia, endothelial cells, and erythrocytes.

4.9 Cosine rhythm analysis based on TOD

Each subject's TOD was converted to Zeitgeber time (ZT) using the sunrise time at the location and date of death (sunrise defined as ZT = 0). Principal component regression was used to obtain residuals, and the data were median-

centred by sex; results obtained without this preprocessing were retained as robustness controls. For each gene, expression values ordered by TOD were fitted to a 24-hour sinusoidal function; the sine and cosine coefficients represented amplitude and peak phase [44], and the percentage of variance explained (PVE) quantified goodness of fit. Gene-specific null distributions of PVE were generated by randomly permuting TOD labels 1,000 times; an inflation factor (λ), estimated from the median chi-square statistics of *P value* distributions across brain regions, was applied to scale *P values*; results from all six regions were integrated using Fisher's combined test to generate cross-region ranking [45]. Phase properties were characterised by the circular mean (phase centre) and mean resultant vector length (phase concentration); the Rayleigh test evaluated whether phases were randomly distributed, and the Watson-Williams test compared phase distributions between groups [46], with background controls provided by null distributions generated through repeated sampling of random gene sets of equal size.

4.10 Analysis of psychiatric diagnosis specificity

Comparative analyses of MDD, BD, and SCZ were performed in parallel using the GSE92538 dataset. As the dataset contains two independent cohorts, control groups were limited to the cohort corresponding to each diagnostic comparison: MDD and BD were compared with 91 controls from the depression cohort, whereas SCZ was compared with 28 controls from the schizophrenia cohort. The primary model adjusted for age, sex, post-mortem interval, brain tissue pH, and agonal state score; three additional specifications, unadjusted, cohort-adjusted, and suicide-history-adjusted, were also used to observe robustness. Formal gene-set inference used the camera and fry rotation tests.

4.11 Statistical inference and result reproducibility

All *P values* were two-sided, and results involving multiple comparisons are reported as Benjamini-Hochberg-adjusted false discovery rates (FDR) [28]. Permutation-test *P values* were rounded to 0.001; values reported as 0.001 indicate that this lower bound was attained. Analyses were implemented in R v4.6.1, primarily using the limma, WGCNA, and ggplot2 packages, together with base R plotting functions. Stochastic procedures used fixed seeds (20261001; seed 20261002 for the GSE92538 analysis layer), and analysis scripts were organized by analytical level to support sequential reproducibility. The analysis code is provided in File S3.

4.12 Covariate-adjusted gene-set testing and single-gene attribution

Score-level comparisons reduce a set to one number per sample and discard the coherence that motivates defining the set in the first place. Robustness to cellular composition was therefore assessed at set level as well, entering the adjustment directly into the design matrix. Two linear models were fitted per gene: one containing disease status alone, and one containing disease status together with the seven estimated cell-type scores, batch, age, sex and anxiety comorbidity. Genes were left unfiltered, so that coherence rather than selection determined the outcome. Rotation-based tests were then applied to the full design with a contrast isolating the disease coefficient, thereby carrying all covariates while preserving inter-gene correlation. Because a surviving result could in principle reflect no more than a higher burden of individually significant genes, the proportion of genes with $P < 0.05$ within each set was compared

with the same proportion computed over all genes detectable in the same model. Single-gene moderated t-statistics from the adjusted model identified which members carried the remaining signal, and leave-one-out sensitivity removed genes in order of their adjusted significance, recomputing the set test each time, to establish whether the result depended on individual members. These analyses test whether a set moves coherently once covariates are accounted for. They do not establish that any surviving signal originates in the tissue of interest, in a particular cell population, or in ECM protein rather than in its transcripts.

4.13 Direct diagnostic contrasts and cross-cohort pooling

A difference that attains significance in one diagnosis and fails to in another does not establish that the diagnoses differ. Specificity was therefore tested by direct contrast. All contrasts were estimated within a single linear model of the standardised score on diagnosis, retaining controls as the reference, so that the MDD-minus-bipolar difference and its standard error were taken from the model covariance matrix, with age, sex, post-mortem interval and brain pH as covariates; a second specification added the estimated cell-type scores. The contrast was repeated at set level with rotation tests. Because the first cohort contained only 42 bipolar samples, the analysis sought to increase power rather than to reinterpret a borderline result. The second brain cohort contributes MDD, bipolar, schizophrenia and control subjects across three regions and was analysed under the same framework, each region treated as a separate estimate because individual subjects contributed tissue to more than one region. Standardised coefficients and standard errors were combined by inverse-variance weighting under a fixed-effect model; heterogeneity was assessed with Cochran's Q and the derived I-squared statistic, and random-effects estimates were inspected where heterogeneity was appreciable. Because the two brain cohorts differ in brain bank, platform, regional sampling and diagnostic ascertainment, pooling is presented as an increase in power for the contrast rather than as evidence that the cohorts measure one common quantity.

4.14 Common-variant gene-set analysis of external genome-wide association summary statistics

Independence from the transcriptomic cohorts requires a resource differing in biological modality as well as in population. Summary statistics for the FinnGen release 12 phenotype F5_DEPRESSIO [47] (59,333 cases) were queried remotely through their tabix index without local download. Gene boundaries were taken from the UCSC refGene annotation for GRCh38, merging transcripts by gene symbol, and each single nucleotide polymorphism with minor allele frequency of at least 0.01 was assigned to every gene whose boundary it fell within, extended by 50 kb on both sides. Each gene was scored by the sum of single-degree-of-freedom chi-square statistics over its assigned variants, a quantity that responds to several weak signals in combination as well as to one strong signal. Because this score increases mechanically with the number of assigned variants, background genes were stratified into twenty bins by variant count and each gene's score converted to its empirical quantile among background genes within the same bin; under the null the mean quantile is 0.5 with standard error equal to the square root of one divided by twelve times the number of genes. Empirical two-sided *P* values were obtained by resampling random gene sets of equal size 2,000

times and recomputing the identical statistic. Four clock-related gene sets served as negative controls, having already returned non-significant results at the transcript level in these cohorts, so that their behaviour indicates the sensitivity of the pipeline rather than a property of the ECM sets.

5. Results

5.1 DEGs in peripheral blood

In the 192 samples in the GSE98793 dataset, 1,151 DEGs were identified among 22,876 genes; these comprised 512 upregulated and 639 downregulated genes (Figures 4a, b). Further filtering based on FDR identified 100 genes at FDR < 0.25 and one gene at FDR < 0.05.

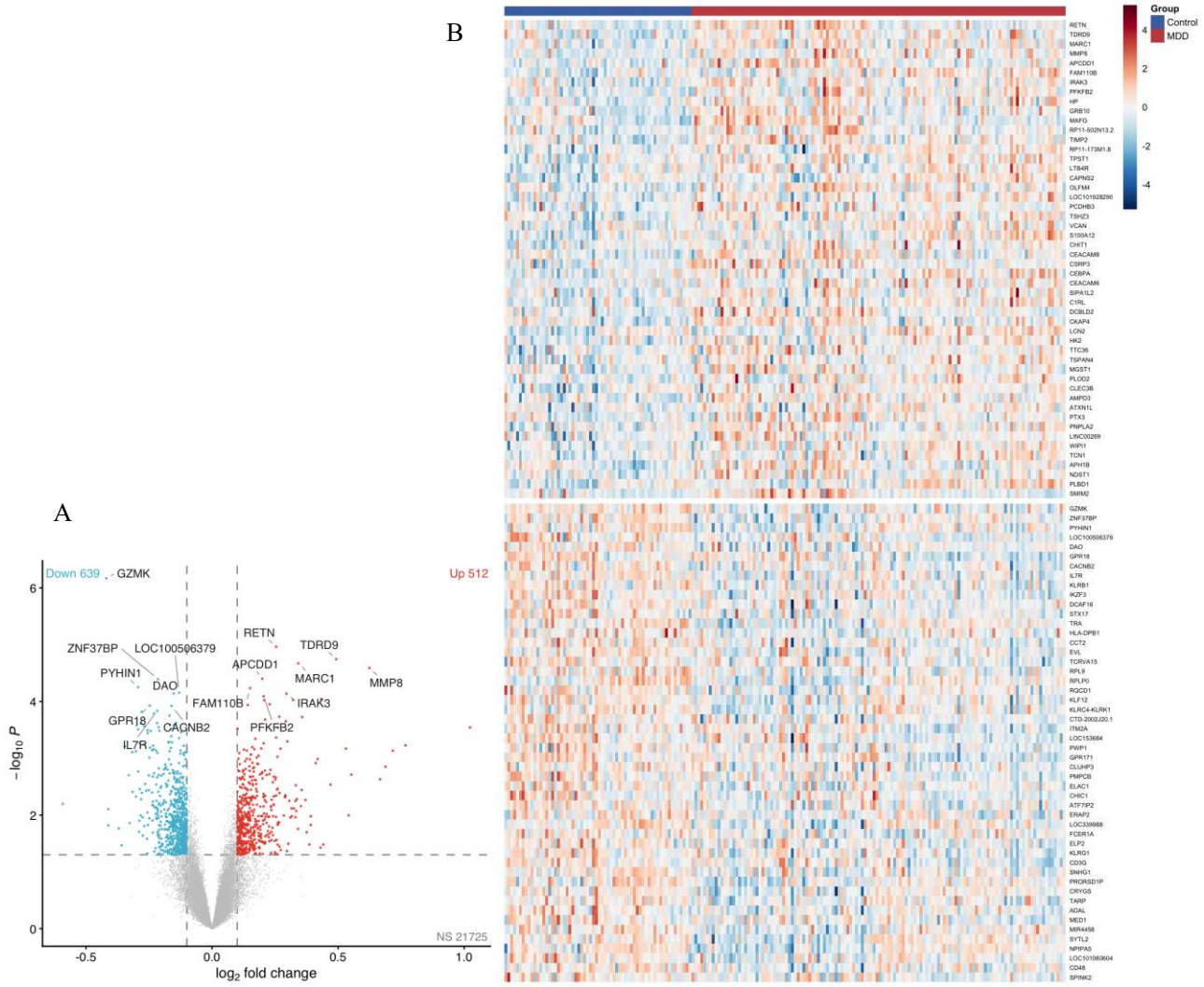


Figure 4. Analysis of differential gene expression in peripheral blood.

(a) Volcano plot of peripheral blood differentially expressed genes (DEGs) (GSE98793, n = 192); 1,151 DEGs were identified (512 upregulated and 639 downregulated). (b) DEG expression heatmap arranged by study group.

To test whether these expression findings reflected demographic or technical artifacts, a model adjusting for batch, age, and sex was fitted; the DEG count increased to 1,510, with 1,148 genes (99.7%) overlapping the original

findings, and the Spearman correlation coefficient for logFC was 1.000. Full differential expression results, including the covariate-adjusted model, are given in Table S4.

5.2 Co-expression network identifies modules related to matrix and mechanotransduction

After adjustment for batch effects, the seed gene network resolved into 7 modules, with no genes assigned to the unclassified module. Among modules significantly associated with disease status, correlation coefficients were -0.270 for the red module (285 genes; $P = 1.5 \times 10^{-4}$), +0.217 for the blue module (1,026 genes; $P = 0.0025$), -0.187 for the pink module (62 genes; $P = 0.0098$), +0.161 for the black module (80 genes; $P = 0.026$), and -0.161 for the brown module (1,479 genes; $P = 0.026$). After adjustment, the absolute module-batch correlation coefficients were uniformly below 0.016; associations with sex and anxiety scores were also non-significant, with only the pink module showing an age association at nominal significance ($P = 0.038$) (Figure 5a, b). The "black" module was enriched for the matrix stiffness literature set (21 overlapping genes; expected: 5.0; OR = 5.58; permutation $P = 0.001$; FDR = 0.011), as well as the mechanotransduction set (OR = 9.97) and ECM core set (OR = 4.57); the "turquoise" module showed enrichment for the ECM core set (OR = 2.27) and ECM stiffness set (OR = 2.39). The "stiffness" set enriched in the turquoise module is the manually curated ECM stiffness set (47 detectable genes in this dataset), distinct from the matrix stiffness literature set enriched in the black module. The literature set comprises 1,417 genes in total, of which 1,314 are detectable on the GSE98793 platform; signature scoring (Table 3) used all 1,314 genes, whereas module enrichment was computed within the 6,765-gene seed-network universe, in which the same set is represented by 422 genes. These two counts therefore denote different gene universes rather than different gene sets.

The whole-genome network contained 39 modules, and the absolute correlation coefficients between these modules and the batch effect were uniformly below 0.020. The modules most strongly associated with disease status were B2.greenyellow (-0.249), B1.pink (-0.234), B2.pink (-0.209), and B2.purple (+0.148). The B2.purple module comprised 152 genes and was enriched for the matrix stiffness literature set (OR = 3.93; permutation $P = 0.001$), mechanotransduction set (OR = 17.56; permutation $P = 0.001$), ECM core set (OR = 8.05), and ECM stiffness set (OR = 10.37). Both analytical strategies identified modules enriched for matrix and mechanotransduction gene sets (Figures 5c, d); despite their associations with disease status, the pink and brown modules contained individual members of these sets but showed no significant enrichment for any of the five matrix and mechanotransduction gene sets. Key module-level measures are presented in Table 2. Enrichment analysis of the seed gene network was a hypothesis-directed process of module identification rather than independent replication; independent evidence was provided by the whole-genome network.

Table 2. Association of modules with disease status and enrichment of matrix/mechanical gene sets (GSE98793)

Network strategy	Module (gene count)	Correlation coefficient (*P value*)	Enriched matrix or mechanical gene sets (OR)
Seed gene network	red (285)	-0.270 (1.5×10^{-4})	—
	blue (1,026)	+0.217 (0.0025)	—
	pink (62)	-0.187 (0.0098)	—
	turquoise (3,543)	-0.067 (0.355)	ECM core (2.27), ECM stiffness (2.39)

	black (80)	+0.161 (0.026)	Stiffness literature set (5.58), Mechanotransduction (9.97), ECM core (4.57)
	brown (1,479)	-0.161 (0.026)	—
Whole-genome network	B2.greenyellow	-0.249	—
	B1.pink	-0.234	—
	B2.purple (152)	+0.148	Stiffness literature set (3.93), Mechanotransduction (17.56), ECM core (8.05), ECM stiffness (10.37)

Seed gene network enrichment was evaluated using permutation *P values*; the black module–stiffness literature set comparison yielded an FDR of 0.011. Whole-genome module enrichment *P values* were uniformly 0.001, corresponding to the permutation test's lower resolution limit. Throughout this table an em dash (—) indicates that no matrix or mechanotransduction gene set showed significant overlap with that module (permutation $P \geq 0.05$); odds ratios are reported only for comparisons meeting this threshold.

Enrichment analyses for circadian rhythm gene sets yielded negative findings: among 126 comparisons comprising 7 modules and 18 gene sets in the seed gene network, no circadian rhythm comparison survived multiple-testing correction; similarly, no circadian gene set survived correction in the 39-module whole-genome network, where the strongest comparison, B2. blue against REACTOME Circadian Clock, yielded a permutation *P value* of 0.013 and an FDR of 0.174. Module assignments, module–trait statistics and all network enrichment results are given in Table S5.

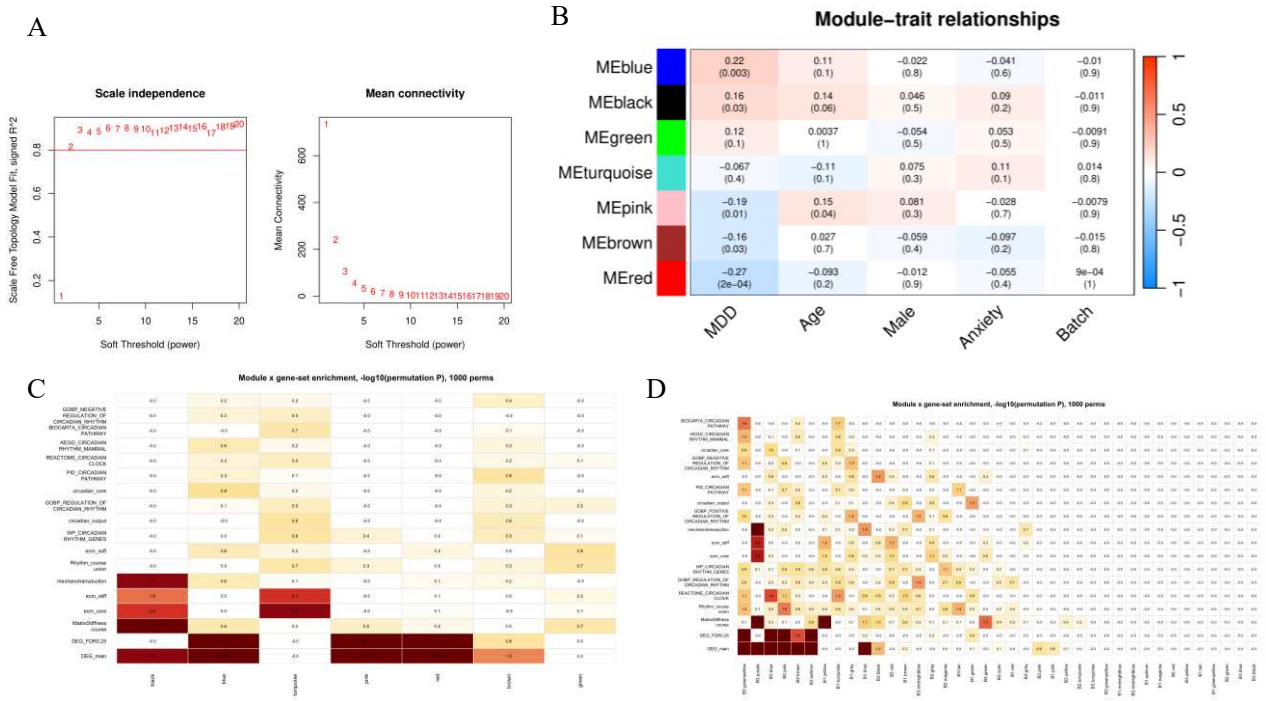


Figure 5. Co-expression network analysis.

(a) Goodness of fit for scale independence observed using the soft-thresholding power function. (b) Heatmap of seed gene network module-trait correlations after batch correction. (c) Heatmap of module-gene set enrichment in the seed gene network (-log10 permutation *P value*). (d) Heatmap of module-gene set enrichment in the whole-genome network (-log10 permutation *P value*).

5.3 Sample-level scoring and gene set definitions

The preceding comparisons were repeated at the sample level: disease-minus-control score differences were +0.1296 for the ECM core set ($P = 4.4 \times 10^{-5}$), +0.1344 for the ECM stiffness set ($P = 7.9 \times 10^{-4}$), +0.0448 for the matrix stiffness literature set ($P = 0.029$), and +0.0772 for the mechanotransduction set ($P = 0.036$); corresponding differences were +0.0076 for the circadian pathway set (broad definition; $P = 0.588$) and -0.0295 for the core clock set ($P = 0.386$) (Figure 6a, Table 3). Because stiffness-associated scores correlated strongly with estimated cell-type abundance, the same comparisons were repeated with adjustment for the seven cell-type scores together with age, sex, batch and anxiety comorbidity: the ECM core (β = 0.079, $P = 0.0076$) and ECM stiffness (β = 0.087, $P = 0.025$) effects persisted, whereas the matrix stiffness literature (β = 0.018, $P = 0.28$) and mechanotransduction (β = 0.004, $P = 0.91$) effects did not. Estimated cell-type scores alone explained 52% and 56% of the variance of the ECM stiffness and matrix stiffness literature scores, respectively, with CD4 T-cell and platelet scores the dominant contributors (Table S7).

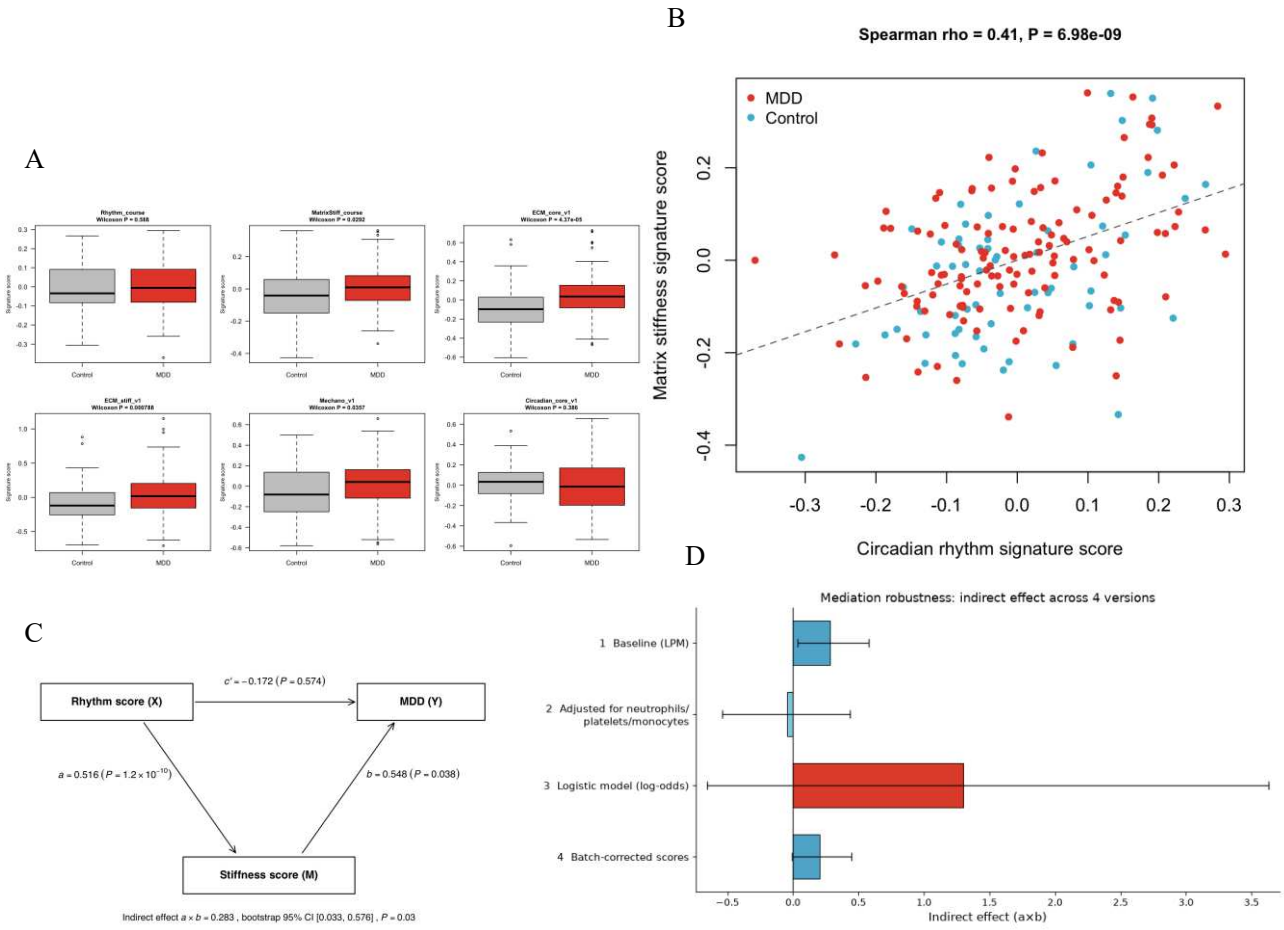


Figure 6. Signature scores and exploratory path analysis.

(a) Between-group box plot comparisons of six signature scores (Wilcoxon rank-sum test). (b) Scatter plot relating the broad-scope rhythm score to the stiffness score (Spearman ρ = 0.41). (c) Path diagram of the pre-specified

path model, shown for completeness. (d) Indirect-effect estimates and 95% CI for four model specifications. Panels (c) and (d) are descriptive and are not interpreted as evidence of mediation.

Analysis of correlations among the signature scores showed that the broad-scope rhythm pathway set correlated withthe stiffness score at Spearman ρ = 0.406 ($P = 7.0 \times 10^{-9}$), with the ECM stiffness set at 0.415, and with the mechanotransduction set at 0.425; in comparison, the core clock set, comprising 18 detectable genes, was negatively correlated with the stiffness score (ρ = -0.458, $P = 3.3 \times 10^{-11}$) while showing no between-group difference (difference -0.0295, $P = 0.386$). Two quantities must therefore be distingui*shed: the* betwe*en-gr*oup comparison, in which the core clock set was null, and the inter-score correlation, in which it was significant and opposite in sign to that of the broad rhythm set. The broad rhythm set and the core clock set were themselves essentially uncorrelated (ρ = 0.059, $P = 0.42$), indicating that the positive correlation between the broad rhythm score and stiffness is attributable to non-clock components of that set rather than to the molecular clock itself; these components include immune and metabolic genes such as IL6, PTEN and HIF1A, which are downstream targets or regulated elements rather than core clock genes.

Table 3. Between-group comparisons and inter-score correlations of signature scores (GSE98793)

Gene set (number of genes)	Between-group difference	*P* value	Correlation with stiffness score
ECM core set (80)	+0.1296	4.4×10^{-5}	—
ECM stiffness set (47)	+0.1344	7.9×10^{-4}	—
Matrix stiffness literature set (1314)	+0.0448	0.029	—
Mechanotransduction set (29)	+0.0772	0.036	—
Circadian pathway set (broad definition, 246)	+0.0076	0.588	0.406 ($P = 7.0 \times 10^{-9}$)
Core clock set (18)	-0.0295	0.386	-0.458 ($P = 3.3 \times 10^{-11}$)

Differences represent the mean standardised score in the disease group minus the mean standardised score in the control group. Between-group comparisons were performed using the Wilcoxon rank-sum test, and associations were assessed using Spearman rank correlation. Correlation coefficients are reported only for circadian gene sets, with *P* values from Spearman rank correlation tests; an em dash (—) indicates that a comparison was not applicable, because correlations were computed only between circadian scores and the stiffness score.

5.4 Exploratory path analysis and its robustness

The hypothesised circadian-to-matrix-to-disease pathway was not supported. In the baseline specification of the exploratory path analysis, path a was 0.516 ($P = 1.2 \times 10^{-10}$), path b was 0.548 ($P = 0.038$), the indirect effect was 0.283 (95% CI: 0.033 to 0.576; resampling $P = 0.030$), the total effect was 0.110 ($P = 0.69$), and the direct effect was -0.172 ($P = 0.574$); at α = 0.05, path a, path b, and the indirect effect were statistically significant, whereas neither the total nor the direct effect was (Figure 6c). None of the three robustness analyses supported the indirect effect (Figure 6d, Table 4): adjustment for cell composition scores yielded an indirect ef*fe*ct of -0.044 ($P = 0.88$); the logistic outcome model produced a CI ranging from -0.656 to 3.629 ($P = 0.16$) and generated a complete separation warning*;* analysis of batch-corrected scores yielded an indirect effect of 0.206 (95% CI: -0.008 to 0.444; $P = 0.062$). These results are reported as exploratory. Neither the total nor the direct effect was significant in the baseline model. The

indirect effect was opposite in direction and similar in magnitude, a pattern consistent with inconsistent mediation and suggesting that collinearity or confounding may influence the model specification. The widening of the confidence interval and the complete separation warning after replacing the linear probability model with logistic regression, given the binary outcome, further indicate the proposed pathway's sensitivity to model specification. Because the independent variable, the intermediate variable and the outcome were measured simultaneously in a case-control sample, this decomposition cannot distinguish the direction of effect; the observed pattern is consistent with a misspecified or confounded model rather than with mediation, and no causal claim is made from it anywhere in this paper.

Table 4. Exploratory path analysis and three robustness versions

Version	Path a	Path b	Indirect effect (95% CI)	*P* value
1 Baseline	0.516	0.548	0.283 (0.033 to 0.576)	0.030
2 Adjusted for cell composition scores	0.658	-0.067	-0.044 (-0.540 to 0.436)	0.88
3 Logistic outcome model	0.516	2.522	1.301 (-0.656 to 3.629)	0.16
4 Batch-corrected scores	0.357	0.576	0.206 (-0.008 to 0.444)	0.062

The baseline specification used a linear probability model; Path a and Path b *P* values were 1.2×10^{-10} and 0.038, respectively; the total effect was 0.110 ($P = 0.69$), and the direct effect was -0.172 ($P = 0.574$). Version 2 adjusted for neutrophil, platelet, and monocyte scores as covariates. Version 3 generated a complete separation warning. Reported as an exploratory analysis. Sample-level scores, between-group comparisons and all four path specifications are given in Table S6.

5.5 Cellular carriers of the matrix signal

Genes associated with ECM biology and stiffness include the neutrophil granule proteins MMP8, MMP9, ELANE, and LCN2; RETN, OLFM4, and GZMK were also among the leading DEGs. Therefore, we evaluated whether higher signature scores were associated with changes in estimated peripheral blood cell abundance; these marker-gene scores were used because directly measured cell counts were not available in this dataset.Cellular co*mp*osition differed between groups: neutrophil scores incr*ea*sed in the disease group ($P = 0.002$), as did monocyte scores ($P = 0.038$), whereas CD4 and CD8 T-cell scores decreased ($P = 0.0024$ and 5.3×10^{-5}); platelet scores showed an upward trend ($P = 0.068$) (Figure 7a). Further examination of cell-score relationships with stiffness scores *s*howed that, despite higher disease-group neutrophil scores, the correlation with stiffness scores was weak ($\rho = 0.069$, $P = 0.34$; Figure 7b). Platelet scores were positively associated with stiffness scores ($\rho = 0.607$), whereas lymphocyte subsets showed inverse associations (CD4 $\rho = -0.555$, CD8 $\rho = -0.365$, natural killer cells $\rho = -0.285$) (Table 5).

A B

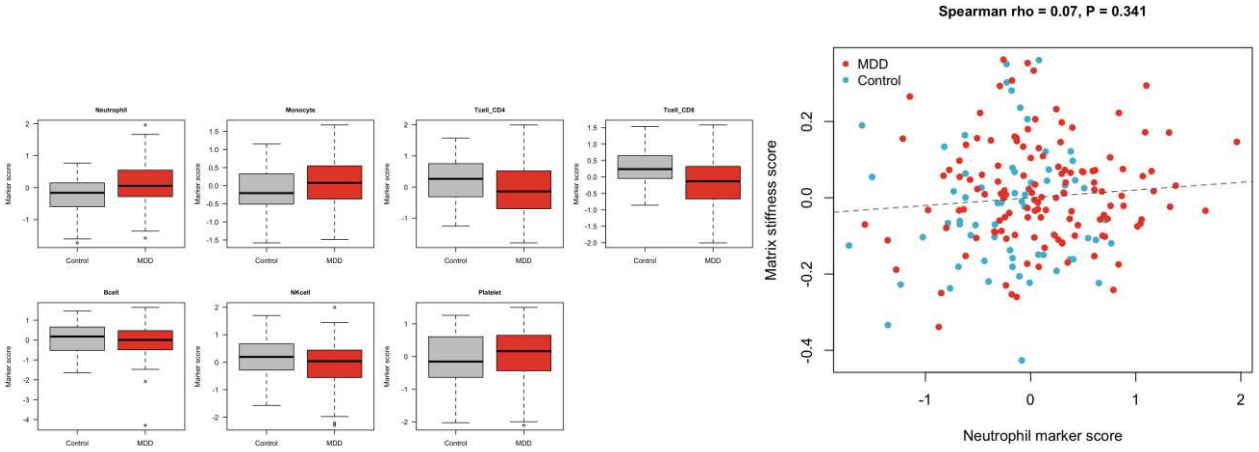


Figure 7. Relationship between peripheral blood cell composition and the matrix signature.

(a) Box plots comparing scores for seven blood cell markers across groups. (b) Scatter plot of stiffness scores vs neutrophil scores ($\rho = 0.069$, $P = 0.34$).

Platelet involvement is supported by independent cytological evidence: the "black" module of the seed gene network is organized around platelet-associated genes, including *PF4V1, ITGA2B, SELP*, and *CLU*, providing a cytological basis for the association between the stiffness-disease module and platelets. This observation also indicates that stiffness scores partly reflect platelet marker-gene abundance, which substantially influences the path analysis described in Table 4 (Version 2).

Table 5. Associations between peripheral blood cell type scores and stiffness/ECM scores

Cell type	vs. stiffness score (ρ, *P* value)	vs. ECM core set (ρ)
Platelets	0.607 ($<10^{-16}$)	0.606
Neutrophils	0.069 (0.34)	0.150
Monocytes	0.113 (0.119)	0.082
CD4 T cells	-0.555 ($<10^{-16}$)	-0.536
CD8 T cells	-0.365 (2.5×10^{-7})	-0.425
Natural killer cells	-0.285 (6.7×10^{-5})	-0.274
B cells	-0.095 (0.191)	-0.167

All cell-type quantities are marker-gene scores estimating relative abundance, not directly measured cell counts; no complete blood counts were available for this dataset. Associations were quantified using Spearman rank correlation.

5.6 Expression patterns and loss of rhythmicity in brain tissue

In control brain samples from the GSE45642 dataset, the nine gene-set signatures associated with rhythmicity and ECM/mechanotransduction differed substantially across the six brain regions (Kruskal-Wallis test *P values* ranging from 2.99×10^{-64} to 4.45×10^{-59}).

Case-control analyses yielded negative findings. In the GSE53987 dataset, MDD-associated DEGs numbered 1,443 in the prefrontal cortex, 1,393 in the striatum, and 630 in the hippocampus; after covariate adjustment, however, none of the nine gene-set signatures, including the highly differentially expressed blood gene set used for cross-tissue comparison, differed significantly between groups, and neither did the 18 core clock genes. Prefrontal cortex DEGs

showed nominal enrichment for the rhythmicity set (22 observed vs. 15 expected; OR = 1.52; permutation $P = 0.040$; FDR = 0.62). Cross-tissue analyses showed that absolute correlation coefficients between disease-associated blood $\log_2$ fold changes and those in the three GSE53987 brain regions were no greater than 0.09. Complete results for this cohort are given in Table S8.

As sampling-time data were unavailable in these cohorts, rhythmicity was evaluated using TOD as the phase coordinate.

5.7 Cosine rhythmicity analysis using TOD as the coordinate

Using TOD from 55 control subjects as the phase coordinate and Fisher's combined probability analysis across the six brain regions, the most rhythmic genes, ARNTL (FDR = 4.0×10^{-6}), PER2, NR1D1, PER3, DBP, and PER1, were identified, consistent with the ranking in the original report [6]. Of the 27 most rhythmic genes reported in the original study, this analysis recovered 22 (OR = 185.5; permutation $P = 0.001$) (Figures 8a, 8b, and 8c). This combined result was used only for cross-region ranking and visualization and does not support statistical inference. Per-gene rhythmicity statistics and phase estimates are given in Table S9.

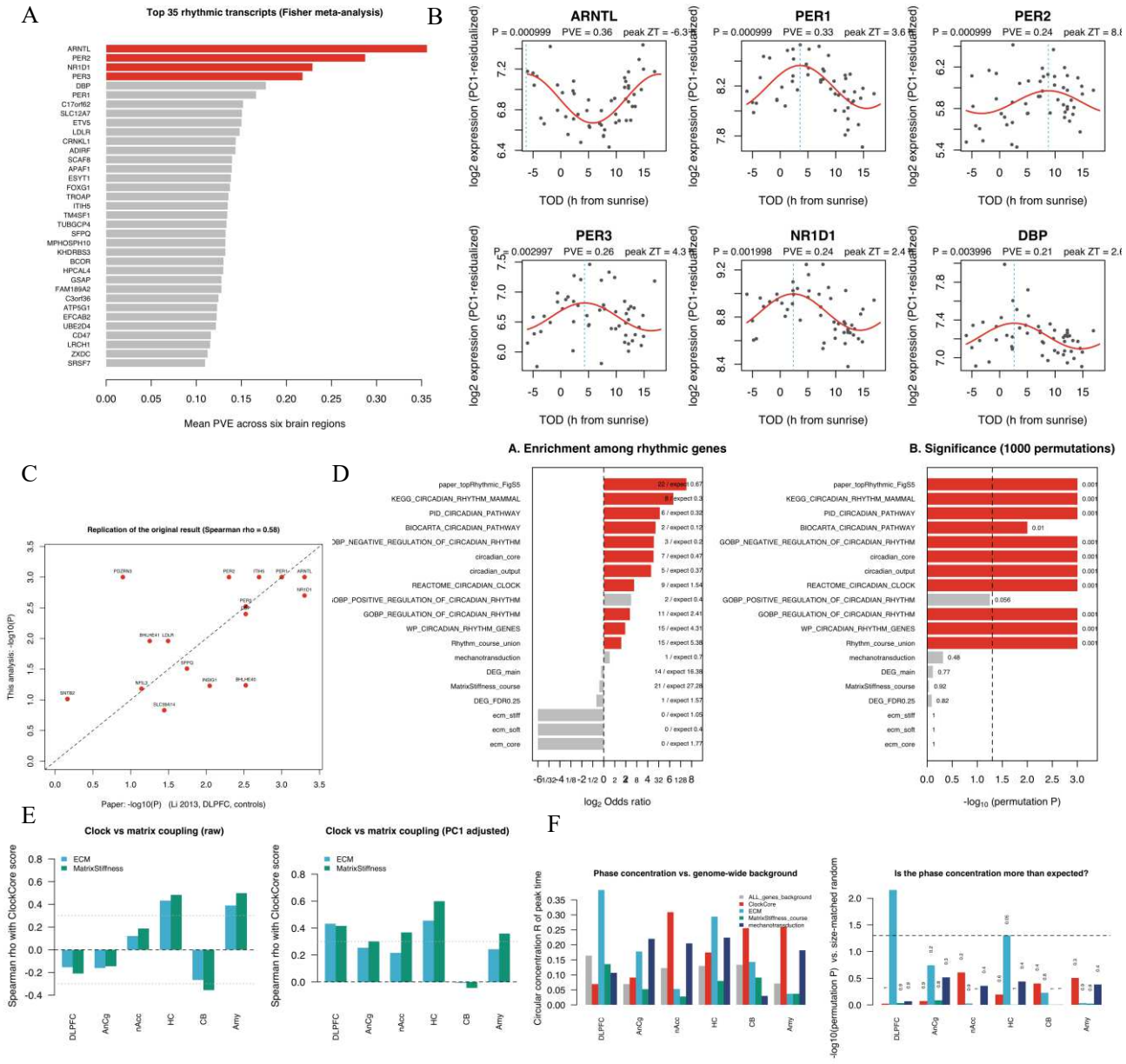


Figure 8. Cosinor rhythmicity analysis of control brain tissue.

(a) The 35 transcripts explaining the highest proportion of variance across six brain regions (Fisher's combined analysis); clock genes are highlighted in red. (b) Cosinor fit for six core clock genes in the dorsolateral prefrontal cortex. (c) Gene-by-gene validation comparing this analysis with the original report; 22 of the 27 most rhythmic genes were reproduced. (d) Gene-set rhythmicity enrichment; left: odds ratio (OR); right: permutation *P value*. (e) Coupling of clock and matrix signatures across brain regions; left: unadjusted correlation; right: correlation after adjustment for the first principal component. (f) Background phase control; phase concentration before and after preprocessing, and the significance of the match in relative magnitude compared to random gene sets.

Using the same analytical pipeline, rhythmicity enrichment was not detected for ECM or mechanotransduction gene sets (Figure 8d, Table 6): no rhythmic genes were identified in the ECM stiffness set (42 genes), ECM core set

(71 genes), or ECM low-stiffness set (16 genes), compared with expected counts of 1.05, 1.77, and 0.40, respectively; the matrix stiffness literature set (1,096 genes) contained 21 rhythmic genes against an expectation of 27.28 (OR = 0.75); only one rhythmic gene was identified in the mechanotransduction set, which was not statistically significant. As reference comparisons within the same testing series, the OR was 80.2 for the KEGG mammalian circadian rhythm set and 23.4 for the core clock set. Accordingly, ECM and mechanotransduction gene sets do not exhibit 24-hour rhythmicity in human brain tissue.

Sporadic rhythmicity was detected in individual mechanosensor genes: *CD47, CAV1, LATS1, ITGA9*, and *GPC3* reached nominal significance but did not survive FDR correction; no matrix synthesis gene reached nominal significance in the same series of tests. Coupling between the clock and matrix signatures was restricted to specific brain regions: stable associations were observed only in the hippocampus ($\rho = 0.433$, $P = 0.0023$; $\rho = 0.454$, $P = 0.0013$ after first-principal-component adjustment), whereas cortical and other regional correlations either became positive only after correction or remained negative (Figure 8e). Correlations between the ECM and stiffness sets were consistently high across all brain regions ($\rho = 0.92$ to 0.98).

In expression matrices without the original preprocessing steps, ECM gene peak phases were strongly concentrated; resultant vector lengths (R) ranged from 0.416 to 0.700 across the six brain regions, and Rayleigh test *P values* were as low as 7.6×10^{-15}. After principal component residual regression and sex-stratified median centring, R values decreased to 0.037 to 0.383; only the dorsolateral prefrontal cortex (R = 0.383, permutation $P = 0.007$) and hippocampus (R = 0.294, permutation $P = 0.050$) retained higher concentration than random gene sets of similar size. The remaining four regions were indistinguishable from background. This widespread loss of concentration after preprocessing indicates that the initial phase clustering was driven primarily by microarray-level technical bias (Figure 8f).

Table 6. Rhythmic enrichment of gene sets in control brain tissue

Gene set	Genes	Rhythmic genes (expected value)	OR	Permutation *P* value	FDR
Top 27 rhythmic genes from original study	27	22 (0.67)	185.5	0.001	3.2×10^{-30}
KEGG Mammalian Circadian Rhythm	12	8 (0.30)	80.2	0.001	5.8×10^{-10}
PID Circadian Rhythm Pathway	13	6 (0.32)	34.2	0.001	1.6×10^{-6}
Core Clock Set	19	7 (0.47)	23.4	0.001	1.4×10^{-6}
Clock Output Set	15	5 (0.37)	19.9	0.001	6.6×10^{-5}
REACTOME Circadian Clock	62	9 (1.54)	6.83	0.001	6.6×10^{-5}
Union of Rhythmic Pathway Sets	216	15 (5.38)	3.03	0.001	7.0×10^{-4}
Mechanotransduction Set	28	1 (0.70)	1.45	0.485	0.740
ECM Stiffness Set	42	0 (1.05)	0	1.00	1.00
ECM Core Set	71	0 (1.77)	0	1.00	1.00
ECM Low Stiffness Set	16	0 (0.40)	0	1.00	1.00
Matrix Stiffness Literature Set	1096	21 (27.28)	0.75	0.92	1.00
Blood differentially expressed gene set	658	14 (16.38)	0.84	0.766	1.00

Permutation *P values* in this table were derived from null distributions of random gene sets of the same size as the target set (1,000 resamplings). They are used solely for descriptive ranking across gene sets and differ from the criteria used for formal inference (e.g., camera or fry rotation tests); thus, they do not establish conclusions about statistical significance. FDR values reflect Benjamini-Hochberg adjustment of Fisher's exact test *P values* within the circadian gene-set family. Permutation *P values* for all members of this family reached the lower limit of testing resolution (0.001). Differences in gene-set size across datasets show platform-specific restrictions on the number of detectable genes.

5.8 Diagnostic specificity of altered matrix programmes

All 235 samples in the GSE92538 dataset were obtained from the dorsolateral prefrontal cortex and were analysed on a common platform; including BD and SCZ samples enabled differentiation between shared and diagnosis-specific features.

Differential expression showed clear diagnostic separation: MDD was associated with 273 DEGs (226 upregulated and 47 downregulated) (Figure 9a). Among them, 234 were MDD-specific, while only four were shared across all three diagnoses. Log-fold-change correlation coefficients among the four models ranged from 0.775 to 1.000; the model adding suicide history showed the greatest deviation (0.775 for all genes and 0.888 for DEGs), reducing the number of DEGs from 273 to 107. Mean core clock gene expression was unchanged: among the 24 core clock genes detectable in the dataset, only HLF attained nominal significance (P = 0.033, FDR = 0.51). The 24 individually examined core clock genes represent a separate gene list from the "core clock set" used for enrichment analysis, whose size varies with the number of genes detectable on the platform.

Gene-set test findings are presented in Table 7 and Figure 9. In MDD, the ECM stiffness programme showed coordinated changes with consistent directionality at nominal significance: 92.9% of test statistics for the 42 genes were negative (mean t = -0.922; camera test P = 0.0445; fry test P = 0.0439); the ECM core set showed similarly concordant changes (90.1% of 71 genes were negative; mean t = -0.850; *P values* for the corresponding tests were 0.0454 and 0.0444). For the broader matrix stiffness literature set (1,093 genes), the mean t-statistic was -0.325 and did not attain significance (P = 0.163). FDR adjustment used the Benjamini-Hochberg procedure, separately by diagnostic group, across the family of ten gene sets within each group; in the MDD family, camera_FDR and fry_FDR for the stiffness and core sets were 0.227 and 0.222, respectively, and neither retained significance after within-family correction. This pattern is supported by concordance in effect direction (92.9% negative test statistics) and by the contrast across diagnoses (absence of this pattern in BD and SCZ).

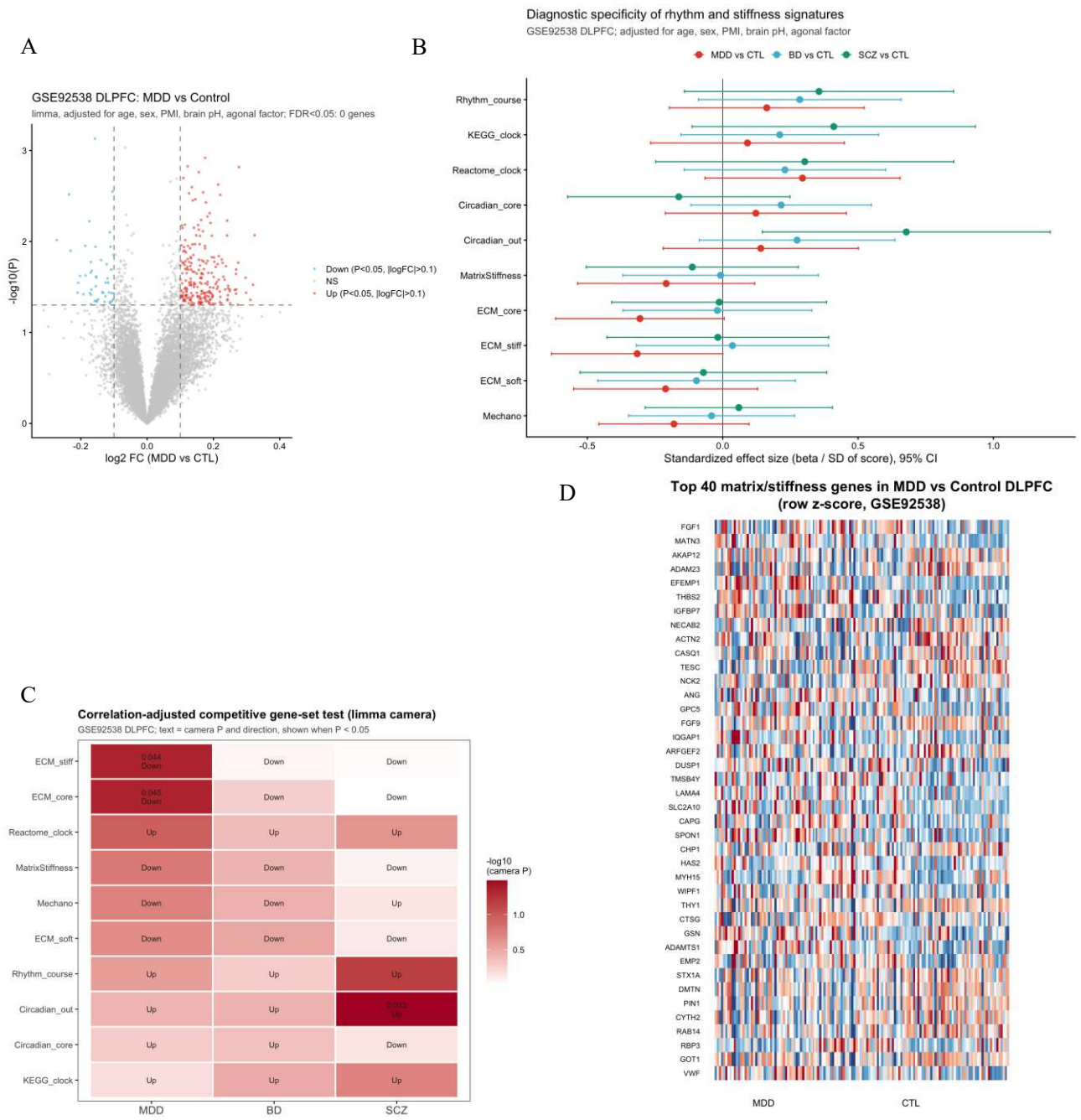


Figure 9. Diagnostic specificity of matrix programme alterations.

(a) Volcano plot of differential expression in MDD relative to controls (GSE92538). (b) Effect estimates for diagnostic specificity of rhythm and stiffness signatures (standardised coefficients and 95% CIs). (c) Heatmap of camera rotation test findings for ten gene sets across three diagnoses. (d) Expression heatmap (row z-scores) comparing the 40 most highly expressed matrix and stiffness genes between MDD and controls.

No significant changes were detected in BD across the ten gene sets; the mean t-statistic for its ECM stiffness set was +0.086, with 47.6% of t-values below zero. The only positive SCZ finding was upregulation of the rhythm output set (mean t = +0.692, P = 0.033), while its ECM stiffness set had a mean t-statistic of -0.112. Directionally

concordant change was confined to MDD (Figures 9b, 9c). Figure 9d presents expression profiles of the 40 most highly expressed matrix and stiffness genes in MDD compared with controls.

After adjusting for marker scores representing six cell types, namely neurons, oligodendrocytes, astrocytes, microglia, endothelial cells and erythrocytes, the ECM core and stiffness sets retained concordant alterations (P = 0.012 and 0.017, without multiple-testing correction); the standardised coefficients for the ECM stiffness set changed from -0.1495 to -0.0996, and for the ECM core set from -0.1387 to -0.1015. In MDD brain tissue, neuronal scores were increased (difference +0.413, P = 0.004), whereas astrocyte scores (-0.338) and endothelial cell scores (-0.445) were decreased; the ECM signature correlated negatively with neuronal scores (ρ = -0.63) and positively with astrocyte and endothelial cell scores. Cross-cohort analysis showed a Spearman correlation coefficient of +0.101 (n = 10,915, $P = 3.5 \times 10^{-26}$) between all-gene log-transformed fold changes in the prefrontal cortex (GSE53987) and dorsolateral prefrontal cortex (GSE92538); this correlation supports only the reproducibility of effect direction across brain regions, explains less than 1% of variance, and is insufficient as independent validation of the underlying molecular programme itself. The correlation coefficient for disease-associated log-transformed fold changes between peripheral blood (GSE98793) and dorsolateral prefrontal cortex (GSE92538) was -0.195. Intracerebral rhythmicity was not associated with disease effect size (correlation coefficient: 0.069; comparison of absolute effect sizes between rhythmic and non-rhythmic genes: Wilcoxon test P = 0.785).

Table 7. Rotation test results for gene sets across three psychiatric diagnoses (GSE92538, dorsolateral prefrontal cortex)

Diagnosis	Gene set (number of genes)	Mean t	Proportion with t < 0	camera/fry *P* value
MDD	ECM Stiffness Set (42)	-0.922	92.9%	0.0445/0.0439
MDD	ECM Core Set (71)	-0.850	90.1%	0.0454/0.0444
MDD	Matrix Stiffness Literature Set (1093)	-0.325	69.4%	0.163/0.184
MDD	Mechanotransduction Set (27)	-0.420	70.4%	0.187/0.171
MDD	REACTOME Circadian Clock (60)	+0.418	35.0%	0.110/0.133
MDD	Core Clock Set (17)	+0.174	47.1%	0.510/0.558
BD	ECM Stiffness Set (42)	+0.086	47.6%	0.861/0.869
SCZ	Clock Output Set (15)	+0.692	33.3%	0.033/0.024
SCZ	ECM Stiffness Set (42)	-0.112	52.4%	0.943/0.773

"Mean t" shows the average disease-effect statistic for genes within the set, and "Proportion with t < 0" shows the percentage of genes with negative statistics; "camera" and "fry" are rotation-based tests that account for inter-gene correlations; the other nine gene sets examined in BD were non-significant; for MDD, the four sets not listed (broad rhythm set, clock output set, ECM low-stiffness set and KEGG clock set) were likewise non-significant (P > 0.2). FDR adjustment used the Benjamini-Hochberg method within each diagnostic group's family of ten gene sets: in MDD, the ECM stiffness set yielded camera_FDR = 0.227 and fry_FDR = 0.222, and the ECM core set yielded camera_FDR = 0.227 and fry_FDR = 0.222; in SCZ, the clock output set yielded camera_FDR = 0.329 and fry_FDR = 0.240. All other tests had FDRs > 0.3; therefore, no gene set remained significant within its diagnostic group after within-family adjustment. As the ECM stiffness set was the study's primary hypothesis, established a priori, a single pre-specified test might not require within-family correction; results are presented under both approaches to maintain

transparency. Complete camera and fry results for all ten gene sets in the three diagnostic groups are given in Table S10.

5.9 Robustness of the matrix programme to adjustment for cellular composition

Testing at set level rather than score level changes which findings survive adjustment. With covariates entered directly into the design matrix, the ECM core set retained a coherent positive shift: across its 80 genes the mean moderated t-statistic was +0.47 and 68.8% of statistics were positive, with a rotation test *P* value of 0.020. The ECM stiffness set behaved likewise: across 47 genes the mean statistic was +0.52, 74.5% were positive, $P = 0.046$. The matrix stiffness literature set did not; over 1,314 genes the mean statistic fell to +0.11 and the set test gave $P = 0.55$, having been $P = 0.071$ without adjustment, and the mechanotransduction set gave $P = 0.78$ over 29 genes, having been $P = 0.073$. At score level the effect estimates of these two sets attenuated by 87% and 88%. Survival therefore did not track the size of the set, since the largest set was among the first to fail. Nor did it track the burden of individually significant genes: within the adjusted model, 6.25% of ECM core genes and 4.26% of ECM stiffness genes reached $P < 0.05$, against 5.77% of all detectable genes in the same model, so the surviving sets were not enriched for nominal significance and their results reflect coordinated direction rather than a few strong members. Leave-one-gene-out sensitivity qualified the conclusion. Removing COL6A2 moved the ECM core *P value* from 0.020 to 0.031 and the stiffness *P value* from 0.046 to 0.079; removing COL6A2 together with LOXL2 gave 0.042 and 0.117, and further removals carried the stiffness set to 0.26. The adjusted result therefore rests on a small number of collagen and cross-linking genes. This is informative rather than merely limiting, because COL6A2 and LOXL2 are matrix-deposition and cross-linking effectors rather than generic markers of cell composition. Set-level adjustment supports the ECM core finding and renders the stiffness finding provisional. Complete set-level results, per-gene statistics and the full leave-one-gene-out sequence are given in Table S11. (Figure 10)

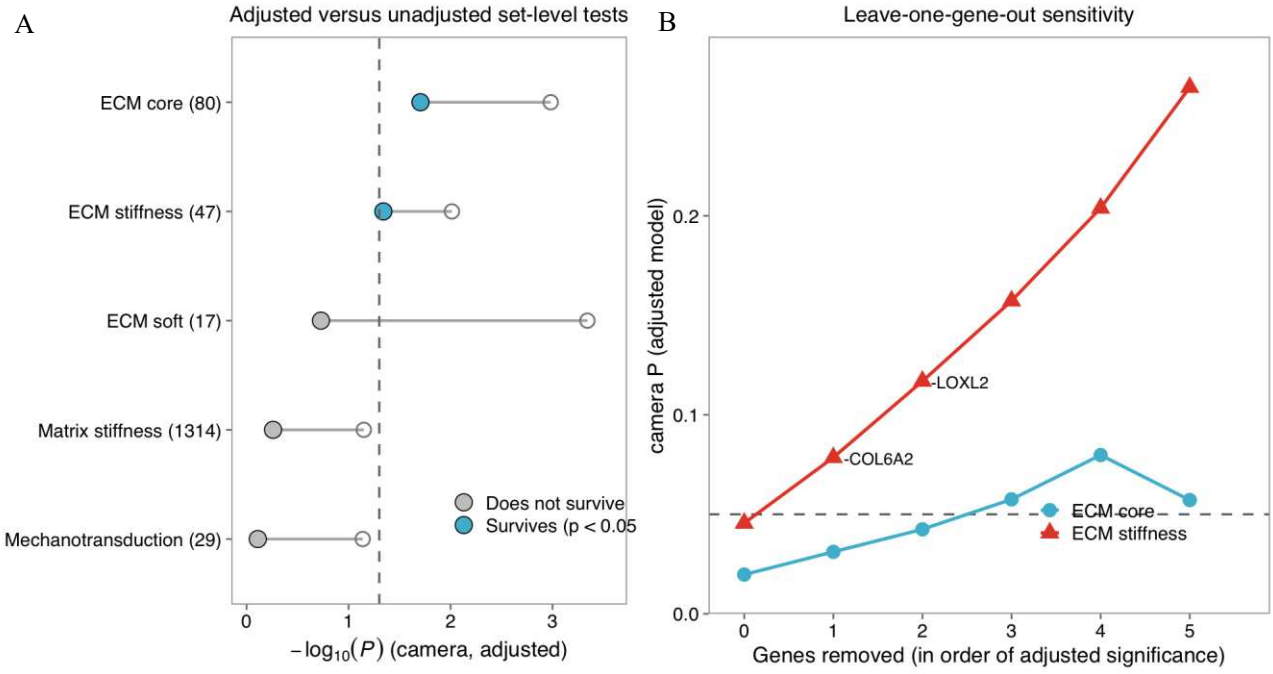


Figure 10. Set-level robustness of the matrix programme to adjustment for cellular composition.

(a) Camera *P* values for five gene sets before and after entering estimated cell-type composition, batch, age, sex and anxiety comorbidity directly into the design matrix (whole blood, GSE98793). Open circles: unadjusted model; filled circles: adjusted model; blue: significant after adjustment ($P < 0.05$); grey: not significant. Dashed line: $P = 0.05$. (b) Leave-one-gene-out sensitivity for the ECM core and stiffness sets. Genes were removed in order of their adjusted single-gene significance and the set-level test was recomputed after each removal; the camera *P* of the stiffness set crosses 0.05 once COL6A2 is removed and rises further with LOXL2, whereas the core set remains at or below 0.05 until three genes have been removed.

5.10 Direct contrasts between diagnoses and pooling across brain cohorts

Specificity was tested by direct contrast rather than inferred from differing significance statuses. In GSE92538, which contributes 91 controls, 47 MDD and 42 bipolar subjects from one platform and one brain region, the ECM stiffness score was reduced in MDD ($\beta = -0.22$, $P = 0.013$) but not in bipolar disorder ($\beta = -0.02$, $P = 0.80$). The direct MDD-minus-bipolar contrast was nevertheless not significant (difference -0.20, standard error 0.105, $P = 0.057$; $P = 0.077$ with estimated cell-type scores added), and the ECM core contrast was not significant either (difference -0.16, $P = 0.098$). At set level the contrast gave $P = 0.125$ by camera and 0.065 by fry for the stiffness set, and $P = 0.216$ and 0.111 for the core set, with 83.3% and 80.3% of statistics negative. The direction is consistent throughout; the power is not. Pooling did not resolve this. Combining the three brain regions of GSE53987 with the dorsolateral prefrontal cortex of GSE92538, giving four independent estimates, the pooled MDD-minus-bipolar difference was -0.073 for the ECM stiffness set (standard error 0.066, $P = 0.27$; $I^2 = 41\%$), -0.058 for the ECM core set ($P = 0.30$; $I^2 = 31\%$) and -0.089 for the mechanotransduction set ($P = 0.080$; $I^2 = 0\%$). Heterogeneity was low for the mechanotransduction set and moderate for the others, so the null result is not explained by inconsistency between cohorts; the estimates are small relative to their standard errors. Diagnostic specificity of the matrix programme is therefore not supported by these data. What the data support is narrower and still worth stating: within one brain region, one platform and one cohort, the programme is altered in MDD and unchanged in bipolar disorder. Contrast estimates with standard errors, per-region estimates and heterogeneity statistics are given in Table S12. (Figure 11)

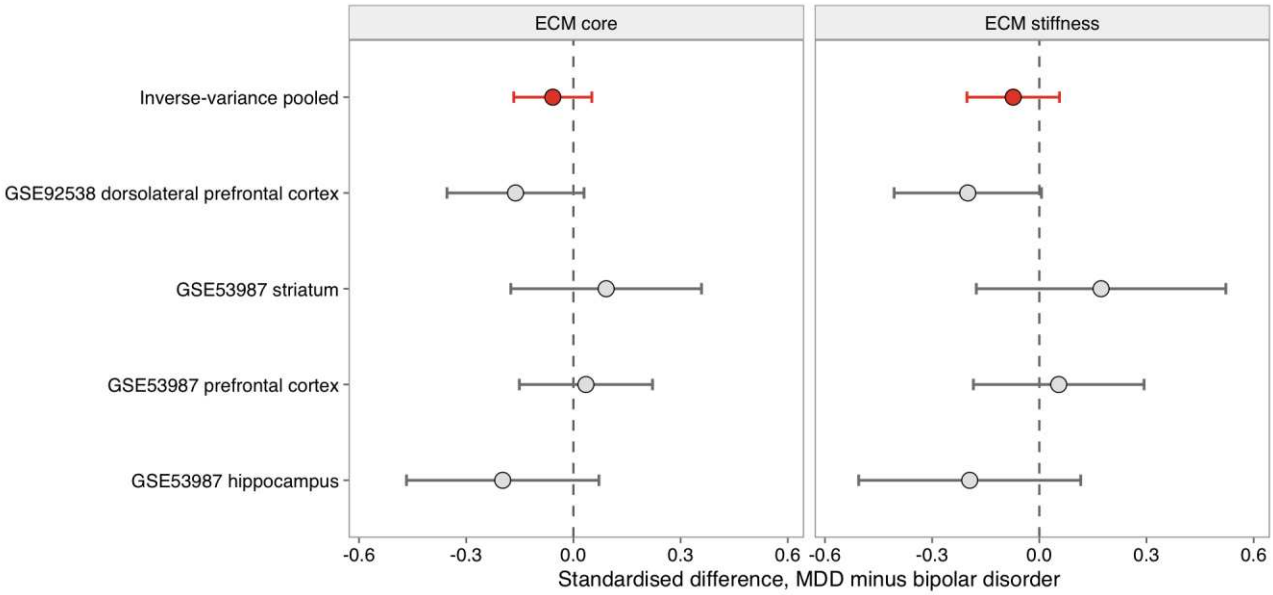


Figure 11. Direct MDD-versus-bipolar contrasts across brain cohorts.

Standardised score differences (MDD minus bipolar disorder) with 95% confidence intervals, adjusted for age, sex, post-mortem interval and brain pH. Each GSE53987 region is treated as a separate estimate because individual subjects contributed tissue to more than one region; red marks the inverse-variance pooled estimate across the four region-level estimates. Neither pooled contrast reached significance, and the pooled estimate is small relative to the between-cohort spread.

5.11 Independent evidence from common genetic variation

Concordance among transcriptomic cohorts cannot establish independence, because the cohorts share modality and largely share platform, preprocessing and analytical pipeline. Gene-set analysis of depression genome-wide association summary statistics from FinnGen release 12 (F5_DEPRESSIO, 59,333 cases) supplied evidence differing in modality as well as in population. After stratifying background genes into twenty bins by variant count so that the score was not confounded by gene length, no ECM set showed enrichment of common-variant association: the ECM core set gave a mean quantile of 0.456 (z = -1.34, empirical $P = 0.18$), the stiffness set 0.461 (z = -0.90, $P = 0.36$), the soft-matrix set 0.446 (z = -0.78, $P = 0.45$), the broad literature set 0.507 (z = +0.92, $P = 0.34$) and the mechanotransduction set 0.524 (z = +0.45, $P = 0.65$). Negative controls did not behave as inert comparators but as an index of sensitivity: the core clock set gave the largest positive deviation of any set examined (mean quantile 0.610, z = +1.62, $P = 0.10$), followed by the KEGG clock set (z = +1.04, $P = 0.29$). The pipeline can therefore register a signal where one exists, and the absence of enrichment across the ECM sets is evidence of absence at the resolution these data afford rather than an artefact of the procedure. Two readings remain consistent with this result. The ECM transcriptional programme of MDD may be a downstream or state-dependent phenomenon with little common-variant determination, or its genetic contribution may be distributed too thinly across the sets as defined here for a set-level statistic to detect. Neither reading supports using common-variant enrichment as corroboration of the transcriptomic

finding, and the result is reported as negative. Per-gene scores, variant-count stratification and permutation results are given in Table S13. (Figure 12)

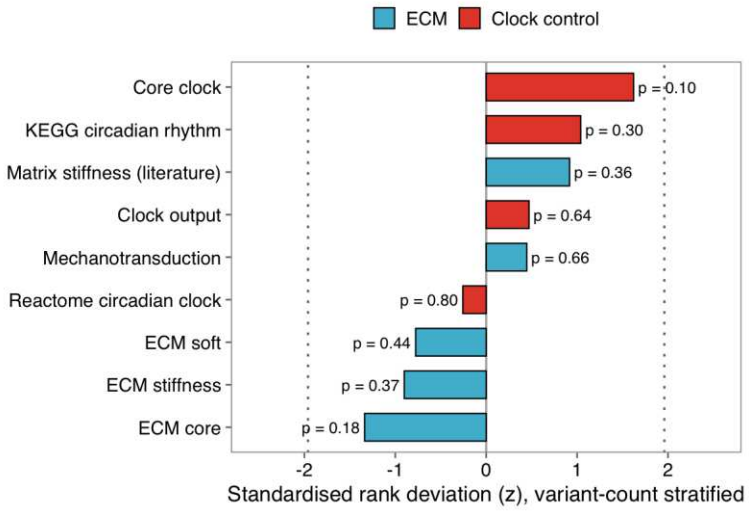


Figure 12. Common-variant gene-set analysis of external genome-wide association summary statistics.

Standardised rank deviation (z) for nine gene sets in FinnGen release 12 depression (59,333 cases), with background genes stratified into bins by assigned variant count so that gene length could not drive the result. Positive z indicates a stronger association than background genes of equal variant count; dotted lines mark z = ±1.96. No ECM set departs from the null, whereas the core clock set, an expected negative control that was also non-significant at the transcript level, shows the largest positive deviation in the analysis, indicating that the null reflects the data rather than the sensitivity of the pipeline.

6. Discussion

6.1 Key findings and their implications

The study tested three pre-specified predictions about the ECM stiffness transcriptional programme across four independent human transcriptomic cohorts and one external genome-wide association resource: that the programme is altered in MDD, that any alteration is specific to MDD rather than shared with bipolar disorder and schizophrenia, and that it is coupled to circadian transcriptional output. The association between the ECM stiffness programme and disease was examined at two complementary levels: peripheral blood showed increased co-expression modules and sample-level scores, whereas post-mortem dorsolateral prefrontal cortex showed a coordinated shift in the opposite direction. The two tissues therefore disagreed: disease-associated log-fold changes in blood and brain were negatively correlated (ρ = -0.195, 623 genes, $P = 1.0 \times 10^{-6}$), so a peripheral ECM transcript score cannot be read as a surrogate for central ECM state. Brain changes were identified only in the GSE92538 dataset; corresponding findings in the GSE53987 prefrontal cortex were not replicated after adjustment for covariates. Previous multi-omics studies have attributed MDD-associated molecular alterations to synaptic and immune pathways [48]; the matrix programme introduces a mechanical dimension to this molecular platform that can be evaluated within the same dataset.

The clinical and pathological relevance of the positive findings has two dimensions: robustness and scope. Observations previously dispersed across separate studies or gene sets are consolidated into a programme-level alteration that can be evaluated at several analytical levels, and the ECM core set, the ECM stiffness set, the mechanotransduction set and the broad literature-derived set show concordant directional trends across three frameworks: modules, scoring and rotation tests. Their fate under adjustment is not uniform, and the differences are informative. The broad literature-derived set was the most sensitive to adjustment for estimated cell composition, its effect estimate attenuating by 87% and its set-level test falling from $P = 0.071$ to 0.55, indicating that it contains many genes unrelated to the core programme and that part of its apparent signal tracks cell composition. The mechanotransduction set behaved similarly. The signal is concentrated in the smaller defined sets, and within those it is carried by collagen and cross-linking genes rather than distributed evenly, as the leave-one-gene-out results show. Within the same cohort, platform and brain region, the programme was altered in MDD while bipolar disorder showed no significant change in any of the ten gene sets and schizophrenia showed upregulation only of circadian output genes. This pattern excludes a shared response to psychiatric illness or neurological injury as a sufficient explanation. It does not establish diagnostic specificity: the direct MDD-versus-bipolar contrast did not reach significance in that cohort, and pooling with a second brain cohort did not rescue it. The defensible statement is therefore that the programme is altered in MDD and unchanged in bipolar disorder, not that it discriminates between them.

This interpretation aligns with recent cross-diagnostic comparative evidence indicating that shared genetic, cognitive, and neuroimaging characteristics of MDD, BD, and SCZ predominate over diagnosis-specific features, and that post-mortem brain studies rarely identify molecular programmes capable of discriminating these diagnoses [49]. Peripheral immunity is among the systems most frequently implicated in the literature; however, meta-analyses and reviews report small effect sizes, substantial heterogeneity, and confounding by leukocyte composition [50–52]. Whether the matrix programme can contribute to such discrimination remains unresolved: the alteration is detectable in MDD and absent in bipolar disorder, but the direct contrast between the two did not reach significance, and no enrichment was detected at the level of common genetic variation. The matrix programme should therefore be regarded as a candidate component of a multivariate marker rather than as a discriminator in its own right.

The methodological approach of inferring central nervous system states from peripheral signals parallels the traditional Chinese medicine (TCM) diagnostic principle of “inferring internal conditions from external manifestations” (si wai chuai nei). The Lingshu, specifically its “Waichuai” or “External Inference” chapter, states: “For distant phenomena, infer the internal from the external; for near phenomena, infer the external from the internal”; that is, observable external manifestations are used to inform assessment of internal physiological alterations. As central nervous system tissues cannot be sampled in living subjects, peripheral blood provides a principal means of investigating intracerebral processes. This approach parallels TCM, in which external signs, including pulse and complexion, inform assessment of the state of internal organs; in either framework, correspondence between external

manifestations and internal alterations cannot be assumed a priori and must be validated for each specific component. For investigators working across disciplines, the study's transferable contribution is its analytical approach: using co-expression modules and gene-set programmes as units of analysis and testing each conclusion's applicability with background controls and independent cohorts. The workflow is not restricted to a particular disease and can be used for different studies of inaccessible organ states using accessible tissues.

6.2 Potential mechanisms underlying changes in the matrix programme

Chronic stress enhances deposition of perineuronal nets (PNNs) around parvalbumin (PV) interneurons, whereas antidepressant treatment re-establishes excitation/inhibition balance and gamma oscillation power through matrix metalloproteinase-dependent degradation [11,12]. In this study, the programme showed coordinated changes in brain tissue, but in the opposite direction of this framework; at the transcriptional level, this could reflect reduced net components, enhanced degradation, or suppressed synthesis, processes that transcriptomic data cannot distinguish.

For cell-type localization, the brain matrix signature correlated negatively with neuronal scores ($\rho = -0.63$) and positively with astrocyte and endothelial cell scores. The peripheral blood signature correlated positively with platelet scores ($\rho = 0.607$) and negatively with CD4 T-cell scores ($\rho = -0.555$); estimated cell-type scores alone explained 52% of the variance of the stiffness score. These observations consistently link alterations in the matrix programme to specific cell populations. Glial and vascular cells are key contributors to PNN secretion and remodelling; substantial evidence implicates astrocytic dysfunction in MDD [53], components of the blood-brain barrier display intrinsic rhythmicity [54], and microglial functional states interact with circadian systems [55]. Platelets additionally contribute to peripheral matrix signalling through serotonin uptake, P-selectin expression, and release of matrix-associated granules, observations consistent with the established platelet-depression hypothesis [56,57]. These observations are consistent with a platelet contribution to peripheral matrix signalling, but the present cross-sectional data cannot establish that this contribution is driven by circadian change.

Control brain analyses also identified sporadic rhythmic expression of mechanosensor genes (e.g., *CD47, CAV1, LATS1, ITGA9, GPC3*), whereas no matrix synthesis gene showed nominally significant rhythmicity. This divergence suggests a link between the molecular clock and cellular mechanosensing, the direction of which the present data cannot establish. Through the integrin-actin cytoskeleton-MRTF/SRF pathway, matrix stiffness information can be conveyed to core clock repressors, including *PER2, NR1D1 and NFIL3,* altering the circadian rhythm. Direct evaluation of this mechanism could combine mechanical loading with clock reporter systems in cellular models.

Histological evidence does not fully align with these findings: Alcaide et al. reported reduced PNN density in the dorsolateral prefrontal cortex in patients with BD and SCZ but found no corresponding difference in patients with MDD [14]. Transcript abundance does not necessarily correlate linearly with ECM protein deposition or cross-linking. However, discordance across biological levels is expected; these discrepancies preclude describing the present results

simply as "downregulated matrix stiffness." The observed phenomenon reflects coordinated changes in transcriptional programme rather than a physical measurement that can be directly translated into tissue stiffness.

6.3 Implications of Negative Results Regarding the Circadian Axis

Using control brain tissue with established phase coordinates, the study reproduced the original report's list of highest-ranked circadian genes, with a recovery rate of 22/27 (OR = 185.5), showing that the circadian analysis pipeline can identify known circadian genes. In this case, the negative findings for ECM and mechanotransduction gene sets are informative: circadian regulation of matrix synthesis at the transcriptional level should yield a corresponding expression signature, yet none was detected.

Three additional observations support these negative findings: enrichment analyses of circadian gene sets in both peripheral blood networks were negative; mean core clock gene expression in brain tissue did not differ between disease and control groups; and disease-associated log-transformed fold changes in peripheral blood (GSE98793) correlated negatively (-0.195) with those in the dorsolateral prefrontal cortex (GSE92538). As mean core clock gene expression was unchanged in peripheral blood and brain tissue in the disease group, any circadian abnormalities would likely involve phase or amplitude, changes that cannot be resolved by tests based on expression abundance. Core molecular clock components undergo translational and post-translational regulation [20,21]; protein modifications and translational states can vary independently of mRNA abundance, providing a mechanistic explanation for the absence of detectable circadian phenotypes at the transcriptomic level. Therefore, negative findings from public peripheral blood datasets cannot be interpreted as evidence excluding peripheral circadian rhythm involvement, because these datasets lack sampling-time information and thus convert phase-associated differences into measurement noise.

Recent reviews examining circadian rhythms in relation to depression report that circadian assessments generally rely on behavioural scales, sleep diaries, hormonal profiles, or neuroimaging phenotypes [4,58,59]; molecular-level disease assessments incorporating actual phase coordinates remain uncommon. This study evaluated the relationship between circadian rhythms and the ECM at the transcriptomic level in post-mortem brain samples with documented times of death; however, phase data were unavailable for the disease group. This asymmetry defines the scope and limits of the study's conclusions.

7. Conclusion

Across four independent human transcriptomic cohorts and one external genome-wide association resource, three pre-specified predictions about the ECM stiffness transcriptional programme were tested. One was supported, one was not supported, and one was refuted in a way that changes how the finding should be used. The programme is altered in MDD: the association survived adjustment for estimated cellular composition at set level for the ECM core set ($P = 0.020$) and provisionally for the stiffness set ($P = 0.046$), although leave-one-gene-out sensitivity showed the latter to depend on COL6A2 and LOXL2. The programme is not diagnostically specific at the available sample size: it was altered in MDD and unchanged in bipolar disorder, but the direct contrast did not reach significance ($P =$

0.057) and pooling across brain regions and cohorts did not rescue it ($P = 0.27$). The programme is not coupled to circadian transcriptional output: with confirmed phase coordinates and a positive control recovering 22 of 27 known circadian genes, no 24 h rhythmicity was detected in ECM or mechanotransduction gene sets, and mean core clock expression was unchanged in both tissues.

Two further results constrain interpretation. Peripheral and central estimates of the same programme moved in opposite directions, with disease-associated log-fold changes negatively correlated between blood and brain ($\rho = -0.195$, 623 genes, $P = 1.0 \times 10^{-6}$), so a peripheral ECM transcript score is not a surrogate for central ECM state; transcript abundance does not in any case measure matrix protein deposition or cross-linking. Common-variant analysis supplied no corroboration, with no ECM set enriched for depression risk (all $P > 0.18$) while a clock-gene comparator gave the largest deviation observed, indicating that the null reflects the data rather than the pipeline. Because phase coordinates were available only for control brain, circadian involvement at the level of phase or amplitude cannot be excluded, and phase-resolved sampling of patient tissue is required before it can be. Three directions follow: identifying the cellular carriers of the signal, with platelets in blood and astrocytes and endothelial cells in brain as candidates; measuring matrix protein and cross-linking directly, since the present evidence is transcriptional; and testing clock regulation of mechanosensors through mechanical loading combined with circadian reporters in cellular models.

References

1. Santomauro DF, Vos T, Whiteford HA, Chisholm D, Saxena S, Ferrari AJ. Service coverage for major depressive disorder: estimated rates of minimally adequate treatment for 204 countries and territories in 2021. *Lancet Psychiatry*. 2024;11(12):1012-1021. doi:10.1016/S2215-0366(24)00317-1
2. Krishnan V, Nestler EJ. The molecular neurobiology of depression. *Nature*. 2008;455(7215):894-902. doi:10.1038/nature07455
3. Leday GGR, Vértes PE, Richardson S, et al. Replicable and coupled changes in innate and adaptive immune gene expression in two case-control studies of blood microarrays in major depressive disorder. *Biol Psychiatry*. 2018;83(1):70-80. doi:10.1016/j.biopsych.2017.01.021
4. de Leeuw M, Verhoeve SI, van der Wee NJA, van Hemert AM, Vreugdenhil E, Coomans CP. The role of the circadian system in the etiology of depression. *Neurosci Biobehav Rev*. 2023;153:105383. doi:10.1016/j.neubiorev.2023.105383
5. Fellinger M, Waldhör T, Serretti A, et al. Seasonality in Major Depressive Disorder: Effect of Sex and Age. *J Affect Disord*. 2022;296:111-116. doi:10.1016/j.jad.2021.09.051
6. Li JZ, Bunney BG, Meng F, et al. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. *Proc Natl Acad Sci U S A*. 2013;110(24):9950-9955. doi:10.1073/pnas.1305814110

7. Saraswathibhatla A, Indana D, Chaudhuri O. Cell–extracellular matrix mechanotransduction in 3D. *Nat Rev Mol Cell Biol*. 2023;24(7):495-516. doi:10.1038/s41580-023-00583-1
8. Hynes RO. Extracellular matrix: not just pretty fibrils. *Sci (N Y NY)*. 2009;326(5957):1216-1219. doi:10.1126/science.1176009
9. Reichelt AC, Hare DJ, Bussey TJ, Saksida LM. Perineuronal Nets: Plasticity, Protection, and Therapeutic Potential. *Trends Neurosci*. 2019;42(7):458-470. doi:10.1016/j.tins.2019.04.003
10. Favuzzi E, Marques-Smith A, Deogracias R, et al. Activity-Dependent Gating of Parvalbumin Interneuron Function by the Perineuronal Net Protein Brevican. *Neuron*. 2017;95(3):639-655.e10. doi:10.1016/j.neuron.2017.06.028
11. Spijker S, Koskinen MK, Riga D. Incubation of depression: ECM assembly and parvalbumin interneurons after stress. *Neurosci Biobehav Rev*. 2020;118:65-79. doi:10.1016/j.neubiorev.2020.07.015
12. Blanco I, Conant K. Extracellular Matrix Remodeling with Stress and Depression: Studies in Human, Rodent, and Zebrafish Models. *Eur J Neurosci*. 2021;53(12):3879-3888. doi:10.1111/ejn.14910
13. Théberge S, Belliveau C, Xie D, et al. Parvalbumin interneurons in human ventromedial prefrontal cortex: a comprehensive post-mortem study of myelination and perineuronal nets in neurotypical individuals and depressed suicides with and without a history of child abuse. *Cereb Cortex (N Y NY)*. 2024;34(5):bhae197. doi:10.1093/cercor/bhae197
14. Alcaide J, Guirado R, Crespo C, et al. Alterations of perineuronal nets in the dorsolateral prefrontal cortex of neuropsychiatric patients. *Int J Bipolar Disord*. 2019;7:24. doi:10.1186/s40345-019-0161-0
15. Xiong X, Li W, Nam J, Qu M, Kay SA, Ma K. The actin cytoskeleton-MRTF/SRF cascade transduces cellular physical niche cues to entrain the circadian clock. *J Cell Sci*. 2022;135(19):jcs260094. doi:10.1242/jcs.260094
16. Dudek M, Gossan N, Yang N, et al. The chondrocyte clock gene Bmal1 controls cartilage homeostasis and integrity. *J Clin Invest*. 2015;126(1):365-376. doi:10.1172/JCI82755
17. Yeung CYC, Garva R, Pickard A, et al. Mmp14 is required for matrisome homeostasis and circadian rhythm in fibroblasts. *Matrix biol : j Int Soc Matrix Biol*. 2023;124:8-22. doi:10.1016/j.matbio.2023.10.002
18. Pantazopoulos H, Gisabella B, Rexrode L, et al. Circadian Rhythms of Perineuronal Net Composition. *eNeuro*. 2020;7(4):ENEURO.0034-19.2020. doi:10.1523/ENEURO.0034-19.2020
19. Partch CL, Green CB, Takahashi JS. Molecular architecture of the mammalian circadian clock. *Trends Cell Biol*. 2014;24(2):90-99. doi:10.1016/j.tcb.2013.07.002
20. Zhuang Y, Li Z, Xiong S, et al. Circadian clocks are modulated by compartmentalized oscillating translation. *Cell*. 2023;186(15):3245-3260.e23. doi:10.1016/j.cell.2023.05.045
21. Lyu J, Zhuang Y, Lin Y. Circadian regulation of translation. *RNA Biol*. 2024;21(1):959-969. doi:10.1080/15476286.2024.2408524

22. Clough E, Barrett T, Wilhite SE, et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. *Nucleic Acids Res*. 2023;52(D1):D138-D144. doi:10.1093/nar/gkad965
23. Lanz TA, Joshi JJ, Reinhart V, Johnson K, Grantham II LE, Volfson D. STEP Levels Are Unchanged in Pre-Frontal Cortex and Associative Striatum in Post-Mortem Human Brain Samples from Subjects with Schizophrenia, Bipolar Disorder and Major Depressive Disorder. *PLOS One*. 2015;10(3):e0121744. doi:10.1371/journal.pone.0121744
24. Lanz TA, Reinhart V, Sheehan MJ, et al. Postmortem transcriptional profiling reveals widespread increase in inflammation in schizophrenia: a comparison of prefrontal cortex, striatum, and hippocampus among matched tetrads of controls with subjects diagnosed with schizophrenia, bipolar or major depressive disorder. *Transl Psychiatry*. 2019;9(1):151. doi:10.1038/s41398-019-0492-8
25. Hagenauer MH, Schulmann A, Li JZ, et al. Inference of cell type content from human brain transcriptomic datasets illuminates the effects of age, manner of death, dissection, and psychiatric diagnosis. *PLOS One*. 2018;13(7):e0200003. doi:10.1371/journal.pone.0200003
26. Bolstad BM, Irizarry RA, Åstrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185-193. doi:10.1093/bioinformatics/19.2.185
27. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 2015;43(7):e47. doi:10.1093/nar/gkv007
28. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc: B (Methodol)*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x
29. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559. doi:10.1186/1471-2105-9-559
30. Zhang B, Horvath S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat Appl Genet Mol Biol*. 2005;4(1). doi:10.2202/1544-6115.1128
31. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425. doi:10.1016/j.cels.2015.12.004
32. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556
33. Gene Ontology Consortium, Aleksander SA, Balhoff J, et al. The Gene Ontology knowledgebase in 2023. *Genetics*. 2023;224(1):iyad031. doi:10.1093/genetics/iyad031
34. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28(1):27-30. doi:10.1093/nar/28.1.27

35. Schaefer CF, Anthony K, Krupa S, et al. PID: the Pathway Interaction Database. *Nucleic Acids Res*. 2009;37(Database issue):D674-D679. doi:10.1093/nar/gkn653
36. Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2018;46(Database issue):D649-D655. doi:10.1093/nar/gkx1132
37. Slenter DN, Kutmon M, Hanspers K, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res*. 2018;46(D1):D661-D667. doi:10.1093/nar/gkx1064
38. Nishimura D. BioCarta. *Biotech Softw Internet Rep*. 2001;2(3):117-120. doi:10.1089/152791601750294344
39. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740. doi:10.1093/bioinformatics/btr260
40. Li H, Zhang J, Shi Y, et al. Identification of matrix stiffness-related molecular subtypes in HCC via integrating multi-omics analysis and machine learning algorithms. *J Transl Med*. 2025;23(1):716. doi:10.1186/s12967-025-06733-7
41. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res*. 2012;40(17):e133. doi:10.1093/nar/gks461
42. Preacher KJ, Hayes AF. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behav Res Methods Instrum Comput*. 2004;36(4):717-731. doi:10.3758/BF03206553
43. Danaher P, Warren S, Dennis L, et al. Gene expression markers of Tumor Infiltrating Leukocytes. *J ImmunoTher Cancer*. 2017;5:18. doi:10.1186/s40425-017-0215-8
44. Refinetti R, Lissen GC, Halberg F. Procedures for numerical analysis of circadian rhythms. *Biol Rhythm Res*. 2007;38(4):275-325. doi:10.1080/09291010600903692
45. Fisher RA. *Statistical Methods for Research Workers*. Vol 24. 4th ed. Oliver and Boyd; 1932. doi:10.2134/agronj1932.00021962002400120011x
46. Berens P. CircStat: A MATLAB Toolbox for Circular Statistics. *J Stat Softw*. 2009;31:1-21. doi:10.18637/jss.v031.i10
47. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613(7944):508-518. doi:10.1038/s41586-022-05473-8
48. Fries GR, Saldana VA, Finnstein J, Rein T. Molecular pathways of major depressive disorder converge on the synapse. *Mol Psychiatry*. 2023;28(1):284-297. doi:10.1038/s41380-022-01806-1
49. Stein F, Kircher T. Transdiagnostic findings across major depressive disorder, bipolar disorder and schizophrenia: A qualitative review. *J Affect Disord*. 2025;387:119464. doi:10.1016/j.jad.2025.119464
50. Jiao W, Lin J, Deng Y, et al. The immunological perspective of major depressive disorder: unveiling the interactions between central and peripheral immune mechanisms. *J Neuroinflammation*. 2025;22(1):10. doi:10.1186/s12974-024-03312-3

51. Ye X, Ho YS, Chang RCC. Re-Examination of Inflammation in Major Depressive Disorder: Bridging Systemic and Neuroinflammatory Insights. *Biomolecules*. 2025;15(11):1556. doi:10.3390/biom15111556
52. Wenming Zhao, Daomin Zhu, Yongqi Zhang, et al. Neural Correlates of Peripheral Inflammation in Major Depressive Disorder and Their Transcriptomic Architecture, Neurochemical Basis, and Behavioral Relevance. *Hum Brain Mapp*. 2025;46(14):e70371. doi:10.1002/hbm.70371
53. González-Arias C, Perea G. When the Stars Misfire: Astrocytic Dysfunctions in Major Depressive Disorder. *Neurochem Res*. 2025;50(4):231. doi:10.1007/s11064-025-04483-y
54. Schurhoff N, Toborek M. Circadian rhythms in the blood–brain barrier: impact on neurological disorders and stress responses. *Mol Brain*. 2023;16(1):5. doi:10.1186/s13041-023-00997-0
55. Jiao H, Kalsbeek A, Yi CX. Microglia, circadian rhythm and lifestyle factors. *Neuropharmacology*. 2024;257:110029. doi:10.1016/j.neuropharm.2024.110029
56. Izzi B, Tirozzi A, Cerletti C, et al. Beyond Haemostasis and Thrombosis: Platelets in Depression and Its Co-Morbidities. *Int J Mol Sci*. 2020;21(22):8817. doi:10.3390/ijms21228817
57. Williams MS. Platelets and depression in cardiovascular disease: A brief review of the current literature. *World J Psychiatry*. 2012;2(6):114-123. doi:10.5498/wjp.v2.i6.114
58. Fagiani F, Di Marino D, Romagnoli A, et al. Molecular regulations of circadian rhythm and implications for physiology and diseases. *Signal Transduction Targeted Ther*. 2022;7(1):41. doi:10.1038/s41392-022-00899-y
59. Saeed S, Sang R, Zhixin L, et al. Circadian rhythms in major depressive disorder: mechanistic insights and therapeutic frontiers. *Ann Med*. 2026;58(1):2671594. doi:10.1080/07853890.2026.2671594

Supplementary Materials

The following supplementary files are submitted with this manuscript.

Supplementary Table S1: Complete gene sets used in this study, with their sources, member genes, and the number of detectable genes in each dataset (XLSX).

Supplementary File S2: The same gene sets as plain-text lists, one file per set, together with two comma-separated summary tables (ZIP, 34 files).

Supplementary File S3: Analysis code, organized by analytical layer (ZIP, 21 R scripts and 1 Python script).

Supplementary Table S4: Differential expression results for the discovery cohort GSE98793, including the covariate-adjusted model and the FDR-sensitivity subset (XLSX).

Supplementary Table S5: Co-expression network results for GSE98793: variance filtering, seed-gene provenance, module membership, module-trait associations, and module-level gene-set enrichment under both network strategies (XLSX).

Supplementary Table S6: Sample-level signature scores, group comparisons, inter-signature correlations, group comparisons adjusted for estimated cell composition and clinical covariates, and the exploratory path analysis (XLSX).

Supplementary Table S7: Estimated peripheral-blood cell composition, stiffness-cell correlations, cell-adjusted and fully adjusted group comparisons of signature scores with attenuation estimates, and robustness checks of the exploratory path analysis (XLSX).

Supplementary Table S8: Brain-tissue validation in GSE53987: regional signature scores, differential expression, and enrichment results (XLSX).

Supplementary Table S9: Circadian rhythmicity and phase analyses in the GSE45642 control brains, including preprocessing comparisons, genomic control, meta-Fisher rhythmicity tests, gene-set rhythmicity enrichment, and phase estimates (XLSX).

Supplementary Table S10: Diagnostic specificity analyses in GSE92538: sample characteristics, confounding checks, differential expression by diagnosis, signature scores and competitive gene-set tests for MDD, BD and SCZ, stiffness-gene statistics, clock-matrix coupling, cell-composition adjustment, and cross-dataset comparison (XLSX).

Supplementary Table S11: Set-level robustness of the matrix programme to adjustment for cellular composition in GSE98793: rotation tests with and without covariates in the design matrix, per-gene moderated statistics from the adjusted model, nominal-significance enrichment against background, and leave-one-gene-out sensitivity for every set (XLSX).

Supplementary Table S12: Direct diagnostic contrasts and cross-cohort pooling: MDD-versus-bipolar contrast estimates with standard errors at score and set level in GSE92538, cell-composition-adjusted specifications, per-region estimates from GSE53987, and inverse-variance pooling with heterogeneity statistics (XLSX).

Supplementary Table S13: Common-variant gene-set analysis of FinnGen release 12 depression summary statistics: per-gene association scores, variant-count stratification of background genes, observed and empirical quantiles for nine gene sets, and permutation results (XLSX).

Gene sets were intersected with the expression matrix used at each analytical layer, so the number of set genes entering a given test may be smaller than the number detectable on the platform; the counts reported in the tables of this paper are those produced by the analysis scripts, and platform-level counts are given in Table S1.

Author Contributions

Y-SL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. Z-LG: Data curation, Formal analysis, Investigation, Project administration, Writing – original draft. LZ: Data curation, Formal analysis, Investigation, Project administration, Software,

Visualization, Writing – original draft. C-LD: Methodology, Resources, Supervision, Validation, Writing – review and editing. W-HL: Funding acquisition, Supervision, Validation, Writing – review and editing. J-FY: Conceptualization, Methodology, Supervision, Validation, Writing – review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (grant No. 82574912); the Natural Science Foundation of Hunan Province (grant Nos. 2023JJ30451 and 2025JJ50719); and the Hunan Provincial Health Commission Research Project (grant No. 20256419).

Conflicts of Interest

The authors declare no conflict of interest.

Ethics Approval

Not applicable. This study used exclusively publicly available, de-identified transcriptomic data from the Gene Expression Omnibus; no new human subjects were involved, and ethical approval was therefore not required.

Informed Consent Statement

Not applicable.

Data Availability Statement

All expression data were obtained from the public database Gene Expression Omnibus (accession numbers GSE98793, GSE45642, GSE53987 and GSE92538). The FinnGen release 12 summary statistics analysed in this study are publicly available from the FinnGen results portal upon registration; no individual-level genotype data were accessed or transferred. All analysis code is provided as File S3.

Acknowledgments

We would like to thank Hunan University of Chinese Medicine for its support and for providing a good research environment. We are especially grateful to Prof. Junfeng Yan and Prof. Wanghua Liu for their guidance, patience, and many helpful discussions throughout this study. We also wish to thank all the students who participated in this project for their time, effort, and contributions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS13commonvariantFinnGen.xlsx
- TableS7cellcompositionsensitivity.xlsx
- TableS12directcontrastsandpooling.xlsx
- TableS6signaturescoresandmediation.xlsx
- FileS2genelistsplaintext.zip
- TableS1genesets.xlsx
- FileS3analysiscode.zip
- TableS4differentialexpressionGSE98793.xlsx
- TableS10diagnosticspecificityGSE92538.xlsx
- TableS9circadianrhythmicityGSE45642.xlsx
- TableS5coexpressionnetworkGSE98793.xlsx
- TableS8brainvalidationGSE53987.xlsx
- TableS11setlevelrobustness.xlsx