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.

Identification of shared blood transcriptomic features and candidate sedative targets in chronic insomnia with comorbid anxiety: an in-silico study

Weixian Xie

weixian1986@126.com

Affiliated Qingyuan Hospital,Guangzhou Medical University,Qingyuan People's Hospital

Wenyue Xiang

Affiliated Qingyuan Hospital,Guangzhou Medical University,Qingyuan People's Hospital

Zhiming Huang

Affiliated Qingyuan Hospital,Guangzhou Medical University,Qingyuan People's Hospital

Research Article

Keywords: chronic insomnia, anxiety, comorbidity, bioinformatics, weighted gene co-expression network analysis, machine learning, molecular docking, C3AR1

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10985453/v1

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

Additional Declarations: No competing interests reported.

Identification of shared blood transcriptomic features and candidate sedative targets in chronic insomnia with comorbid anxiety: an in-silico study

Weixian Xie,Wenyue Xiang,Zhiming Huang

Affiliated Qingyuan Hospital, Guangzhou Medical University ,Qingyuan People's Hospital, Qingyuan, Guangdong 511518, China

Corresponding author: Weixian Xie. Email: weixian1986@126.com

Keywords: chronic insomnia; anxiety; comorbidity; bioinformatics; weighted gene co-expression network analysis; machine learning; molecular docking; C3AR1

Abstract

Background: Chronic insomnia and anxiety co-occur frequently and reciprocally worsen each other's clinical trajectory, yet the molecular features shared by the two conditions remain poorly defined. We aimed to extract a peripheral blood transcriptomic signature linking the two disorders and to predict sedative/anxiolytic intervention targets.

Methods: PBMC profiles from GSE208668 (chronic insomnia, 17 vs 25) and whole-blood profiles from GSE61672 (generalized anxiety disorder, 157 vs 179) were analysed with limma. Because no anxiety gene survived Benjamini–Hochberg correction, a directional concordance screen combining nominal anxiety signals with insomnia significance produced 80 comorbidity candidate genes. Immune landscapes were inferred by CIBERSORT and ssGSEA. Hub genes were selected by LASSO, SVM-RFE and random forest; a logistic regression signature was trained in GSE61672. Sensitivity analyses probed screening thresholds and missing sex information. A curated 49-compound sedative/anxiolytic library was used for target overlap, Connectivity-Map-style connectivity scoring and AutoDock Vina docking of 14 hub proteins, with inert-ligand negative controls. Two independent sleep-loss cohorts (GSE39445, GSE56931) served as external validation.

Results: Insomnia PBMCs showed expansion of resting memory CD4 T cells and contraction of γδ T, Th1 and activated dendritic cell fractions. The pink WGCNA module was strongly enriched with the comorbidity genes ($P = 5.6 \times 10^{-9}$), and insomnia-cohort WGCNA independently recovered a trait-associated module ($r = 0.73$). Three machine-learning methods converged on 13 hub genes (training AUC 0.764); the cross-cohort AUC of 1.000 is uninterpretable because of circularity. The gene set was stable across threshold and sex-covariate sensitivity analyses. C3AR1 was the only hub gene overlapping a known drug target ($P = 0.0024$); docking predicted favourable C3AR1 binding for zolpidem (−8.53 kcal/mol), zaleplon (−8.21) and midazolam (−8.10), beyond the inert-ligand baseline. Directional consistency in acute sleep-loss cohorts was limited (48.7% and 46.8%), with GRINA and TRIM58 significant in both.

Conclusion: A candidate immune-oriented comorbidity signature links chronic insomnia and generalized anxiety, with C3AR1-centred complement signalling a candidate bridge to sedative pharmacology. The findings are exploratory, hypothesis-generating and entirely computational, and require validation in tissue-uniform, age-matched chronic-insomnia cohorts.

1. Background

Chronic insomnia affects roughly one in ten adults and is one of the most common reasons for primary care visits [1]. Generalized anxiety disorder (GAD) is similarly prevalent and carries comparable functional burden [2, 3]. When the two conditions coexist, outcomes are worse than for either alone: persistence is longer, suicidal ideation is more frequent, and treatment response is reduced [4, 5]. The longitudinal literature points to a bidirectional relationship, with insomnia often preceding the first anxiety episode and pre-existing anxiety lengthening the natural history of insomnia [4].

A mechanistic bridge between the two disorders is suggested by psychoneuroimmunology. Experimental sleep restriction increases circulating IL-6, TNF-α and C-reactive protein [6, 7, 29]. Anxiety disorders, in turn, are associated with elevated systemic inflammation in large cohort studies [8, 9]. Sleep and innate immunity are reciprocally coupled through cholinergic signalling, hormonal rhythms, and the complement cascade, the last of which has received comparatively little attention in this literature [7, 10, 28]. Complement factors such as C3 and C3aR have emerged as regulators of synaptic remodelling and microglial tone, but their peripheral blood transcriptomic footprints in insomnia–anxiety comorbidity have not been systematically queried [11].

Pharmacotherapy for the two conditions remains dominated by benzodiazepines, Z-drugs, melatonin receptor agonists and orexin antagonists, all of which target central nervous system effectors. Off-target immune effects are recognised but underexploited [12]. Public microarray cohorts provide an opportunity to look for shared peripheral signatures and to nominate repositionable targets.

Here we integrated two GEO microarray datasets that cover chronic insomnia (PBMC) and generalized anxiety disorder (whole blood), built a comorbidity gene list on the basis of cross-cohort directional concordance, characterised the cellular composition of the immune compartment in both datasets, and applied a three-method machine-learning funnel to derive a hub gene panel (Figure 1). We then asked whether curated sedative/anxiolytic compounds and immune-neuromodulators could be re-prioritised on the basis of that panel, including by molecular docking against AlphaFold and experimentally solved receptor structures. The pipeline ends with external validation in two independent sleep-loss cohorts (GSE39445 and GSE56931), whose limitations for modelling chronic insomnia are explicitly addressed. Throughout, the work is framed as an in-silico, hypothesis-generating analysis: no wet-laboratory experiment was performed, and all results require experimental confirmation before clinical interpretation.

2. Methods

2.1 Data sources and preprocessing

Four datasets were downloaded from the NCBI Gene Expression Omnibus (GEO). GSE208668 contains peripheral blood mononuclear cell (PBMC) profiles from 17 patients with chronic insomnia and 25 controls, profiled on the GPL10904 Illumina bead array. GSE61672 contains whole-blood profiles from 546 individuals with a generalized anxiety disorder diagnosis label, of which 336 had complete clinical annotations (157 GAD, 179 controls); these were profiled on the GPL10558 Illumina bead array. Two independent cohorts served as external validation: GSE39445 (sleep extension vs sleep restriction, 438 samples profiled on the GPL15331 Affymetrix Human Gene 2.0 ST array, 427 retained after paired-subject filtering) and GSE56931 (sleep deprivation vs baseline, 249 samples on the GPL10379 Affymetrix Human Genome U133 Plus 2.0 array, 209 retained). Demographic characteristics of the two discovery cohorts are summarised in Table S1. The insomnia cohort is substantially older (mean 68 years) than the anxiety cohort (mean 48 years); age was included as a covariate in the anxiety linear model but could not be modelled in the smaller insomnia cohort, and this tissue (PBMC vs whole blood) and demographic heterogeneity is

acknowledged as a limitation and addressed through cross-cohort directional concordance, which requires only the sign (not the magnitude) of effects to agree.

Series matrices were parsed manually; probe-level expression was background-corrected and quantile-normalised, $\log_2$-transformed, and mapped to gene symbols. When multiple probes targeted the same gene, the probe with the highest interquartile range was kept. Samples lacking complete clinical covariates in GSE61672 were excluded (210 of 546); missing sex information in the remaining 336 samples was encoded as a separate "unknown" factor level rather than imputed. The influence of this encoding was probed by a sensitivity analysis restricted to the 195 samples with known sex (see section 2.10).

2.2 Differential expression

Linear models were fitted with limma [12]. For GSE208668, the design included only the insomnia-vs-control group term. For GSE61672, the design was ~ age + sex + group, where "sex" was a factor with a missing-data level. Genes with Benjamini–Hochberg adjusted $P < 0.05$ and $|\log_2 \text{fold change}| \geq 0.5$ were called insomnia DEGs (7,670 in total; 4,506 up and 3,164 down). The anxiety cohort yielded no FDR-significant genes; nominal $P < 0.05$ was therefore used as the anxiety screening criterion for downstream comorbidity analysis. Global concordance of insomnia and anxiety $\log_2$ fold changes across 7,973 shared genes was assessed by Spearman correlation.

2.3 Comorbidity core genes

A gene was retained as a comorbidity candidate if it was (i) an insomnia DEG by the criteria above, (ii) nominally significant in the anxiety cohort ($P < 0.05$), and (iii) shared the same direction of regulation in both cohorts. Eighty genes (49 up, 31 down) met these criteria and formed the comorbidity core set (Table S4). Because criterion (ii) relies on nominal rather than FDR-controlled significance, the robustness of the core set to the choice of threshold was explicitly evaluated by repeating the entire screen at $P < 0.01$ and $P < 0.10$ and measuring retention of the 80 genes and of the 13 consensus hub genes (section 3.8, Table S2).

2.4 Functional enrichment

Over-representation analysis (ORA) against GO Biological Process, Cellular Component and Molecular Function, and against KEGG pathways, was performed with a hypergeometric test; P values were adjusted with Benjamini–Hochberg [13]. The background gene set was the intersection of genes measured in both GSE208668 and GSE61672.

2.5 Immune infiltration and core-gene correlations

CIBERSORT with the LM22 signature matrix [14] was applied to both cohorts after pre-filtering to remove low-variance probes. Samples whose deconvolution P value exceeded 0.05 were excluded from cell-fraction comparisons (34 of 42 in insomnia, 194 of 336 in anxiety). Single-sample gene-set enrichment analysis (ssGSEA) was performed across 28 immune signatures adapted from Charoentong and colleagues [15]; the enrichment score followed the original ssGSEA formulation [16]. Wilcoxon rank-sum tests compared each cell type between case and control within each cohort; Benjamini–Hochberg adjusted $P < 0.05$ defined significance. Spearman correlations between the 80 comorbidity core genes and each immune cell fraction or ssGSEA score were computed within each cohort.

2.6 Weighted gene co-expression network analysis

WGCNA [17] was run on the 336 samples of GSE61672. The input comprised the 5,000 genes with the highest median absolute deviation plus the 80 comorbidity core genes (5,029 in total). A signed hybrid

network was constructed with a soft-thresholding power of 14, which produced a scale-free topology fit index of 0.855. Modules were detected with the default dynamic tree-cut settings and visualised by hierarchical clustering of the topological overlap matrix. Module eigengenes were tested for Pearson correlation with the GAD diagnosis label. Enrichment of the 80 comorbidity genes within each module was assessed by hypergeometric test (Figure 10); module assignments of the hub genes are tabulated in Table S5.

To assess cross-cohort reproducibility of the co-expression architecture, the same WGCNA pipeline was applied independently to the 42 samples of GSE208668 (17 insomnia, 25 controls). Because of the small sample size, the soft-thresholding power was selected adaptively (power = 2) and the minimum module size reduced to 30; all other parameters were unchanged. Module eigengenes were correlated with the insomnia diagnosis label, and enrichment of the 80 comorbidity core genes within each insomnia-cohort module was assessed by the same hypergeometric test (section 3.4, Supplementary Figure S4).

2.7 Machine-learning screening and ROC evaluation

Three feature-selection algorithms were applied to the 80 comorbidity core genes in the GSE61672 training cohort. LASSO logistic regression [18, 19] used 10-fold cross-validation with the binomial deviance objective to select λ.min; selected genes were those with non-zero coefficients at that λ. SVM-RFE with a linear kernel [20] was implemented with the e1071 package and the squared weight magnitude as the elimination criterion. Random forest [21] used 500 trees with the default mtry; genes were ranked by mean decrease in the Gini index. Genes selected by all three methods formed the hub set (Table S5). A logistic regression classification signature was fitted on the z-scored training expression of the 13 hub genes, and the same coefficients were applied to the z-scored expression matrix of the 13 hub genes in GSE208668. We emphasise that this cross-cohort application is not an independent external validation of diagnostic performance: the comorbidity gene list was in part derived from insomnia DEGs of GSE208668, so the two cohorts are not fully independent with respect to feature selection. ROC curves, AUC values, 95% confidence intervals, and P values were obtained with the pROC package [22].

2.8 Drug target prediction and molecular docking

A curated library of 49 sedative, hypnotic, anxiolytic and immune-neuromodulatory compounds, with 44 distinct protein targets, was assembled from published pharmacology and from publicly available structure-activity summaries. Targets whose encoding genes were detected in the insomnia and anxiety expression matrices were retained. Direct overlap between the drug target set and the 80 comorbidity genes was tested by hypergeometric test. Connectivity scoring followed a CMap-style Kolmogorov–Smirnov enrichment statistic with 1,000 permutations against both disease signatures [23].

For molecular docking, three-dimensional receptor structures were retrieved from the Protein Data Bank when available, or from the AlphaFold Protein Structure Database otherwise [24, 25]. AlphaFold models are computational predictions rather than experimentally solved structures; local errors in side-chain geometry or binding-pocket conformation propagate into docking scores, so results based on predicted models are treated as less reliable than those based on experimental structures and are interpreted with corresponding caution. For 14 proteins (the 13 hub gene products plus C3AR1, the hub gene set member that overlapped a known drug target), structures were cleaned of non-standard residues and ligands, and converted to PDBQT format with the meeko toolkit (autodock-vina 1.2.5 pipeline) [26]. Ligand structures were downloaded as 3D conformers from PubChem [27]. Binding pockets were detected by a custom grid-based cavity analysis; for two receptor–ligand pairs in which no cavity was detected (PARP8 with several ligands and TNFRSF9 with

several ligands), a whole-protein bounding box was used and flagged in the output. Vina 1.2.5 was run with exhaustiveness 8 and the centre and size of each box reported in the supplementary table. To provide a non-specific binding baseline, three inert small-molecule reference ligands (glycine, D-glucose and glycerol) were docked against all 14 receptors under an identical protocol; drug-like affinities were interpreted against this negative-control distribution (section 3.6, Table S3). All docking scores are reported as predicted binding affinities and are not experimental measurements.

2.9 External validation

The 80 comorbidity core genes were mapped onto the validation cohorts. Limma with paired-subject designs was applied to GSE39445 (~ subject + sleepprotocol) and to GSE56931 (~ patient + group). Direction consistency was tested both by binomial test (against the null of 50%) and by Spearman correlation between validation and discovery $\log_2$ fold changes.

2.10 Sensitivity analyses

Two pre-specified sensitivity analyses addressed the two screening choices on which the comorbidity gene set depends. First, the nominal-significance threshold used for the anxiety cohort was varied over $P < 0.01$, $P < 0.05$ (baseline) and $P < 0.10$, and the entire comorbidity screen was repeated at each threshold; retention of the 80 core genes and of the 13 consensus hub genes, and the Jaccard index of each gene set against the baseline set, were recorded. Second, to evaluate the influence of the "unknown" sex level in GSE61672, the limma model ~ age + sex + group was refitted on the 195 samples with documented sex (95 GAD, 100 controls), and the $\log_2$ fold changes of the 80 core genes were compared with those of the full-cohort model (direction consistency, nominal significance and Spearman correlation). Results are reported in section 3.8 and Table S2.

2.11 Software and statistics

All analyses were performed in R 4.4.3, Python 3.13, and AutoDock Vina 1.2.5. Continuous variables were compared by Wilcoxon rank-sum tests; categorical associations by χ^2 tests; multiple-testing correction used Benjamini–Hochberg. Two-sided $P < 0.05$ defined significance unless otherwise stated.

Use of large language models. During the preparation of this work the authors used DeepSeek (a large language model developed by DeepSeek-AI, Hangzhou, China) in order to assist with drafting and language editing of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

3. Results

3.1 Differential expression and the comorbidity core

Limma yielded 7,670 DEGs in the insomnia cohort (4,506 up, 3,164 down; FDR < 0.05 and $|\log_2 FC| \geq 0.5$; Figure 2A). The anxiety cohort failed to produce any FDR-significant genes (Figure 2B), with 510 genes reaching nominal $P < 0.05$. Of these, 199 overlapped with the insomnia DEG set (Figure 2C). Filtering for concordant $\log_2 FC$ direction yielded 80 comorbidity core genes (49 up, 31 down), illustrated as a z-score heatmap in Figure 2D and tabulated in Table S4. The highest-ranked genes by summed $\log_2 FC$ across the two cohorts were IL32, BBX, UCP2, GRINA, HDGF and TRIM52. Global concordance of insomnia and anxiety effect sizes was modest but statistically supported (Spearman $\rho = 0.082$, $P = 2.1 \times 10^{-13}$). Demographic and tissue differences between the two discovery cohorts are summarised in Table S1 and considered in the Discussion.

3.2 Pathway enrichment of the comorbidity and disease-specific gene sets

In the insomnia cohort, 1,532 GO Biological Process terms and 86 KEGG pathways were enriched after FDR correction (BH < 0.05). The top KEGG pathways were spliceosome, T-cell receptor signalling, chemokine signalling, lysosome, natural-killer-cell-mediated cytotoxicity, B-cell receptor signalling and antigen processing and presentation (Figure 5A,B). Insomnia-down genes were enriched for extracellular structure organisation, wound healing and positive regulation of peptide hormone secretion (Figure 5C). Anxiety-cohort enrichment relied on nominal significance; among anxiety-down genes the dominant theme was cell-cycle G2/M transition and pyrimidine biosynthesis, while anxiety-up genes pointed to N-acylethanolamine metabolism (Figure 5D). No ORA term survived BH correction for the 80 comorbidity core genes; nominal enrichment nevertheless highlighted negative regulation of IL-2 production (fold enrichment 18.1) and positive regulation of chemokine and granulocyte chemotaxis (Table S4).

3.3 Immune landscape

In GSE208668, CIBERSORT identified four differentially abundant cell types at BH < 0.05: resting memory CD4 T cells were expanded in insomnia (mean 6.89% vs 0.65% in controls), together with memory B cells, whereas γδ T cells and naive CD4 T cells were contracted (Figure 3A). ssGSEA on the same cohort pointed to contraction of type-1 T helper cells, memory B cells, activated dendritic cells and activated CD4 T cells (Figure 3C). In GSE61672 neither CIBERSORT nor ssGSEA returned a BH-significant cell type (Figure 3B,D). The discrepancy between CIBERSORT memory B cells (up) and ssGSEA memory B cells (down) is reported here and not reconciled, because the two methods rely on partially overlapping but non-identical signature definitions.

Core-gene × immune-cell Spearman correlations were denser in the insomnia cohort (817 BH-significant pairs in CIBERSORT, 1,490 in ssGSEA) than in the anxiety cohort (414 and 807 respectively; Figure 4). IL32, the top comorbidity gene by summed effect size, correlated positively with CD8 T cells ($\rho = 0.74$), central memory CD4 T cells ($\rho = 0.50$) and activated CD8 T cells ($\rho = 0.44$) in GSE208668, and negatively with plasma cells ($\rho = -0.53$), activated CD4 T cells ($\rho = -0.47$) and activated CD4 T cells (ssGSEA, $\rho = -0.47$).

3.4 Co-expression module anchored on the comorbidity set

WGCNA on 336 GSE61672 samples yielded 15 modules. The pink module, containing 146 genes, was the most strikingly enriched for the 80 comorbidity core genes: 15 of 80 (18.8%) mapped to pink, against an expected 1.47 by chance, hypergeometric $P = 5.6 \times 10^{-9}$ ($BH = 5.0 \times 10^{-8}$; Figure 10). The module eigengenes of red, grey, pink and salmon showed weak negative or positive correlations with the GAD diagnosis label (red $\rho = -0.105$, $P = 0.054$; grey $\rho = -0.095$, $P = 0.080$; pink $\rho = 0.092$, $P = 0.092$; salmon $\rho = 0.091$, $P = 0.095$), consistent with the absence of FDR-significant DEGs in anxiety.

The same pipeline applied to the insomnia cohort (GSE208668, n = 42) produced three modules (turquoise, 2,824 genes; blue, 2,104; grey, 107). The turquoise module eigengene was strongly correlated with the insomnia diagnosis label ($r = 0.730$, $P = 4.0 \times 10^{-8}$), showing that a disease-associated co-expression signal can be detected even in this small cohort, though module assignments remain provisional given sample-size limitations (Supplementary Figure S4). Two further caveats apply. First, WGCNA is known to be unstable in cohorts of this size — module detection in networks built from fewer than approximately 50 samples is unreliable — so the insomnia-cohort modules should be considered provisional and require replication in larger cohorts. Second, enrichment of the 80 comorbidity core genes within insomnia-cohort modules was not significant (blue: 38 genes observed vs 33.4 expected, $P = 0.176$; turquoise: 42 vs 44.9, $P = 0.778$), in contrast

to the strong pink-module enrichment in the anxiety cohort. This asymmetry is considered further in the Discussion.

3.5 Hub genes and a logistic classification signature

LASSO, SVM-RFE and random forest selected 23, 28 and 26 genes respectively (Figure 8A–C). Thirteen genes were selected by all three methods and formed the hub set (Figure 8D; Table S5): EIF1AY, EIF4EBP3, KCTD18, LRRN3, MAP7, MRPS36, PARP8, RAC1, RNASEH2A, STK17B, TNFRSF9, TRIM32 and UCP2. A further 13 genes received votes from two of the three methods, with C3AR1 the only hub-tier gene that overlapped a curated drug target.

The 13-gene logistic signature achieved an AUC of 0.764 (95% CI 0.713–0.814, $P = 7.0 \times 10^{-17}$) in the GSE61672 training cohort. When the same coefficients were applied to GSE208668, the apparent AUC was 1.000; **this value must not be read as diagnostic evidence**. The comorbidity gene list was partly derived from GSE208668 DEGs, so the feature space is not independent of the evaluation cohort (circularity), and the validation set comprises only 42 PBMC samples, in which ROC estimates are unstable. The cross-cohort application is therefore reported for completeness only (Supplementary Figure S1) and was removed from the main figures accordingly. Single-gene ROC analysis showed modest training-set discriminative ability (AUC 0.56–0.62 across hub genes; Supplementary Figure S1B) and larger cross-cohort variance in the small validation cohort (KCTD18 1.00, PARP8 0.99, STK17B 0.92, LRRN3 0.84, UCP2 0.80, MRPS36 0.76), again reflecting the small sample size rather than generalisable performance.

3.6 Drug target overlap, connectivity scoring and molecular docking

Of 44 curated drug targets, only C3AR1 overlapped a hub gene (hypergeometric $P = 0.0024$ against the comorbidity core, BH = 0.0024 because of the single test), implicating the C3a receptor antagonist SB290157 as a mechanistically relevant probe. KS connectivity scoring against the insomnia signature ranked apremilast (PDE4 inhibitor) as the top hit (enrichment score 0.57, $P = 0.062$); connectivity against the anxiety signature produced no compound with sufficient permutation support (Figure 6A). No compound survived multiple-testing correction in the CMap-style connectivity analysis (all BH-adjusted $P > 0.05$); these results are exploratory only, and no drug-repositioning conclusion is drawn from them.

Across 140 protein–ligand pairs, the top five predicted binding affinities (Figure 9A,B) were C3AR1 with zolpidem (−8.53 kcal/mol), zaleplon (−8.21) and midazolam (−8.10), TRIM32 with zaleplon (−7.29), and RAC1 with midazolam (−6.31); the best ligand for each protein is summarised in Figure 9C. Fifty-seven pairs reached an affinity below −5 kcal/mol, a conventional threshold for in silico prioritisation. Six pairs produced positive-energy artefacts where no cavity was detected and a whole-protein bounding box was used (PARP8 with eszopiclone, midazolam and suvorexant; TNFRSF9 with eszopiclone and suvorexant; TRIM32 with suvorexant); these values are reported for completeness but were excluded from interpretation and from the group comparisons. The full matrix is in Table S7.

To calibrate these values, three inert reference ligands (glycine, D-glucose, glycerol) were docked against all 14 receptors under the identical protocol (42 pairs; Table S3). Negative-control affinities ranged from −2.37 to −5.31 kcal/mol (mean −3.44, median −3.37); the single strongest inert pair was TRIM32–glucose (−5.31). On C3AR1 itself, the best inert ligand (glucose) reached only −4.69 kcal/mol, so the predicted affinities of zolpidem, zaleplon and midazolam exceed the receptor-specific inert baseline by 3.4–3.8 kcal/mol. The top drug-like scores therefore lie clearly outside the non-specific binding range of these receptors, although the

margin for mid-tier pairs (−5 to −6 kcal/mol) overlaps the strongest negative controls and such pairs should be interpreted with particular caution (Supplementary Figure S2).

3.7 External validation in sleep-loss cohorts

Seventy-eight of the 80 comorbidity genes were detected in GSE39445 and 77 in GSE56931. Directional consistency against the discovery $\log_2$FC was 38/78 (48.7%) in GSE39445 and 36/77 (46.8%) in GSE56931, neither significant by binomial test (P = 0.91 and P = 0.65). Spearman correlations between discovery and validation $\log_2$FC were $\rho = -0.021$ (P = 0.85) and $\rho = -0.116$ (P = 0.32) respectively (Figure 7A,B). Thirty-eight genes reached $P < 0.05$ in at least one validation cohort. GRINA was the most consistently replicated gene, with $P = 3.3 \times 10^{-4}$ in GSE39445 and $P = 3.3 \times 10^{-4}$ in GSE56931, both with the same direction as in the discovery cohort. TRIM58 was also significant in both ($P = 1.3 \times 10^{-4}$ and P = 0.029). Twenty-four of 76 evaluable genes were direction-consistent in both cohorts (Table S6).

3.8 Sensitivity analyses

Nominal-P threshold

Repeating the comorbidity screen at anxiety-cohort thresholds of $P < 0.01$ and $P < 0.10$ bracketed the baseline $P < 0.05$ set (Table S2). At $P < 0.01$, 12 comorbidity genes were retained, of which 12/12 (100%) belonged to the baseline 80-gene set; 5 of the 13 consensus hub genes (EIF4EBP3, MRPS36, RNASEH2A, TRIM32, UCP2) survived this strictest threshold. At $P < 0.10$, the comorbidity set expanded to 163 genes while retaining all 80 baseline genes and all 13 hub genes (Jaccard index 0.49 against the expanded set). The core signal is therefore stable within the conventional screening range: tightening the threshold prunes the set without changing any retained gene's status, and loosening it leaves the baseline set fully nested within the larger set (Supplementary Figure S3).

Sex covariate

Restricting GSE61672 to the 195 samples with documented sex and refitting ~ age + sex + group preserved the direction of all 80 core genes (100% concordance), with 41/80 (51.2%) individually nominal-significant in the same direction and a near-perfect rank correlation of $\log_2$FC with the full-cohort model (Spearman ρ = 0.920; Table S2). The "unknown" sex level therefore does not materially drive the comorbidity signal.

4. Discussion

This study integrated two publicly available microarray cohorts to identify a candidate peripheral blood signature shared by chronic insomnia and generalized anxiety disorder. The 80 candidate comorbidity-associated genes converged with prior transcriptomic reports on inflammation- and lymphocyte-related pathways [6, 7, 9], with the additional novelty of a complement-receptor (C3AR1) connection that ties in with the underappreciated role of complement in neuropsychiatric phenotypes [11, 28]. The pink WGCNA module, enriched at $P = 5.6 \times 10^{-9}$ for the comorbidity set, suggests that the signature is internally coherent in a larger cohort even where single-gene DEG analysis fails to reach significance. Replication of the WGCNA pipeline in the insomnia cohort yielded a mixed picture that we report transparently: a disease-associated co-expression module was clearly recoverable (turquoise, r = 0.73), yet the 80 comorbidity genes were not significantly over-represented in any insomnia-cohort module. The comorbidity signature is thus anchored in anxiety-cohort network structure, while its insomnia-cohort network anchoring remains unproven — an asymmetry that likely reflects both the small insomnia sample (n = 42) and the fact that the insomnia DEG signal is much

broader (7,670 genes) than a single module can capture. That coherence does not constitute causal evidence, but it does justify the gene-prioritisation step that followed.

The immune profile in insomnia PBMCs points to a redistribution of the CD4 compartment rather than to overt lymphopenia or neutrophilia: resting memory CD4 T cells expanded roughly ten-fold while naïve CD4 T cells and γδ T cells contracted. The simultaneous ssGSEA decrease in activated CD4 T cells and type-1 T helper cells fits a model in which chronic insomnia drives T-cell homeostasis changes rather than classical acute inflammation, consistent with the observation that chronic insomnia is associated with subtle rather than dramatic shifts in blood counts [7, 8]. The opposite directions of memory B cells in CIBERSORT versus ssGSEA underline that immune signatures are method-dependent and should be triangulated with protein-level data before clinical use.

Among the 13 hub genes, several have clear mechanistic hooks. STK17B encodes DRAK2, a serine/threonine kinase that restrains T-cell activation and promotes apoptosis; its nomination in a sleep-and-anxiety context is consistent with the lymphocyte remodelling we observed. TNFRSF9 encodes the costimulatory receptor 4-1BB, a T-cell activation molecule that is targeted by agonist antibodies in tumour immunotherapy. RAC1 is a Rho GTPase central to actin dynamics and synaptic plasticity. EIF4EBP3 belongs to the eIF4E-binding protein family and represses cap-dependent translation under stress. UCP2 uncouples oxidative phosphorylation and is increasingly discussed in the metabolic side of stress-related disorders. The fact that none of these genes carries a very large single-cohort effect size is consistent with the modest cross-cohort directional concordance and argues against any single gene being a stand-alone biomarker.

The drug-target overlap highlighted C3AR1 as the only hub gene that is also a curated sedative, anxiolytic or immune-neuromodulator target. The C3aR antagonist SB290157 has anti-inflammatory activity in animal models, and its cross-cohort enrichment ($P = 0.0024$) is one of the more specific findings of this analysis [10, 11]. By contrast, the CMap-style connectivity analysis produced no compound surviving multiple-testing correction (Table S8); it is reported as exploratory and carries no repurposing claim. The docking experiment predicted favourable binding affinity of zolpidem, zaleplon and midazolam for C3AR1, and the inclusion of inert reference ligands showed that these scores exceed the receptor-specific non-specific baseline by 3.4–3.8 kcal/mol. This calibration strengthens the computational observation, but the predicted affinities are unlikely to reflect a clinically meaningful off-target effect of these compounds, since their canonical pharmacology is GABA-A-receptor mediated; rather, they suggest that the C3aR binding pocket tolerates small hydrophobic scaffolds and could be drugged with better-fitting chemotypes. The docking predictions for TRIM32 and RAC1 are similarly hypothesis-generating rather than clinical, and mid-tier scores (−5 to −6 kcal/mol) overlap the strongest negative controls and warrant scepticism.

The two external cohorts tested a different biological question than the discovery cohorts. GSE39445 and GSE56931 are sleep-loss paradigms, that is acute sleep restriction and one night of total sleep deprivation, and therefore do not model chronic insomnia. The fact that fewer than half of the comorbidity genes reproduce their direction in these acute settings is therefore expected and informative: chronic insomnia cannot be modelled by two to seven nights of restricted sleep at the peripheral transcriptomic level. The two genes that did replicate (GRINA, TRIM58) are reproducible in acute sleep-loss paradigms; it remains unknown whether they are altered in patients with chronic insomnia–anxiety comorbidity, and their candidacy as comorbidity genes must therefore await testing in chronic-disease cohorts. They nonetheless deserve individual follow-up, particularly GRINA given its glutamate-receptor-associated function. We also note that AUC = 1.000 in the

validation cohort does not imply a clinically useful test, because the comorbidity list is partly derived from that same cohort, the sample size is small, and the validation cohort is acute rather than chronic.

Before listing limitations, we summarise what the sensitivity analyses add. The comorbidity set behaved as a nested, monotone family across anxiety-cohort thresholds from $P < 0.01$ to $P < 0.10$, and the five hub genes that survived the strictest threshold (EIF4EBP3, MRPS36, RNASEH2A, TRIM32, UCP2) constitute a conservative inner core for future replication efforts. The sex-covariate analysis (100% direction consistency, $\rho = 0.920$) indicates that the substantial fraction of samples with undocumented sex does not drive the signal. These checks bound, but do not eliminate, the dependence of the gene list on nominal significance.

Several limitations qualify every finding in this work. (i) Tissue and demographic heterogeneity: the insomnia cohort comprises PBMCs from older adults (mean 68 years) and the anxiety cohort whole blood from middle-aged adults (mean 48 years); cell-composition and age effects therefore differ systematically between cohorts. Because the insomnia cohort contains only 42 samples, tissue-of-origin and age effects cannot be formally disentangled in that cohort: age could not be modelled as a covariate, and although the concordance screen uses only the sign of effects and the anxiety model adjusts for age, residual confounding by tissue-of-origin cannot be excluded. The same small sample size renders the insomnia-cohort WGCNA modules unstable (co-expression networks built from fewer than ~50 samples are known to be unreliable), so those modules should be treated as provisional and require replication in larger cohorts. (ii) Reliance on nominal significance: the anxiety cohort yielded no FDR-significant DEGs, so the comorbidity screen depends on nominal $P < 0.05$; the threshold and covariate sensitivity analyses mitigate but do not remove this weakness, and the 80-gene set should be regarded as a screening result. (iii) Circularity of the cross-cohort evaluation: the 13-gene signature was trained on genes selected partly through GSE208668 DEGs, so the apparent validation AUC of 1.000 is uninterpretable, and only the training-cohort AUC (0.764) carries evidential weight. (iv) Mismatch of validation paradigms: GSE39445 and GSE56931 model acute sleep restriction and one night of total sleep deprivation, not chronic insomnia, and the sub-50% directional concordance in these cohorts is consistent with the comorbidity signal being chronicity-specific — or with it being noise; the two designs cannot distinguish these explanations. (v) Docking is static and uncalibrated against experiment: receptor flexibility, solvation and membrane context are ignored, two pairs required a whole-protein fallback box, and although inert reference ligands provide a non-specific baseline, binding-affinity predictions alone cannot establish target engagement. (vi) Method-dependent immune estimates: CIBERSORT and ssGSEA returned opposite directions for memory B cells in the insomnia cohort; we report rather than resolve this discrepancy, which reflects that deconvolution signatures are not interchangeable and require protein-level triangulation. (vii) Absence of experimental validation: no wet-laboratory or clinical experiment supports any gene, module or interaction reported here; the entire analysis is in-silico and hypothesis-generating.

Future work should validate the 13 hub genes — and in the first instance the conservative five-gene core (EIF4EBP3, MRPS36, RNASEH2A, TRIM32, UCP2) that survives the strictest screening threshold — in chronic insomnia–anxiety comorbidity cohorts specifically designed to remove the confounders acknowledged above: a single uniform tissue type (PBMC or whole blood throughout) and age-matched comorbid patients and controls, with RNA-seq and protein-level readouts. Complementary directions include characterising the C3aR axis in insomnia animal models and testing whether GRINA and TRIM58 emerge in prospective clinical cohorts. The CMap-style connectivity scores would also benefit from a larger reference signature library than the curated 49 compounds used here.

5. Conclusions

A peripheral blood transcriptomic analysis identifies a candidate immune-anchored signature shared by chronic insomnia and generalized anxiety disorder, with thirteen hub genes and a coherent pink WGCNA module in the anxiety cohort; a five-gene inner core survives the strictest screening threshold. C3AR1 links the signature to a known pharmacological probe and to docking-predicted favourable binding affinity for sedative scaffolds that exceeds a receptor-specific inert-ligand baseline. Acute sleep-loss validation reproduces only a minority of the genes, the cross-cohort ROC evaluation is circular, and the work is in-silico throughout; clinical translation will require chronic-insomnia-specific cohorts and experimental confirmation.

List of abbreviations

AUC: area under the receiver operating characteristic curve; BH: Benjamini–Hochberg; C3AR1: complement C3a receptor 1; CIBERSORT: cell-type identification by estimating relative subsets of RNA transcripts; CMap: Connectivity Map; DEG: differentially expressed gene; FDR: false discovery rate; GAD: generalized anxiety disorder; GEO: Gene Expression Omnibus; GO: Gene Ontology; IL: interleukin; KEGG: Kyoto Encyclopedia of Genes and Genomes; KS: Kolmogorov–Smirnov; LASSO: least absolute shrinkage and selection operator; ORA: over-representation analysis; PBMC: peripheral blood mononuclear cell; ROC: receiver operating characteristic; ssGSEA: single-sample gene set enrichment analysis; SVM-RFE: support vector machine recursive feature elimination; TNF: tumour necrosis factor; WGCNA: weighted gene co-expression network analysis

Declarations

Ethics approval and consent to participate

Not applicable. This study used only publicly available, de-identified data.

Consent for publication

Not applicable.

Availability of data and materials

All microarray datasets analysed during the current study are available in the NCBI Gene Expression Omnibus repository, https://www.ncbi.nlm.nih.gov/geo/, under accessions GSE208668 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE208668), GSE61672 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE61672), GSE39445 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39445) and GSE56931 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56931). The curated drug-target library, the docking input files, the supplementary tables (S1–S8) and the supplementary figures (S1–S4) are included in this published article and its supplementary information files. Analysis scripts are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work received no external funding.

Authors' contributions

W.X. conceived and designed the study, performed the bioinformatics analyses, and drafted and revised the manuscript. W.Y.X. and Z.H. contributed to the interpretation of the results, provided clinical input, and critically revised the manuscript. All authors read and approved the final version.

Acknowledgements

We thank the investigators who deposited the GSE208668, GSE61672, GSE39445 and GSE56931 datasets in GEO, and the maintainers of the CIBERSORT, ssGSEA, AutoDock Vina, AlphaFold DB and PubChem resources used in this work.

Figure legends

Figure 1. Study design and analysis workflow. The pipeline runs from GEO data acquisition and preprocessing through differential expression analysis, comorbidity gene identification, immune infiltration, functional enrichment, gene–drug network construction, and final integration with WGCNA, machine learning and molecular docking.

Figure 2. Differential expression analysis and the comorbidity core. (A) Volcano plot of insomnia DEGs in GSE208668. (B) Volcano plot of anxiety DEGs in GSE61672; no gene reaches FDR significance. (C) Venn diagram of insomnia DEGs (7,471) and nominal anxiety DEGs (510), with a 199-gene overlap. (D) z-score heatmap of the 80 direction-consistent comorbidity core genes across both cohorts.

Figure 3. Immune cell comparison between disease and control. (A,B) CIBERSORT LM22 cell fractions in GSE208668 and GSE61672. (C,D) ssGSEA scores across 28 immune signatures in the two cohorts. Boxes in red denote the case group.

Figure 4. Gene–immune correlation heatmaps. (A,B) ssGSEA enrichment scores across 28 signatures in GSE208668 and GSE61672. (C,D) Spearman correlations between the 80 comorbidity core genes (rows) and CIBERSORT cell fractions (columns), coloured by direction of regulation (blue: down in both; red: up in both).

Figure 5. Functional enrichment dot plots. (A,B) Top GO Biological Process and KEGG terms enriched in insomnia-upregulated genes. (C) Top GO Biological Process terms enriched in insomnia-downregulated genes. (D) Top KEGG terms enriched in anxiety-upregulated genes (nominal $P < 0.05$).

Figure 6. Drug-target prediction and gene-drug network. (A) Connectivity scores of the curated sedative/anxiolytic library against the insomnia signature. (B) Gene-drug network centred on the 13 comorbidity core genes (orange) and the 10 prioritised drugs (coloured squares).

Figure 7. External validation in sleep-loss cohorts. (A,B) Scatter plots of discovery (GSE208668) $\log_2$FC against validation $\log_2$FC in GSE39445 and GSE56931; consistent genes are in red. (C) Heatmap of core gene $\log_2$FC across the discovery cohort and the two validation cohorts for genes with $P < 0.05$ in at least one validation cohort.

Figure 8. Machine-learning screening of hub genes. (A) LASSO cross-validation curve and selected coefficients. (B) SVM-RFE accuracy curve. (C) Random forest top-30 gene importance. (D) Three-set Venn diagram of the three selection methods (LASSO, SVM-RFE, random forest); the 13 genes in the three-way intersection constitute the hub set. All ROC analyses (13-gene classification signature and single-gene ROC curves) are presented in Supplementary Figure S1.

Figure 9. Molecular docking of hub proteins against sedative ligands. (A) Binding-affinity heatmap across 14 hub proteins and 10 sedative compounds; values prefixed with a plus sign (+) correspond to positive-

energy artefacts from whole-protein bounding boxes and should not be interpreted biologically. (B) Top 15 protein–ligand pairs ranked by affinity. (C) Best ligand per protein.

Figure 10. WGCNA and module-level comorbidity enrichment. (A1,A2) Scale-free topology fit index and mean connectivity across soft-thresholding powers. (B) Hierarchical clustering dendrogram of the 5,029 input genes with module colour bars. (C) Module-trait correlation heatmap for the GAD diagnosis label. (D) Module membership vs gene significance scatter for the red module, with hub genes in red. (E) Comorbidity gene enrichment per module; pink is the most significant.

Supplementary tables

Table S1. Demographic and clinical characteristics of the two discovery cohorts (age, sex distribution, sample sizes) (file: results/sensitivity/cohort_characteristics.csv).

Table S2. Sensitivity analyses of the comorbidity screen. (A) Nominal-P threshold sensitivity ($P < 0.01/0.05/0.10$): gene-set sizes, retention of the 80 core genes and 13 hub genes, Jaccard indices. (B) Sex-covariate sensitivity: limma refitted on the 195 known-sex samples, per-gene direction consistency and $\log_2$FC correlation with the full-cohort model (files: results/sensitivity/threshold_sensitivity.csv; results/sensitivity/sex_sensitivity_coregenes.csv; results/sensitivity/sex_sensitivity_summary.csv).

Table S3. Negative-control molecular docking: predicted affinities of glycine, D-glucose and glycerol against all 14 receptor proteins, with summary statistics contrasting the inert-ligand distribution with the sedative-library distribution (files: results/docking/negative_control_raw.csv; results/docking/negative_control_summary.csv).

Table S4. The 80 comorbidity core genes with insomnia and anxiety $\log_2$FC and P values (file: results/DEG/comorbidity_core_genes.csv).

Table S5. The 13 hub genes with three-method votes, training and validation AUCs, and WGCNA module assignments (file: results/ML/hub_genes.csv; results/WGCNA/hubgene_module_assignment.csv).

Table S6. External validation results in GSE39445 and GSE56931, including direction consistency, P values and binomial statistics for each of the 80 core genes (file: results/validation/validation_summary.csv).

Table S7. Full molecular docking record for the 140 receptor–ligand pairs: predicted binding affinities, grid-box centre coordinates and box dimensions, number of output modes, and docking status ("ok" = pocket-based box; "weak" = whole-protein fallback box used, positive-energy artefact) (file: results/docking/docking_raw.csv).

Table S8. KS connectivity scores of the 49-compound library against the insomnia and anxiety signatures (file: results/drug/connectivity_scores_with_direction.csv).

Supplementary figures

Figure S1. ROC analyses, moved from the main figures following peer review because the cross-cohort evaluation is circular and not interpretable as validation. (A) ROC of the 13-gene logistic classification signature in the GSE61672 training cohort (AUC 0.764) and its application to GSE208668 (apparent AUC 1.000, uninterpretable). (B) Single-gene ROC curves for the 13 hub genes in both cohorts (file: figures/supplementary/S1_validation_ROC.png).

Figure S2. Negative-control docking summary: distribution of predicted affinities for inert reference ligands (glycine, D-glucose, glycerol) versus sedative-library ligands across the 14 receptor proteins (file: figures/supplementary/S2_docking_negative_control.png).

Figure S3. Sensitivity analyses. (A) Comorbidity gene-set size, core-gene retention and hub-gene retention across nominal-P thresholds (0.01/0.05/0.10). (B) Per-gene comparison of full-cohort and known-sex-subset $\log_2$FC for the 80 core genes (file: figures/supplementary/S3_sensitivity.png).

Figure S4. WGCNA replication in the insomnia cohort (GSE208668). Soft-threshold selection, module dendrogram, module–trait correlations (turquoise $r = 0.73$, $P = 4.0 \times 10^{-8}$) and comorbidity-gene enrichment per module (file: figures/supplementary/S4_WGCNA_insomnia.png).

References

1. Morin CM, Benca R. Chronic insomnia. Lancet. 2012;379(9821):1129-1141.
2. Buysse DJ. Insomnia. JAMA. 2013;309(7):706-716.
3. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry. 2005;62(6):593-602.
4. Taylor DJ, Lichstein KL, Durrence HH, Reidel BW, Bush AJ. Epidemiology of insomnia, depression, and anxiety. Sleep. 2005;28(11):1457-1464.
5. Wittchen HU. Generalized anxiety disorder: prevalence, burden, and cost to society. Depress Anxiety. 2002;16(4):162-171.
6. Irwin MR, Olmstead R, Carroll JE. Sleep disturbance, sleep duration, and inflammation: a systematic review and meta-analysis of cohort studies and experimental sleep deprivation. Biol Psychiatry. 2016;80(1):40-52.
7. Besedovsky L, Lange T, Haack M. The sleep-immune crosstalk in health and disease. Physiol Rev. 2019;99(3):1325-1380.
8. Vogelzangs N, Beekman ATF, de Jonge P, Penninx BWJH. Anxiety disorders and inflammation in a large adult cohort. Transl Psychiatry. 2013;3:e256.
9. Irwin MR, Opp MR. Sleep health: reciprocal regulation of sleep and innate immunity. Neuropsychopharmacology. 2017;42(1):129-155.
10. Stephan AH, Barres BA, Stevens B. The complement system: an unexpected role in synaptic pruning during development and disease. Annu Rev Neurosci. 2012;35:369-389.
11. Coulthard LG, Woodruff TM. Is the complement activation product C3a a proinflammatory molecule? Re-evaluating the evidence and the myth. J Immunol. 2015;194(8):3542-3548.
12. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
13. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57(1):289-300.
14. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12(5):453-457.
15. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep. 2017;18(1):248-262.
16. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven tumors require TBK1. Nature. 2009;462(7269):108-112.

17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

18. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010;33(1):1-22.

19. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B. 1996;58(1):267-288.

20. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Mach Learn. 2002;46(1-3):389-422.

21. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32.

22. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.

23. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006;313(5795):1929-1935.

24. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455-461.

25. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589.

26. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database in 2024: providing high-quality experimentally validated structures. Nucleic Acids Res. 2024;52(D1):D368-D375.

27. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res. 2019;47(D1):D1102-D1109.

28. Feuth T. Interactions between sleep, inflammation, immunity and infections: a narrative review. Immun Inflamm Dis. 2024;12(10):e70046.

29. Irwin MR. Sleep and inflammation: partners in sickness and in health. Nat Rev Immunol. 2019;19(11):702-715.

Figures

Figure 1. Study design and analysis workflow.

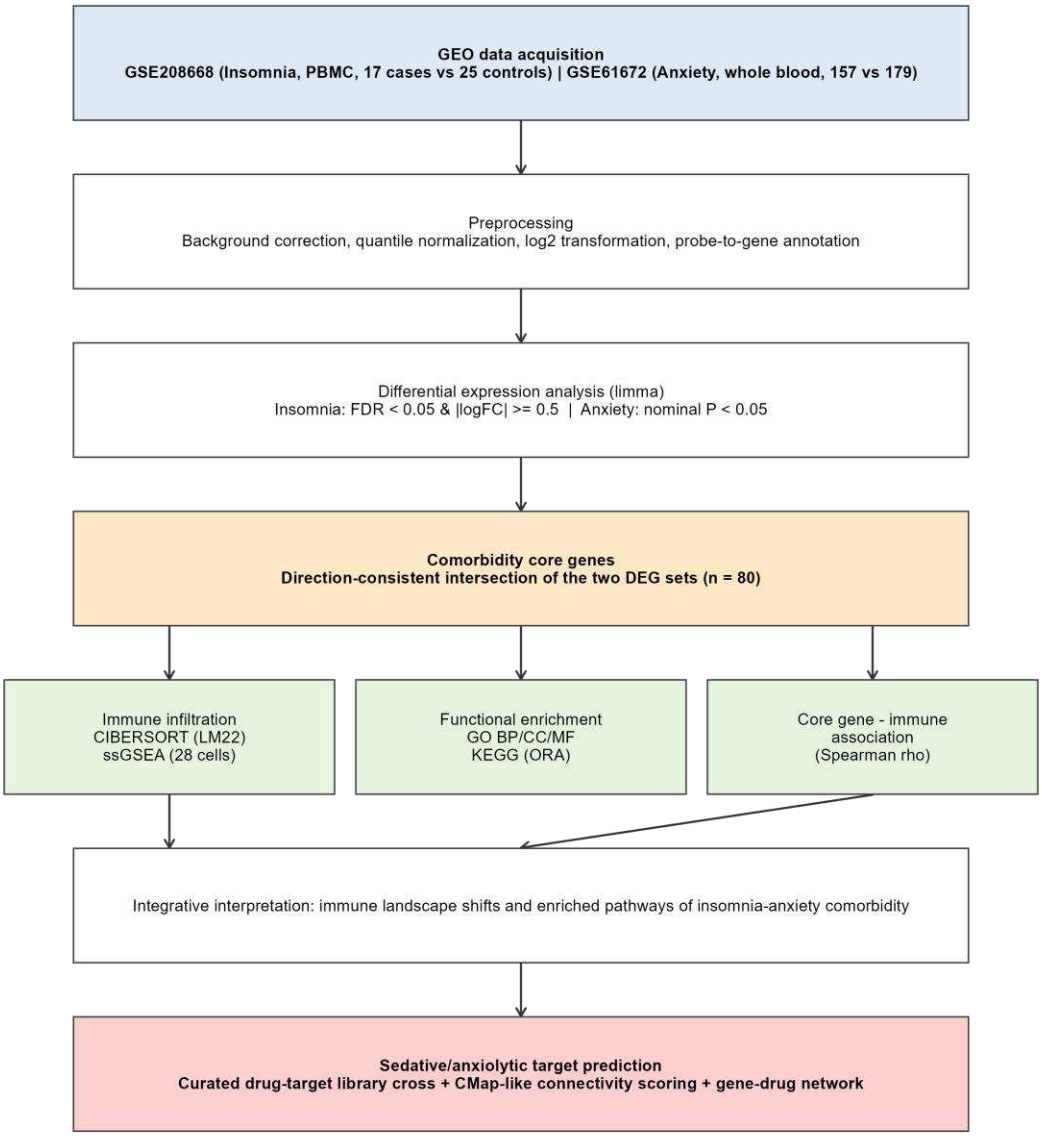
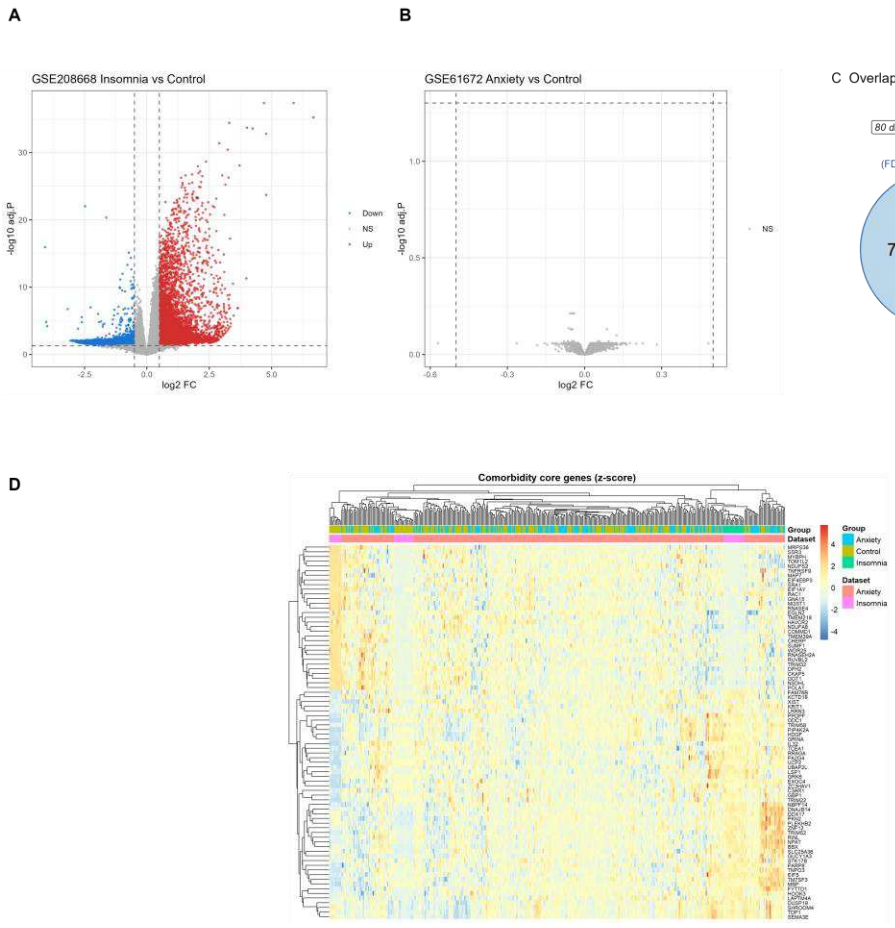


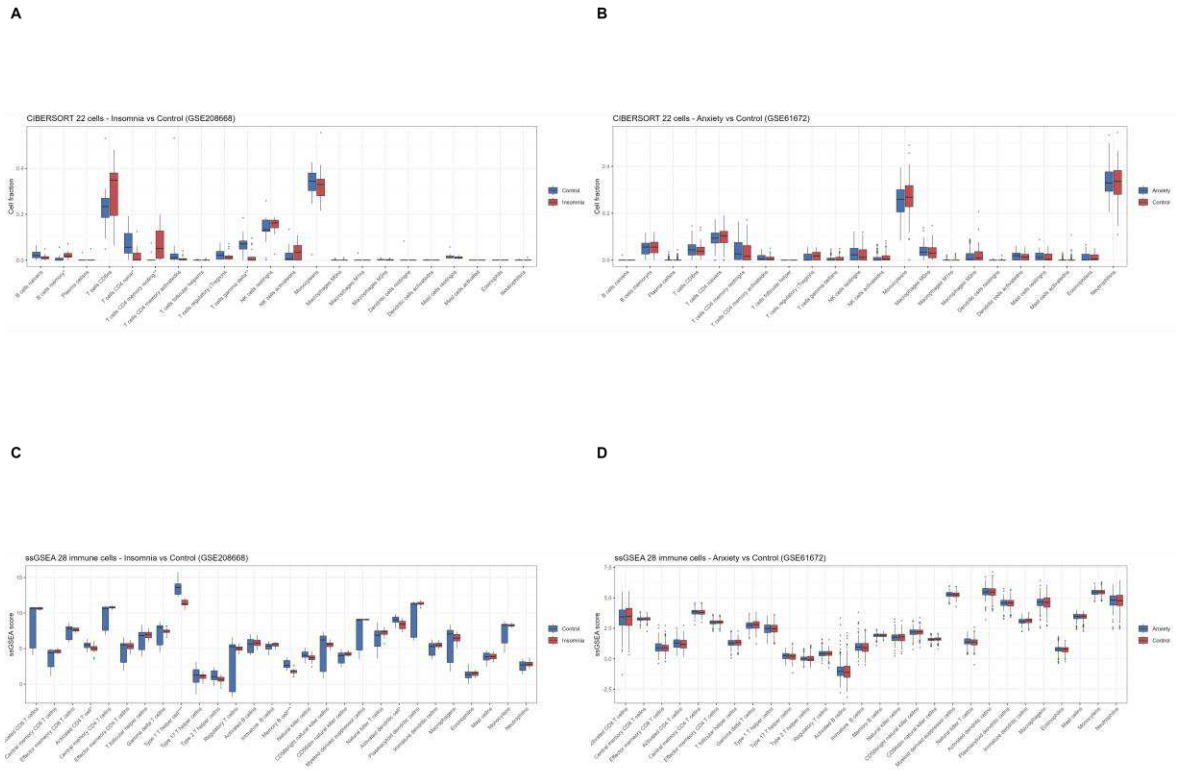
Figure 2. Differential expression analysis and the comorbidity core.



588

589

590 Figure 3. Immune cell comparison between disease and control.



591

592

593 Figure 4. Gene-immune correlation heatmaps.

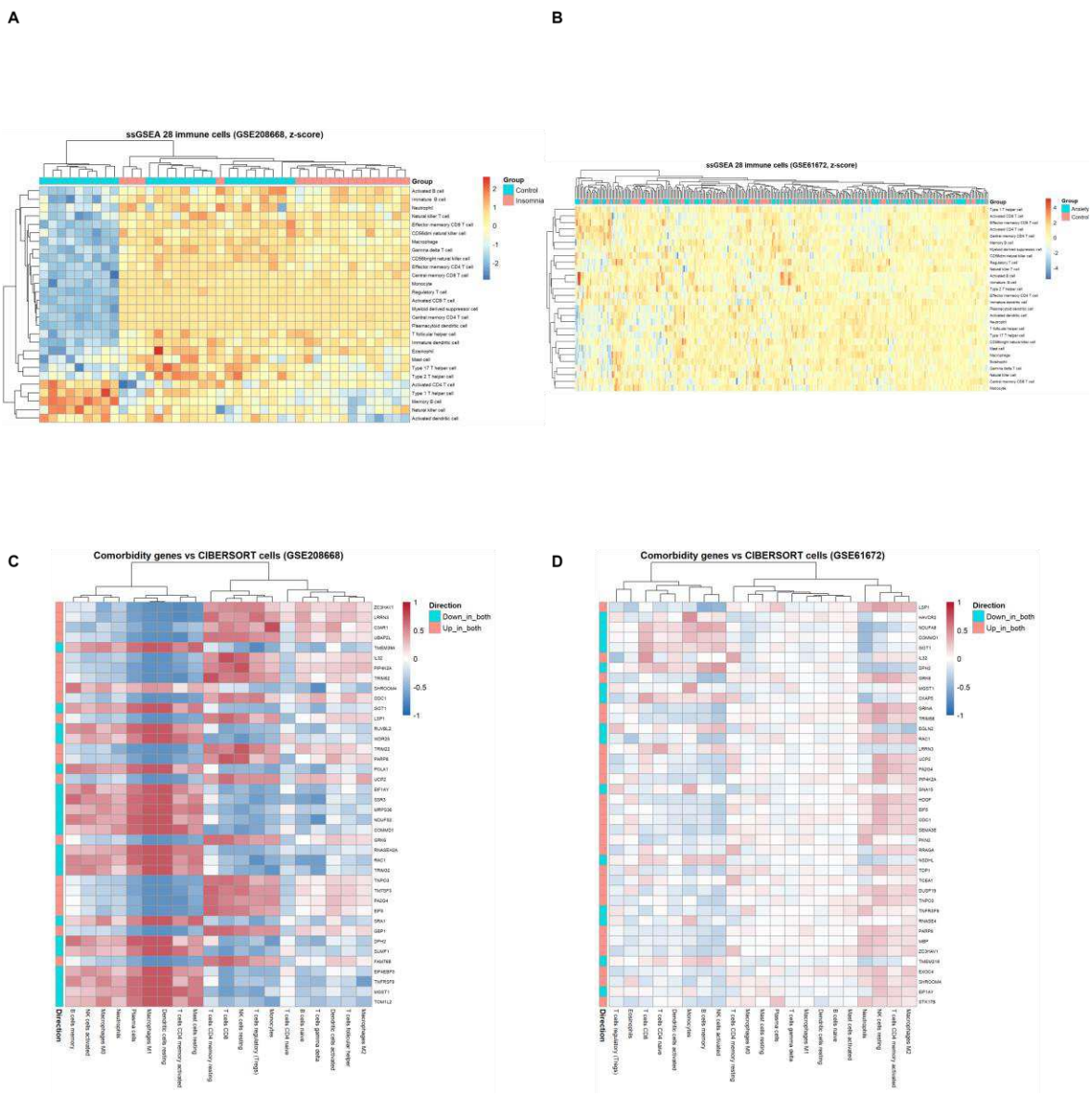


Figure 5. Functional enrichment dot plots.

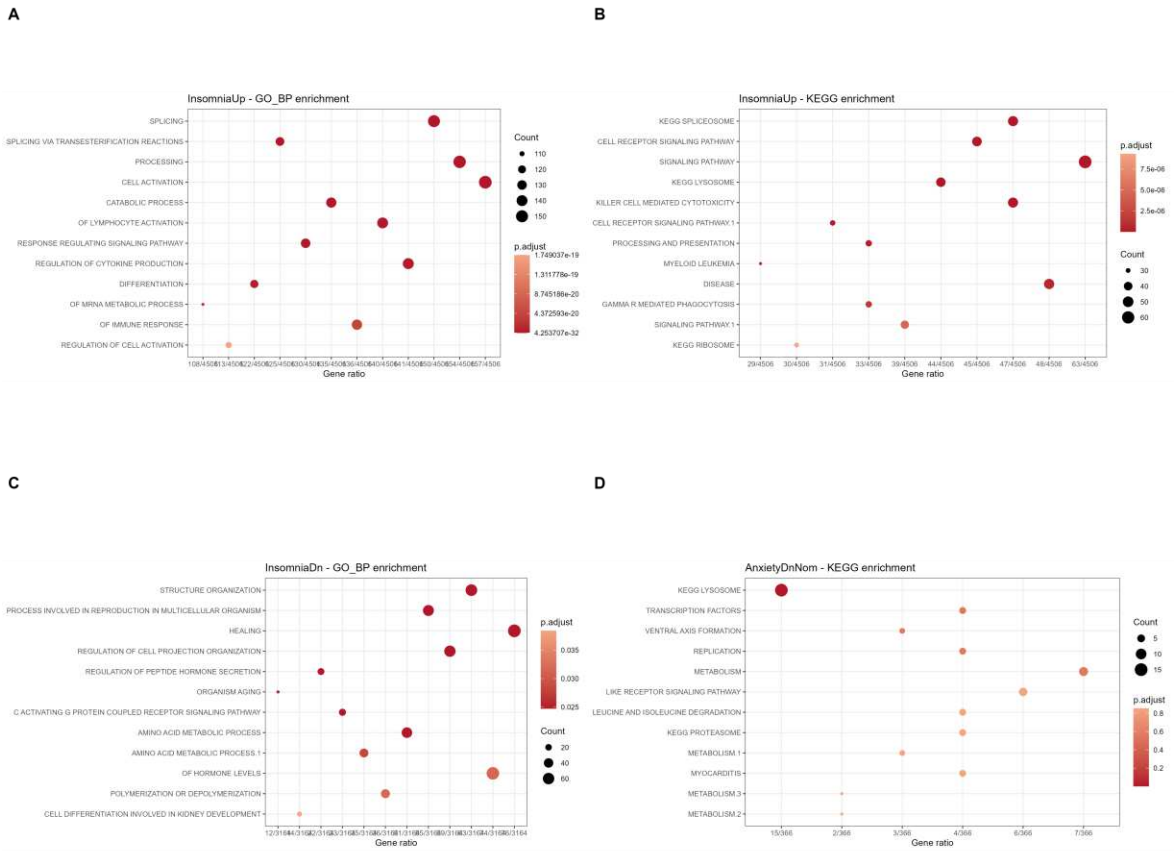


Figure 6. Drug-target prediction and gene-drug network.

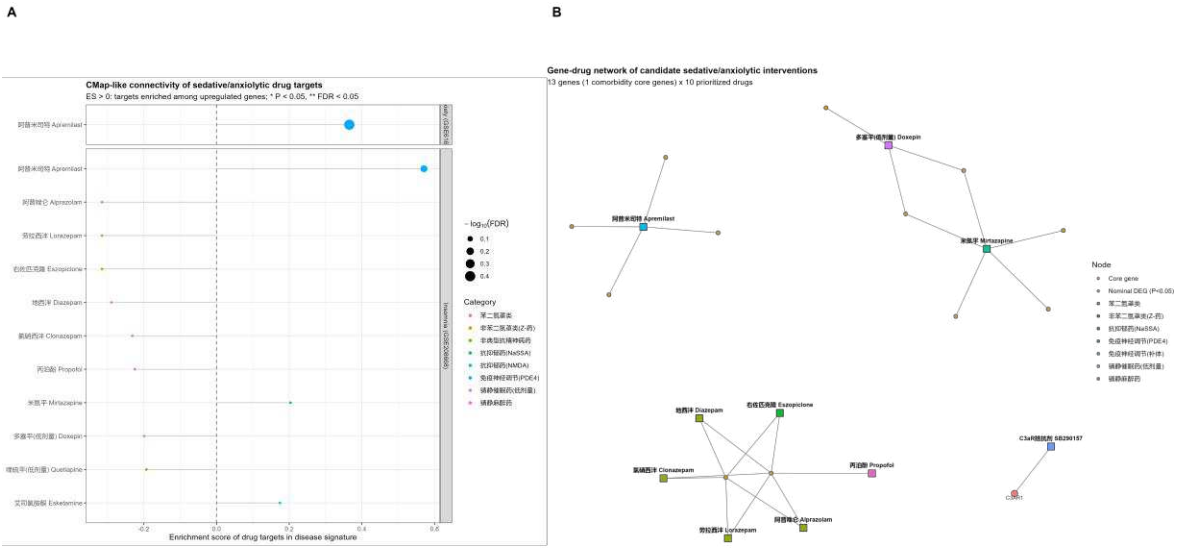


Figure 7. External validation in sleep-loss cohorts.

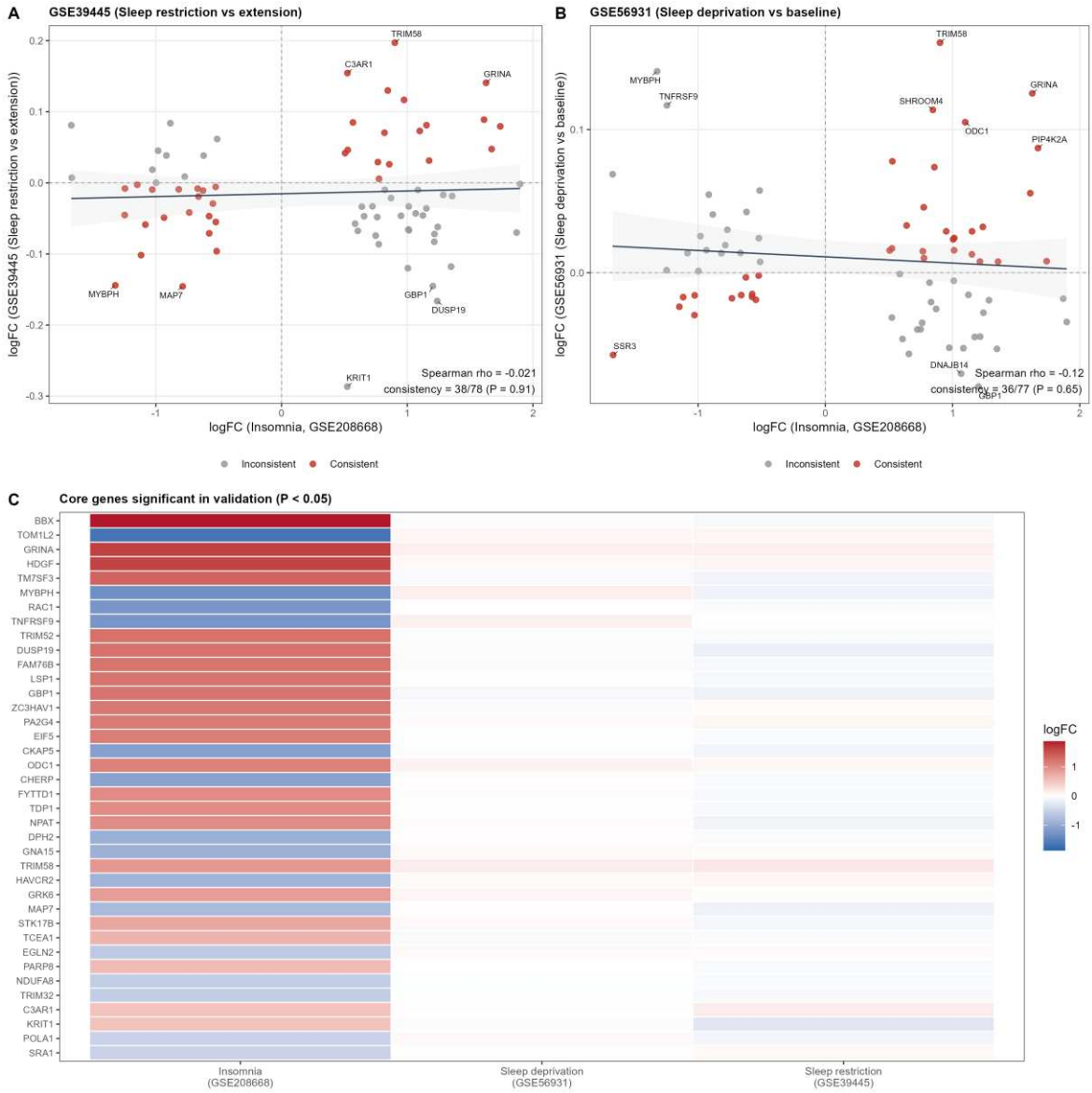
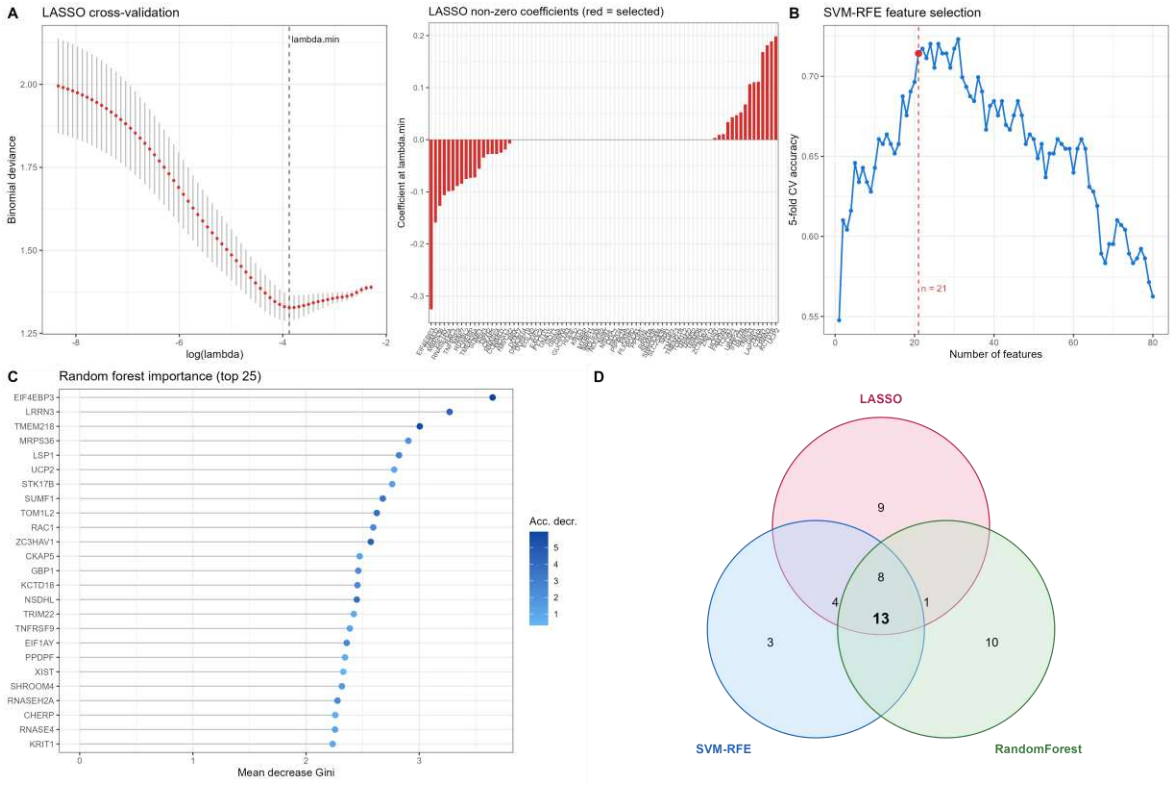


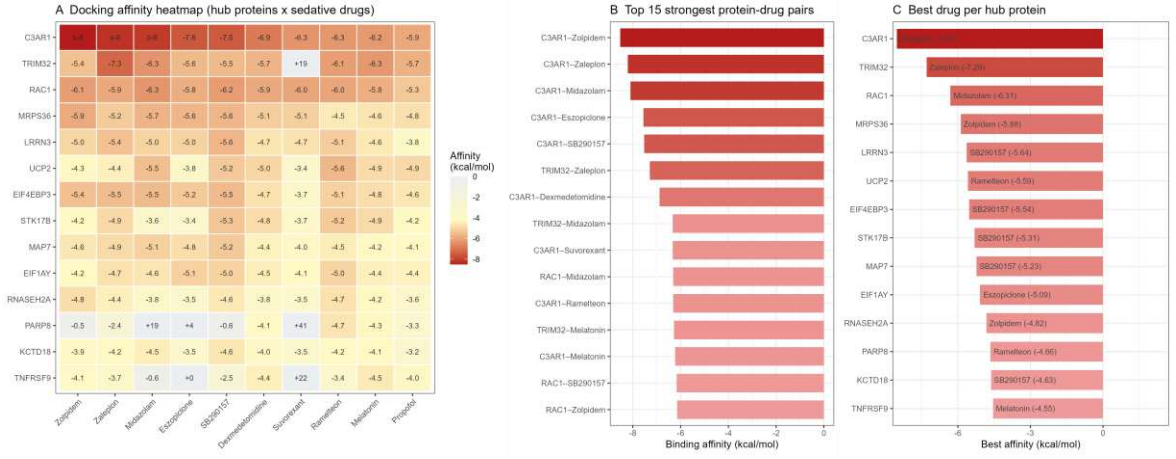
Figure 8. Machine-learning screening of hub genes.



606

607

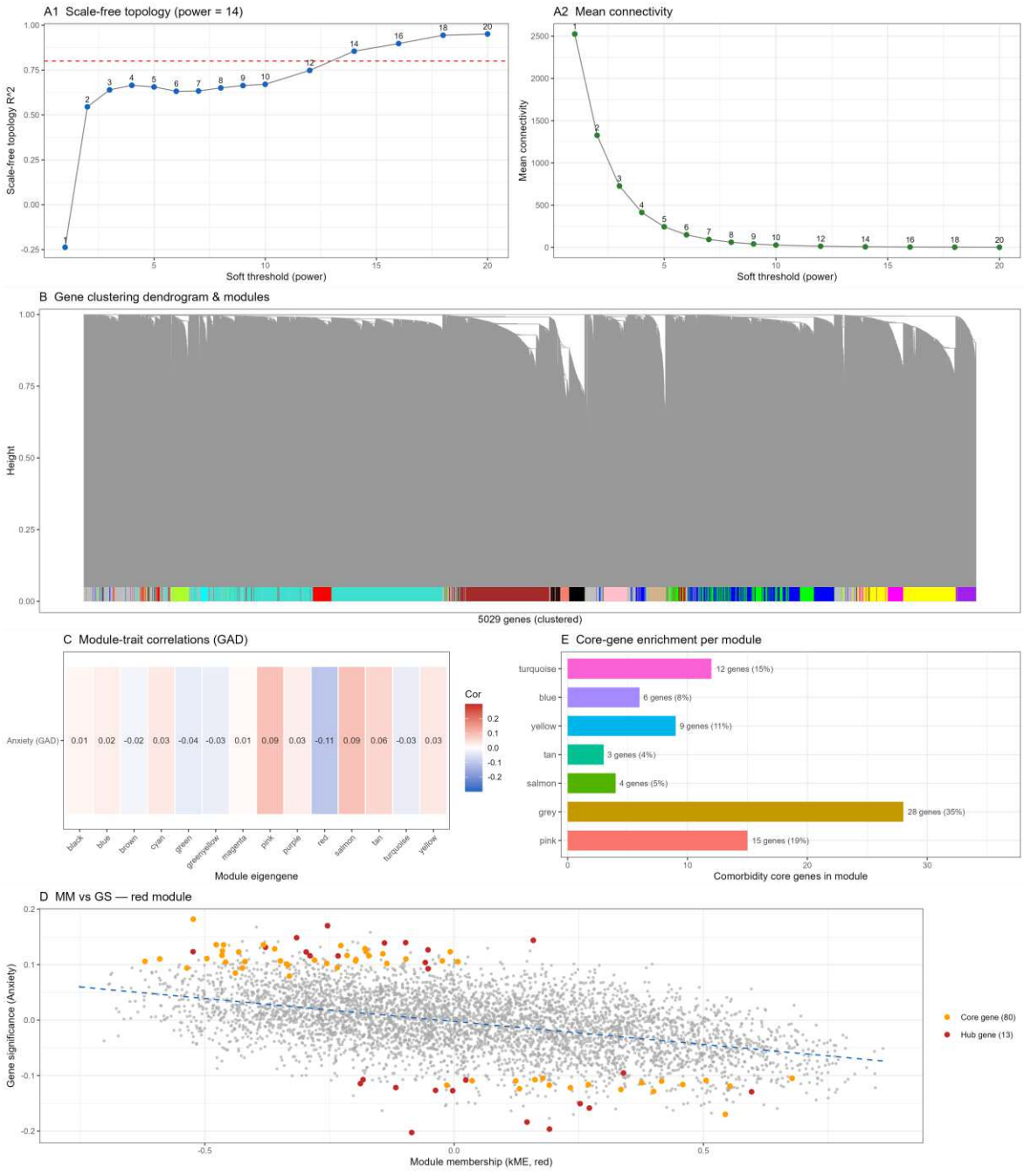
608 Figure 9. Molecular docking of hub proteins against sedative ligands.



609

610

611 Figure 10. WGCNA and module-level comorbidity enrichment.


A1 Scale-free topology (power = 14)
Scale-free topology R^2
Soft threshold (power)
A2 Mean connectivity
Mean connectivity
Soft threshold (power)
B Gene clustering dendrogram & modules
Height
5029 genes (clustered)
C Module-trait correlations (GAD)
Anxiety (GAD)
0.01 0.02 -0.02 0.03 -0.04 -0.03 0.01 0.09 0.03 -0.11 0.09 0.06 -0.03 0.03
black blue brown cyan green greenyellow magenta pink purple red salmon tan turquoise yellow
Module eigengene
Cor
0.2 0.1 0.0 -0.1 -0.2
E Core-gene enrichment per module
turquoise 12 genes (15%)
blue 6 genes (8%)
yellow 9 genes (11%)
tan 3 genes (4%)
salmon 4 genes (5%)
grey 28 genes (35%)
pink 15 genes (19%)
Comorbidity core genes in module
D MM vs GS — red module
Gene significance (Anxiety)
Module membership (kME, red)
Core gene (80)
Hub gene (13)

612

613

Supplementary Files

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

- TableS1cohortcharacteristics.csv
- TableS2Athresholdsensitivity.csv
- TableS2Bsexsensitivitysummary.csv
- TableS3Anegativecontrolraw.csv
- TableS3Bnegativecontrolsummary.csv
- TableS5Bhubgenemoduleassignment.csv
- TableS4comorbiditycoregenes.csv
- TableS2Csexsensitivitycoregenes.csv
- TableS5Ahubgenes.csv
- TableS6validationsummary.csv
- TableS7dockingraw.csv
- SupplementaryMaterials.docx
- TableS8connectivityscoreswithdirection.csv