

# Investigating biomarkers and regulatory mechanisms associated with depression and anxiety in obesity based on transcriptome sequencing and experimental validation

**Xiaochen Ren**

First Teaching Hospital of Tianjin University of Traditional Chinese Medicine

**Ziqi Zhao**

First Teaching Hospital of Tianjin University of Traditional Chinese Medicine

**Ning Xu**

First Teaching Hospital of Tianjin University of Traditional Chinese Medicine

**Qianjing Yang**

First Teaching Hospital of Tianjin University of Traditional Chinese Medicine

**Yuyang Lan**

First Teaching Hospital of Tianjin University of Traditional Chinese Medicine

**Yuntao Ma**

First Teaching Hospital of Tianjin University of Traditional Chinese Medicine

**Qinghua Zhang**

First Teaching Hospital of Tianjin University of Traditional Chinese Medicine

**Bin Wang**

First Teaching Hospital of Tianjin University of Traditional Chinese Medicine

**Bin Wang**

**robin\_e@sina.com**

First Teaching Hospital of Tianjin University of Traditional Chinese Medicine

---

## Article

**Keywords:** Obesity, Depression, Anxiety, Biomarkers

**Posted Date:** September 1st, 2026

**DOI:** <https://doi.org/10.21203/rs.3.rs-10259789/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.  
[Read Full License](#)

**Additional Declarations:** No competing interests reported.

---

# Investigating biomarkers and regulatory mechanisms associated with depression and anxiety in obesity based on transcriptome sequencing and experimental validation

Xiaochen Ren<sup>a,b,#</sup>, Ziqi Zhao<sup>a,b,#</sup>, Ning Xu<sup>a,b</sup>, Qianjing Yang<sup>a,b</sup>, Yuyang Lan<sup>a,b</sup>, Yuntao Ma<sup>c,d</sup>, Qinghua Zhang<sup>a,b</sup>, Bin Wang<sup>a,b</sup>, Bin Wang<sup>a,b\*</sup>

<sup>a</sup> Department of Endocrinology, First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin 300381, China.

<sup>b</sup> Department of Endocrinology, National Clinical Research Center for Chinese Medicine Acupuncture and Moxibustion, Tianjin 300381, China.

<sup>c</sup> International Medical Center, First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin 300381, China.

<sup>d</sup> International Medical Center, National Clinical Research Center for Chinese Medicine Acupuncture and Moxibustion, Tianjin 300381, China.

# These authors contributed equally to this work: Xiaochen Ren and Ziqi Zhao.

\* Corresponding author.

Correspondence to: Professor BinWang, Department of Endocrinology, First Teaching Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin 300381, China. E-mail address: robin\_e@sina.com.

## Abstract

**Background:** Obesity is a chronic metabolic disease. Emerging evidence underscores a intricate, bidirectional association between obesity and comorbid depression and anxiety disorders. However, biomarkers associated with depression and anxiety in obesity have not yet been investigated.

**Methods:** This study utilized public databases and employed machine learning algorithms combined with expression validation to screen for biomarkers. A predictive nomogram was developed using the selected markers, followed by comprehensive characterization via Gene Set Enrichment Analysis (GSEA) and immune infiltration profiling. Furthermore, unsupervised clustering was employed to stratify obesity into distinct molecular subtypes. Potential therapeutic agents for these subtypes were predicted, with binding affinities confirmed through molecular docking simulations. Finally, the expression levels of biomarkers in clinical samples were validated using reverse transcription quantitative polymerase chain reaction (RT-qPCR).

**Results:** Seven feature genes were identified using machine learning. Among these,

CPOX and SLC35F5 were significantly downregulated in obesity groups across 2 datasets ( $P < 0.05$ ) and were defined as biomarkers. A nomogram constructed from these two genes displayed robust discriminative power and calibration accuracy (Area Under the Curve, AUC > 0.7). GSEA analysis showed that the Toll-like receptor signaling pathway was commonly enriched by both biomarkers. An increase in mesenchymal stem cells was observed in obesity samples. Consensus cumulative distribution function (CDF) analysis classified obese patients in the training set into 2 subtypes: C1 and C2. Molecular docking results showed that among potential drugs for subtype C1, mazindol exhibited the strongest binding affinity to CPOX (-13.1 kcal/mol), while for subtype C2, FK-866 showed the strongest binding affinity to CPOX (-12.6 kcal/mol). Finally, RT-qPCR validation demonstrated that CPOX and SLC35F5 were significantly down-regulated in the Obesity group.

**Conclusion:** Through bioinformatic analysis, the roles of CPOX and SLC35F5 in the regulatory mechanisms of obesity were elucidated, providing a new research basis for early intervention in obesity.

**Keywords:** Obesity, Depression, Anxiety, Biomarkers

## 1. Introduction

Obesity is defined as a chronic, relapsing condition marked by excessive adipose tissue accumulation, internationally classified by a body mass index (BMI)  $\geq 30$  kg/m<sup>2</sup> (with  $\geq 25$  kg/m<sup>2</sup> being more sensitive for East Asian populations) [1,2]. Its pathogenesis arises from a confluence of genetic susceptibility and environmental or behavioral determinants, including high-energy-density diets, sleep disruptions, and specific pharmacotherapies, culminating in visceral fat deposition and sustained low-grade inflammation[3,4]. Treatment is based on lifestyle interventions [5], supplemented when necessary by weight management medications represented by GLP-1 or GIP/GLP-1 receptor agonists [6]. For severely obese patients, surgical interventions (such as gastric bypass) are effective options. Therefore, obesity management requires a comprehensive consideration of etiology, mechanisms, and individualized treatment strategies.

Depression is a type of mental disorder characterized by persistent and significant low mood or loss of interest and pleasure [7], while anxiety disorders are a group of mental disorders centered around excessive fear and anxiety, along with related behavioral and cognitive changes [8]. Both have a profound effect on mental well-being, physical health, and social functioning [9]. Recent investigations have uncovered a sophisticated, reciprocal association linking depression, anxiety disorders, and obesity. Specifically, obese persons exhibit a 55% elevated probability of developing depression compared to the broader population; conversely, those suffering from depression face a 58% heightened likelihood of becoming obese [10]. Studies have also shown that obesity amplifies the incidence of depression and anxiety. Metabolic and vascular dysfunctions caused by obesity, such as chronic inflammation, resistance to insulin and leptin, and elevated blood pressure, are now recognized as major contributors to the onset of depression and anxiety disorders [11]. Therefore, exploring biomarkers associated with depression and anxiety in obesity is highly important. This research not only helps uncover the intricate connections among the three factors but also lays the foundation for early diagnosis and targeted treatment. Comprehensive analysis of these markers will facilitate the design of interventions addressing shared pathological mechanisms, ultimately enhancing the physical and psychological outcomes for obese individuals.

In summary, this study integrated transcriptomics and bioinformatics methods to screen obesity-related genes, identified biomarkers through machine learning and expression validation, conducted a series of analyses including nomogram construction, molecular network regulation, and immune infiltration, and validated the related genes via PCR. Emphasizing the potential influence of depression and anxiety within the context of obesity, our work offers a rational basis for timely clinical intervention.

## **2. Materials and methods**

### **2.1 Data source**

Two transcriptome datasets associated with obesity were obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) in order to

thoroughly investigate the expression patterns and underlying processes of depression and anxiety disorders in obesity. These datasets were utilized as the training set and validation set for the study. Based on the GPL570 platform, the training set, GSE109597, consists of 43 control whole blood samples and 41 obesity whole blood samples (with  $\text{BMI} \geq 25 \text{ kg/m}^2$  defined as obesity). The validation set is GSE69039, based on the GPL10558 platform, and includes 14 obesity peripheral blood mononuclear cell samples and 3 control peripheral blood mononuclear cell samples. Additionally, a dataset for comorbid depression and anxiety, GSE98793, was downloaded from the GEO database. This dataset uses the GPL570 platform and consists of 64 whole blood samples from patients with comorbid depression and anxiety and 64 control whole blood samples (note: data from 64 patients with depression without anxiety were excluded).

## **2.2 Acquisition of candidate genes (CGs)**

To identify differentially expressed genes (DEGs1) between obesity and control groups, differential expression analysis was conducted on the training dataset GSE109597 using the "limma" package (v3.56.2)[12]. Genes meeting the criteria of  $|\log_2\text{Fold Change (FC)}| > 0.25$  and  $P < 0.05$  were considered significantly differentially expressed. To illustrate the overall distribution of expression changes, volcano plots and heatmaps were generated using the "ggplot2" package (v 3.4.1)[13] and the "pheatmap" package (v 1.0.12)[14]. The 10 most significantly up-regulated and down-regulated genes, ranked according to  $|\log_2\text{FC}|$ , were annotated in the figures.

Similarly, differentially expressed genes (DEGs2) between anxiety-depression comorbid samples and control samples were identified from the GSE98793 dataset using the same analytical criteria ( $|\log_2\text{FC}| > 0.25$ ,  $P < 0.05$ ). Volcano plots and heatmaps were also constructed to visualize the global expression patterns of these genes. In the corresponding plots, the top 10 up-regulated and down-regulated genes were marked according to their  $|\log_2\text{FC}|$  values.

To further identify potential differentially expressed genes linked to anxiety and depression in obesity, an intersection analysis was performed between the previously

identified DEGs1 and DEGs2. The overlap between the two gene sets was visualized using a Venn diagram created with the "ggvenn" package (v 0.1.10)[15]. Genes that appeared in both sets were extracted and considered as candidate genes (CGs).

### **2.3 Enrichment analysis**

The Gene Ontology (GO) comprises three functional dimensions: Molecular Functions (MFs), Cellular Components (CCs), and Biological Processes (BPs), which are used to systematically describe the functional attributes of genes. Complementarily, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to map metabolic and signaling pathways within complex biological systems[16,17]. To further explore the potential biological functions and pathways involving the CGs in obesity, GO and KEGG enrichment analyses were performed using the "clusterProfiler" package (v 4.2.2)[18]. A significance threshold of adjusted  $P < 0.05$  was applied for screening. The top 5 enriched GO terms and KEGG pathways, ranked according to the number of enriched genes, were subsequently selected and presented to illustrate the potential biological mechanisms involved. KEGG pathway data were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Permission to use KEGG pathway-related materials in this article was obtained from Kanehisa Laboratories.

### **2.4 Identification of biomarkers**

To identify biomarkers associated with obesity from the CGs, two machine learning algorithms were employed for feature selection based on the training dataset GSE109597. First, Least Absolute Shrinkage and Selection Operator (LASSO) regression was applied to the CGs for feature screening and variable selection using the "glmnet" package (v 2.1.1.1)[19]. A five-fold cross-validation procedure was employed, and the model was fitted under the binomial family. The minimum lambda value was selected as the optimal penalty parameter, and genes with non-zero coefficients were retained as LASSO feature genes. Second, Support Vector Machine Recursive Feature Elimination (SVM-RFE) was applied through the "caret" package (v 6.0.94)[20], also using five-fold cross-validation. To improve the robustness of feature screening, the outputs generated by the two machine learning algorithms were

compared and integrated. Their shared candidates were subsequently visualized with the "ggplot2" package (v 3.4.1). The common features identified from both methods were defined as characteristic genes for further downstream analysis.

To further validate expression differences in characteristic genes between obese individuals and controls, and to identify biomarkers with stable expression patterns across datasets, the Wilcoxon test was conducted separately in the training dataset GSE109597 and the validation dataset GSE69039. Expression profiles were evaluated between the two groups, and statistical significance was defined as  $P < 0.05$ . Genes showing significant differential expression in both datasets, together with concordant regulatory trends, were retained as candidate biomarkers for the present study. To further illustrate these findings, graphical visualization was performed using the "ggplot2" package (v 3.4.1), which enabled intuitive presentation of expression variations and overall distribution patterns across the compared groups.

## **2.5 Construction and evaluation of the nomogram**

Based on the expression levels of biomarkers, a nomogram was constructed using all samples in the training set to predict the probability of obesity. The "rms" package (v 6.8.1)[21] was employed to develop a biomarker-based nomogram for disease prediction. In the constructed nomogram, each prognostically relevant gene is assigned a numerical score (Points) based on its relative contribution to disease prediction. The aggregation of individual gene scores yields a composite value (Total Points), which is subsequently mapped to an estimated probability of disease onset or progression. To further assess the discriminatory capacity of the model, Receiver Operating Characteristic (ROC) curves were generated using the "pROC" package (v 1.18.5)[22]. The Area Under the Curve (AUC) was computed as a quantitative index of predictive performance, with values falling within the range of 0.7 to 1.0 considered indicative of satisfactory discriminatory ability. This analytical framework was applied to both the GSE109597 and GSE69039 validation cohorts to ensure cross-dataset generalizability of the nomogram. Additionally, calibration analysis was conducted to examine the agreement between model-predicted probabilities and observed clinical outcomes. Calibration curves were constructed with the aid of the

"rms" package (v 6.5.0), wherein a Hosmer-Lemeshow test  $P > 0.05$  was interpreted as evidence of adequate model calibration and goodness-of-fit.

## **2.6 Tissue Localization and Subcellular Localization Analysis**

Tissue-specific expression profiles of the identified biomarkers were systematically evaluated through the Human Protein Atlas database (HPA, <https://www.proteinatlas.org/>), a comprehensive repository that catalogues protein expression across diverse human tissues and organs. To facilitate a structured and interpretable visualization of the resulting expression data, histogrammic representations were constructed, enabling comparative assessment of distributional heterogeneity among distinct tissue types. For each biomarker, the five tissues showing the highest expression levels were selected and imported into Cytoscape (v 3.8.2) to construct a tissue-biomarker network.

Understanding the subcellular localization information of proteins can help predict their biological functions, thereby providing research directions for elucidating the mechanisms of gene action. Biomarker FASTA sequences were retrieved from NCBI's genomic repository. Subcellular localization was subsequently predicted using the mRNALocator platform (<http://bio-bigdata.cn/mRNALocator/>), wherein the location yielding the highest predictive score was designated as the definitive subcellular compartment for each biomarker. Visualization of the resulting localization distribution was accomplished through construction of a stacked bar chart, implemented via the "ggplot2" package (v 3.4.1).

## **2.7 Gene set enrichment analysis (GSEA)**

To investigate the biological roles of these identified biomarkers within the pathological mechanisms underlying obesity development, Spearman correlation analysis was conducted utilizing the training dataset GSE109597. Specifically, the computational implementation relied upon the "psych" package (v 2.1.6)[23], through which correlation coefficients were systematically derived to quantify the associative relationships between each candidate biomarker and the remaining genes across the dataset. The genes were ranked in descending order based on the correlation coefficients. The c2.cp.kegg.v7.5.1.symbols.gmt dataset, retrieved from

the Molecular Signatures Database (MSigDB, <http://software.broadinstitute.org/gsea/msigdb>), was employed as the reference background for downstream analyses. Gene Set Enrichment Analysis (GSEA) was subsequently performed through the "clusterProfiler" package (version 4.2.2) to identify functionally relevant biological pathways. To facilitate interpretability, the ten pathways demonstrating the most substantial statistical significance, as determined by p-value ranking, were selected and subjected to visualization. Pathways were considered significantly associated with individual biomarkers when the following criteria were simultaneously satisfied:  $|\text{normalized enrichment score (NES)}| > 1$ ,  $P < 0.05$ , and false discovery rate (FDR)  $< 0.25$ . These filtered results were subsequently used to infer potential molecular mechanisms and related pathological processes.

## **2.8 Immune Infiltration Analysis**

The infiltration of immune cells has increasingly been identified as a pivotal determinant underlying the onset and advancement of numerous pathological conditions, with particular relevance observed in malignancies and disorders of immunological origin. Growing evidence suggests that the dynamic interplay between infiltrating immune populations and the surrounding tissue microenvironment plays a fundamental role in shaping disease trajectories, influencing both early-stage pathogenesis and long-term clinical outcomes. This phenomenon extends beyond neoplastic contexts, encompassing a broad spectrum of chronic and systemic conditions in which dysregulated immune cell recruitment serves as a central mechanistic driver. The relationship between the immune system and disease processes can be better understood by evaluating immune cell infiltration within tissues, including its degree, cellular composition, and functional status. Such analyses are considered valuable, as they support the exploration of disease mechanisms, offer potential targets for therapeutic intervention, and contribute to the advancement of drug development.

Immune cell infiltration patterns were systematically compared between obesity-associated and control specimens through application of the ssGSEA algorithm. This computational approach, incorporated within the "GSVA" package (v

1.49.4)[24], facilitated quantitative estimation of relative infiltration levels across 64 distinct immunological cell populations. Abundance profiling was conducted on a per-sample basis, drawing upon the GSE109597 training dataset [25] as the foundational reference dataset for downstream analytical procedures. Based on these results, stacked bar plots illustrating the proportional distribution of the 64 immune cell populations in both groups were generated with the "ggplot2" package (v 3.4.1). Immune cell populations demonstrating statistically meaningful distinctions across both cohorts were identified through infiltration-level analysis conducted via the Wilcoxon test within the training dataset, applying a significance criterion of  $P < 0.05$ . Graphical representation of these comparative statistical analyses was subsequently accomplished through utilization of the "Stats" package (v 4.3.3)[26].

To further explore the correlations among differentially expressed immune cells and between biomarkers and these immune cells, Spearman correlation analysis was performed across all samples in the training set GSE109597 using the "psych" package (v 2.1.6). Correlations with  $P < 0.05$  and  $|\text{correlation coefficient (cor)}| > 0.3$  were considered significant. For visualization purposes, a correlation heatmap was produced with the "psych" package (v 2.1.6).

## **2.9 Unsupervised clustering identified distinct obesity subtypes.**

To further characterize inter-sample heterogeneity and molecular profiles within the patient cohort, unsupervised clustering analysis was performed based on CGs. The consensus cumulative distribution function (CDF) curve was constructed by employing the "ConsensusClusterPlus" package (v 1.64.0)[27]. Within the training dataset GSE109597, obesity-related specimens were partitioned into discrete molecular subgroups, with the optimal cluster number established through systematic clustering evaluation. To elucidate the molecular distinctions underlying these classification patterns, an inter-group differential expression analysis was conducted via the "limma" package (v 3.56.2), culminating in the characterization of differentially expressed genes (DEGs3). The overall distributional landscape of DEGs3 was subsequently illustrated through volcano plots alongside heatmap visualizations, which were rendered respectively by utilizing the "ggplot2" package (v

3.4.1) and the "pheatmap" package (v 1.0.12). The top 10 genes with the highest  $|\text{Log}_2\text{FC}|$  values, both up-regulated and down-regulated, were highlighted in the visualizations.

Filtered differentially expressed genes (DEGs) underwent functional characterization through two complementary bioinformatic frameworks: Gene Ontology (GO) classification and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping. Enrichment computations were executed via the "clusterProfiler" package (v 4.2.2), wherein biological terms and metabolic routes demonstrating adjusted  $P < 0.05$  threshold were deemed statistically meaningful. From these results, the top 5 GO terms and KEGG pathways, ranked according to p-value significance, were subsequently selected for presentation.

Molecular subgroup-associated biological pathways were examined through differential expression analysis conducted on subtype classifications within the training cohort. Employing the "limma" package (v 3.56.2), expression profiles were systematically compared across subtypes, with  $\text{log}_2\text{FC}$  values subsequently derived for the complete gene set to quantify transcriptional variation between groups. Genes were ranked according to  $\text{log}_2\text{FC}$  in descending order. The KEGG gene set (c2.cp.kegg.v7.5.1.symbols.gmt) was downloaded from the Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb>) as the reference gene set. GSEA analysis was conducted by employing the "clusterProfiler" package (v 4.2.2) with thresholds of  $|\text{NES}| > 1$ ,  $\text{FDR} < 0.25$ , and  $P < 0.05$ . To prioritize the most biologically meaningful results, the 10 highest-ranking signaling pathways, as determined by their degree of enrichment significance quantified through p-value ordering, were subsequently selected for visualization and further interpretive analysis.

Subsequently, to elucidate the divergent pathway activities among molecular subgroups and uncover putative biological mechanisms alongside actionable therapeutic targets, an initial differential expression analysis was carried out across subtypes within the training cohort, leveraging the "limma" package, whereby  $\text{log}_2\text{FC}$  values were systematically derived. All candidate genes were subsequently arranged

in descending sequence according to their respective  $\log_2FC$  magnitudes. Drawing upon the KEGG gene collection (c2.cp.kegg.v7.5.1.symbols.gmt) obtained from the MSigDB repository, subtype-specific enrichment scores were estimated via the ssGSEA algorithm as operationalized within the "GSVA" package (v 1.49.4), applied exclusively to obesity-related specimens from the training dataset. A comparative analysis of KEGG pathway-level enrichment scores across the defined subgroups was then executed through the "limma" package (v 3.56.2), enabling the systematic delineation of statistically significant inter-subtype pathway discrepancies. The top 5 pathways showing the greatest and least  $|\log_2FC|$  differences were identified. Pathway enrichment patterns were visualized across samples by constructing a heatmap through implementation of the "pheatmap" package (v 1.0.12), which facilitated systematic representation of score distributions for both the highest- and lowest-ranked biological pathways. Concurrently, comparative analysis of pathway activity variations among experimental groups was conducted via box-and-whisker plots, rendered through utilization of the "ggplot2" package (v 3.4.1).

## **2.10 Drug Prediction and Molecular Docking**

Compound signatures targeting obesity-associated subtypes were retrieved from the Connectivity Map (CMap) repository (<https://clue.io/>) to identify candidate therapeutics. The query signature comprised the 100 most significantly upregulated and downregulated transcripts characterizing the specific subtype. CMap generates an enrichment score spanning from -100 to 100; notably, negative values for small molecules denote a capacity to invert the disease-associated transcriptional profile, thereby implying therapeutic potential. Consequently, the 5 compounds with the most negative scores were prioritized for subsequent analysis.

To further explore the potential therapeutic mechanisms of the predicted drugs, molecular docking analysis was conducted between the selected small-molecule compounds and the identified biomarkers. Three-dimensional molecular conformations of selected candidate compounds were obtained in SDF format via the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>), whereas crystallographic structures of the target proteins of interest were sourced from the Protein Data Bank

(PDB) (<https://www.rcsb.org/>). Prior to computational analysis, all retrieved protein structures underwent systematic preprocessing within the PyMOL molecular visualization environment (v 3.1.3) (Rosignoli and Paiardini, 2022), during which bound water molecules alongside residual solvent components were eliminated to ensure structural integrity suitable for subsequent docking simulations. Hydrogen atoms were added and atomic charges were assigned prior to protein preparation. Protein structures were subsequently converted into PDBQT format using AutoDockTools (v 1.5.7)[28]. Prior to initiating the docking procedure, each pharmaceutical compound underwent structural format conversion into mol2 files via the Open Babel toolkit. Subsequently, these converted structures were loaded into AutoDockTools (v 1.5.7), wherein polar hydrogen atoms were systematically incorporated and the resulting molecular configurations were exported in PDBQT format to ensure compatibility with downstream computational analyses. To define the spatial boundaries of the docking region, a three-dimensional grid box was manually delineated to encompass the putative binding pocket as identified through prior binding site prediction. Following this preparatory phase, molecular docking calculations were executed within the AutoDock simulation framework to evaluate the binding affinity and probable interaction modes between each ligand candidate and the target receptor. Three-dimensional visualization of the docking results was performed using PyMOL. Molecular docking conformations were evaluated and prioritized on the basis of their computed binding free energy values, with thresholds  $\leq -5$  kcal/mol serving as the criterion for classifying a compound as exhibiting robust intermolecular affinity toward the target receptor. Through systematic examination of these energetically favorable binding configurations, the study yields mechanistic insights into how structurally diverse candidate molecules may exert their pharmacological effects at the molecular level.

## **2.12 Reverse transcription-quantitative PCR (RT-qPCR)**

Clinical samples for validation were obtained from First Teaching Hospital of Tianjin University of Traditional Chinese Medicine. Specifically, total RNA was extracted from whole blood specimens obtained from five obese patients and five control

subjects using TRIzol reagent. Complementary DNA (cDNA) was subsequently synthesized with the HP All-in-one qRT Master Mix II kit. Real-time quantitative PCR (RT-qPCR) was carried out using the 2×Universal Blue SYBR Green qPCR Master Mix. The primer sequences for the target biomarkers and the internal reference gene (GAPDH) are provided in **Table 1**. Relative gene expression levels were calculated using the  $2^{-\Delta\Delta C_t}$  method following RT-qPCR. Statistical comparisons between groups were performed using the t-test, with  $P < 0.05$  considered statistically significant. Data visualization was conducted using GraphPad Prism software.

### 2.13 Ethical approval

The clinical validation component of this study was approved by the Institutional Review Board of the First Teaching Hospital of Tianjin University of Traditional Chinese Medicine (approval number: TYLL2025[Z]056). All methods involving human participants and human blood samples were carried out in accordance with relevant guidelines and regulations, including the Declaration of Helsinki. Written informed consent was obtained from all participants before sample collection.

**Table 1 Primer sequences**

| Primer    | Sequence (5′ →3′ )   |
|-----------|----------------------|
| CPOX-F    | TCTACGCCATCCAAGGGACT |
| CPOX-R    | GGAGAAGACTAAGGCACGGG |
| SLC35F5-F | GGTGACAGCTCTGTGAGTCC |
| SLC35F5-R | GAGACCTTCACGGTTTCGGT |
| GAPDH-F   | ATGGGCAGCCGTTAGGAAAG |
| GAPDH-R   | AGGAAAAGCATCACCCGGAG |

### 2.13 Statistical Analysis

The R software (v 4.2.2) was applied to all analyses. Besides, the Wilcoxon test was employed to assess group disparities, setting significance threshold at  $P < 0.05$ .

## 3. Results

### 3.1 Identification of 10 CGs and biological function analysis

Differential expression profiling compared obesity samples against controls within the

training dataset, applying filtering criteria of  $|\log_2FC| > 0.25$  and  $P < 0.05$ . This analysis yielded 236 differentially expressed genes (DEGs), consisting of 38 upregulated and 198 downregulated transcripts (**Supplementary Table S1**). The distribution and significance of DEGs1 were visualized through volcano plots and heatmaps (**Fig 1A**). The 10 most significantly up-regulated and down-regulated genes were subsequently annotated in the visualization. Specifically, compared to control samples, SLIP, FOLR3, RNASE3, PGM5, EIF1AY, GSTM3, C4BPA, DDX3Y, PF4V1, and RPS4Y1 were significantly up-regulated in obesity samples, while IL12RB2, SLC41A2, DSEL, CAV1, HMMR, ROBO1, KLHDC1, CXCL8, MYOM2, and OCLN were significantly down-regulated.

Using the same threshold criteria ( $|\log_2FC| > 0.25$  and  $P < 0.05$ ), a comparative transcriptomic analysis was conducted to examine gene expression differences between specimens exhibiting anxiety-depression comorbidity and healthy control specimens within the GSE98793 cohort. This systematic screening procedure yielded a comprehensive catalogue of 142 genes demonstrating statistically significant differential expression patterns. Among these transcriptomic alterations, 113 genes exhibited notable upregulation, whereas the remaining 29 genes manifested measurable downregulation, collectively providing a molecular foundation for subsequent pathway enrichment investigations (**Supplementary Table S2**). Volcano plots and heatmaps illustrated the distribution and statistical significance of these DEGs2 (**Fig 1B**), highlighting the top 10 most significantly altered genes in each direction. Specifically, compared to control samples, MMP8, OLFM4, CEACAM8, CEACAM6, CRISP3, LTF, LCN2, ARG1, BPI, and DEFA4 were significantly up-regulated in anxiety-depression samples, while GZMK, KLRG1, ERAP2, PYHIN1, HLA-DQA1, FCER1A, TREML4, FAM3B, CCDC91, and CLC were significantly down-regulated.

To identify CGs associated with anxiety and depression in obesity, the intersection of 236 DEGs1 and 142 DEGs2 was analyzed, yielding 10 CGs related to anxiety and depression (**Fig 1C**). These genes are CPOX, TMTC4, TMEM170B, CEP170, ATP8B4, SLC25A40, SCAPER, ZMAT1, NHS, and SLC35F5, providing candidate

targets for understanding the progression from anxiety and depression to obesity.

GO functional enrichment analysis of these 10 CGs revealed significant enrichment in only 6 molecular functions (MFs) terms ( $P_{\text{adjust}} < 0.05$ ) (**Supplementary Table S3**). Functional annotation of the identified terms revealed predominant enrichment in several enzymatic and structural categories. Specifically, oxidoreductase-related functions targeting CH-CH group donors with molecular oxygen serving as the electron acceptor were prominently represented, alongside structural constituents integral to ocular lens architecture. Additional enriched categories encompassed mannosyltransferase catalytic activity, as well as both ATPase-dependent and ATPase-independent intramembrane phospholipid transporter functions (**Fig 1D**). Complementary pathway-level analysis conducted via the KEGG database revealed statistically significant enrichment in 2 discrete metabolic pathways ( $P < 0.05$ ), including Porphyrin metabolism and Biosynthesis of cofactors, suggesting coordinated regulatory involvement of these biological processes in the context under investigation (**Fig 1E**).

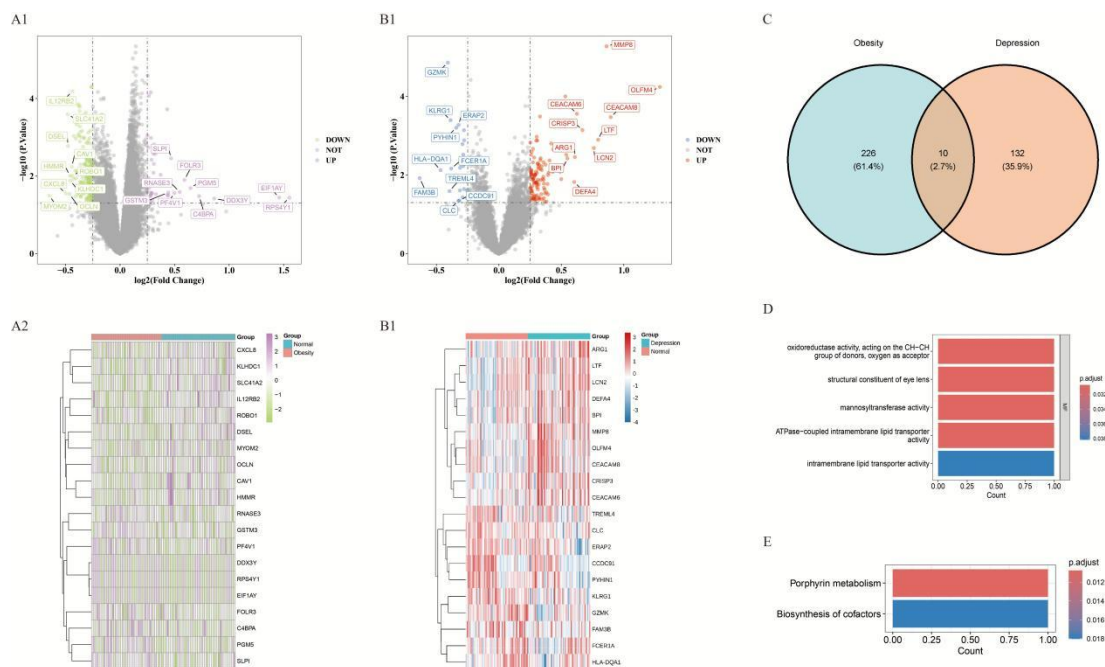


**FIGURE 1 Differential Expression and Biomarker Identification in Obesity.** (A1) Volcano Plot of DEGs in Obesity. (A2) Heatmap of DEGs in Obesity. (B1) Volcano Plot of DEGs in Anxiety-Depression Comorbidity. (B2) Heatmap of DEGs in

Anxiety-Depression Comorbidity. (C) Intersection of DEGs in Obesity and Anxiety-Depression. (D) Functional Enrichment of Biomarkers. (E) Pathway Enrichment Analysis of Biomarkers. KEGG pathway data used in this figure were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database with permission from Kanehisa Laboratories.

### 3.2 CPOX, SLC35F5 as biomarkers

Subsequently, to further identify characteristic genes, 2 machine learning methods were employed for gene screening. First, LASSO regression was used for variable compression and selection. At  $\log(\lambda_{\min}) = -4.19$ , the model's prediction error (i.e., cross-validation error) reached its minimum, identifying 8 LASSO feature genes: CPOX, TMTC4, TMEM170B, ATP8B4, SLC25A40, SCAPER, NHS, and SLC35F5 (**Fig 2A**). Second, SVM-RFE analysis identified 7 SVM-RFE feature genes (**Fig 2B**), including CPOX, TMTC4, NHS, SCAPER, ATP8B4, SLC25A40, and SLC35F5. Finally, the intersection of both methods was taken, resulting in 7 common feature genes: CPOX, TMTC4, NHS, SCAPER, ATP8B4, SLC25A40, and SLC35F5 (**Fig 2C**). These were defined as characteristic genes and served as core molecules for subsequent diagnostic and mechanistic studies.

To assess the consistency of expression patterns for these 7 characteristic genes across the training set GSE109597 and the validation dataset GSE69039, comparative expression analysis between obesity and control samples was performed. Significant differential expression was identified for two genes, CPOX and SLC35F5 ( $P < 0.05$ ). Both genes were consistently down-regulated in obesity samples across the two datasets. These 2 genes were identified as biomarkers for subsequent studies (**Fig 2D**).

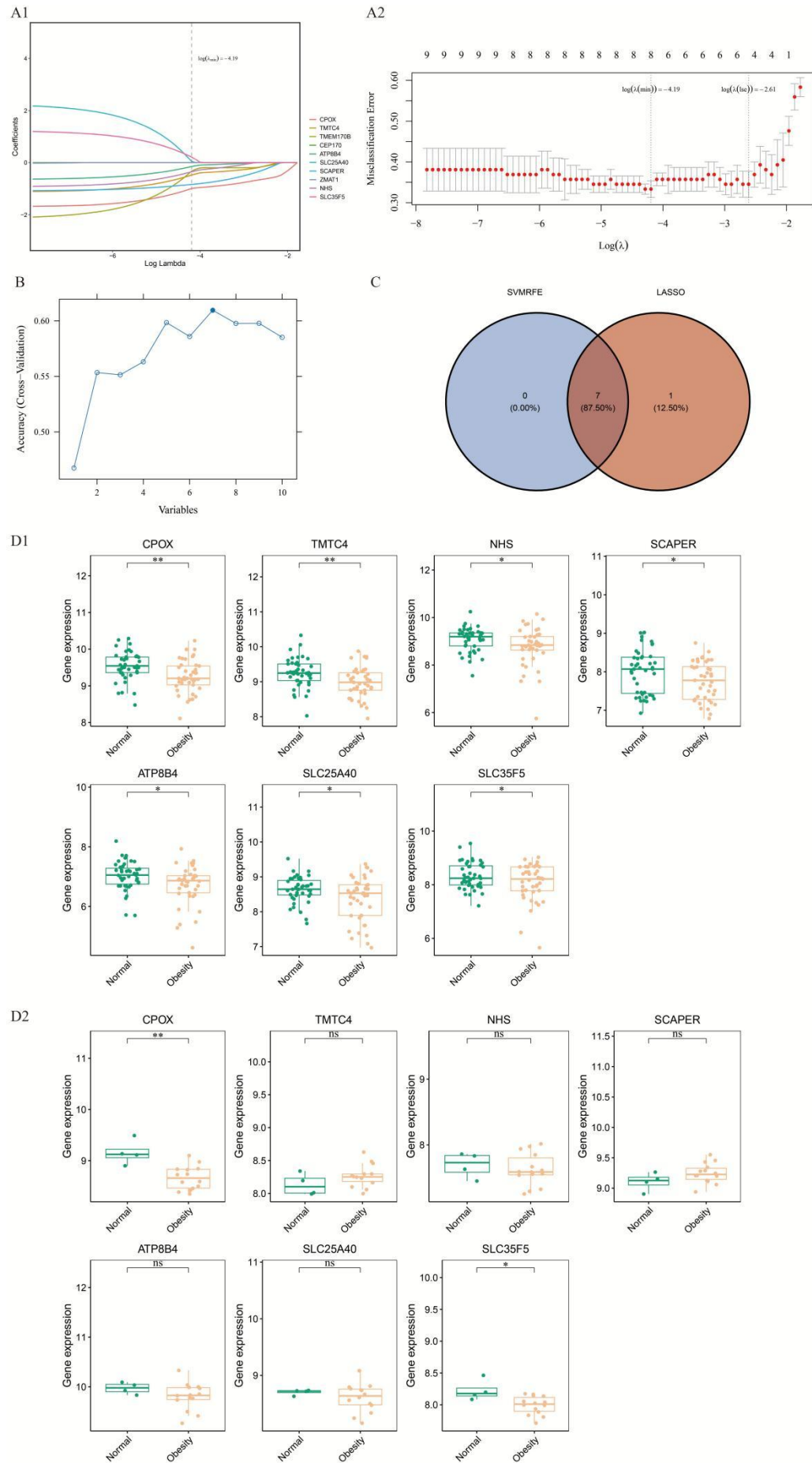


**FIGURE 2 Biomarker Identification Using Machine Learning in Obesity.** (A1) LASSO Regression Feature Selection. (A2) SVM-RFE Feature Selection. (B) Venn Diagram of Common Feature Genes. (C) Comparison of Gene Expression. (D1) Obesity Prediction Nomogram. (D2) Nomogram Calibration Curve.

### **3.3 Nomogram prediction model for CPOX and SLC35F5**

To further predict the probability of obesity, a nomogram model was constructed by integrating biomarkers. Within the nomographic framework, discrete point allocations were systematically assigned to each incorporated variable, encompassing both CPOX and SLC35F5, whereby every individual gene corresponded to a specifically delineated scoring interval. An aggregate score was subsequently derived through the summation of all constituent biomarker values, such that progressively elevated cumulative totals served as indicators of heightened disease susceptibility. The results showed that the total score for obese individuals was 84.2, with an odds ratio of 1.05 (**Fig 3A**). The predictive performance of the model was evaluated through calibration analysis, wherein the calibration curve demonstrated substantial concordance between the nomogram-derived probability estimates and the ideal reference standard, indicating robust agreement across the predicted risk spectrum. A p-value of 0.708, which is greater than 0.05, indicated high predictive accuracy (**Fig 3B**). Additionally, the Hosmer-Lemeshow (HL) test was subsequently employed to evaluate the concordance between model-generated predictions and empirically observed values. The resulting p-value of 0.562 surpassed the conventional significance threshold of 0.05, indicating an absence of statistically meaningful discrepancy between predicted probabilities and actual outcomes. This finding provides evidence of satisfactory calibration and confirms that the logistic regression model demonstrates adequate fit to the observed data (**Fig 3C**). In conclusion, CPOX and SLC35F5 serve as biomarkers with high predictive value for the early diagnosis of obesity. The nomogram model further enhances their potential for clinical application.

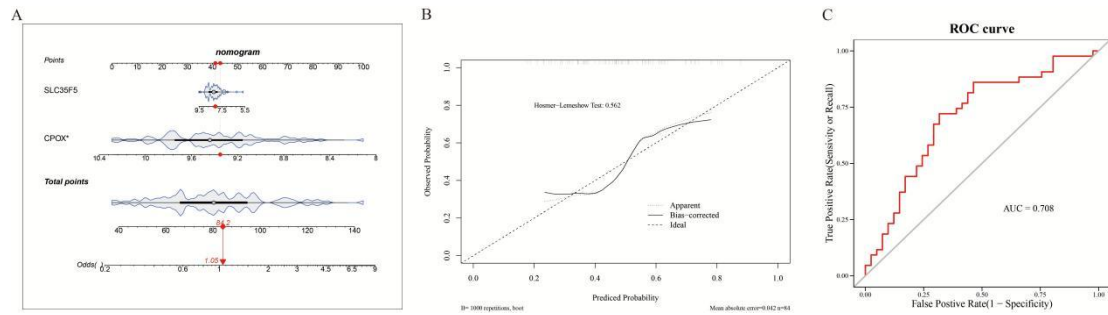


**FIGURE 3 Biomarker Identification and Validation in Obesity and Comorbidities.** (A) Venn Diagram of Key Genes. (B) Differential Gene Expression Heatmap. (C) Feature Gene Selection by Machine Learning.

### 3.4 Tissue-specific expression analysis of CPOX and SLC35F5

The expression patterns of biomarkers CPOX and SLC35F5 across various tissues were analyzed. CPOX showed high expression in the bone marrow, liver, duodenum, placenta, and small intestine, while SLC35F5 was predominantly expressed in the thyroid gland, adrenal gland, choroid plexus, skin, and kidneys. (**Fig 4A**).

Deeper investigation into the intracellular compartmentalization of both CPOX and SLC35F5 demonstrated distinct spatial segregation between these two proteins. Specifically, immunolocalization data indicated that CPOX exhibited predominant accumulation within the cytoplasmic compartment, while SLC35F5 displayed preferential enrichment in the nuclear fraction (**Fig 4B**).

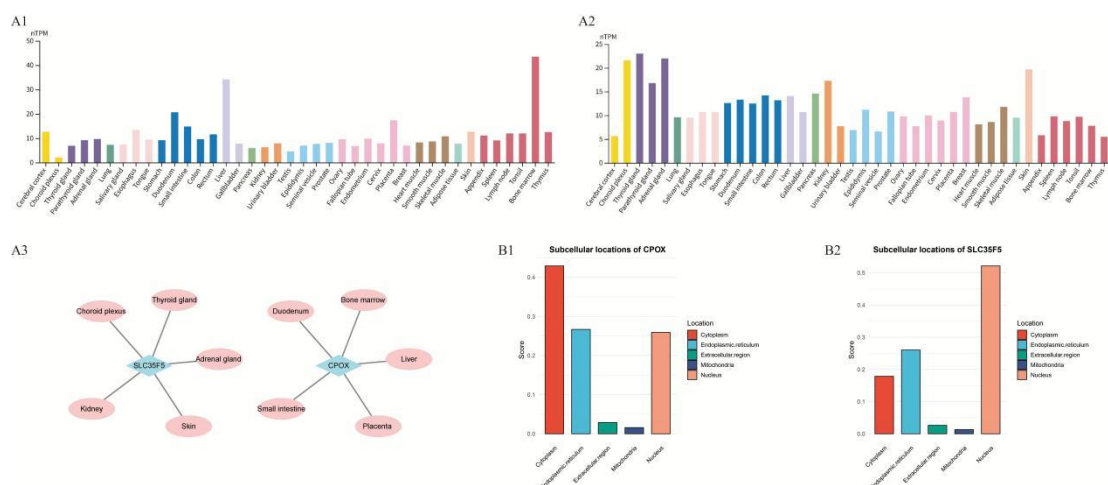


**FIGURE 4 Immune Profiling and Molecular Pathway Analysis in Obesity. (A1)**

Volcano Plot of DEGs in Obesity. (A2) Heatmap of Differential Expression in Obesity. (A3) Venn Diagram of Overlapping DEGs. (B1) KEGG Pathway Enrichment Analysis. (B2) Gene Set Enrichment Analysis (GSEA) Results.

### **3.5 GSEA and immune infiltration analysis of CPOX and SLC35F5**

GSEA analysis was performed to investigate the potential biological functions of the biomarkers CPOX and SLC35F5 in obesity pathogenesis. Significant enrichment of both genes was observed in multiple KEGG pathways, including ubiquitin-mediated proteolysis, Huntington's disease, glycolysis and gluconeogenesis, proteasome, and Toll-like receptor signaling pathway. Most of these pathways are closely associated with obesity (**Fig 5A, Supplementary Table S4**).

To explore the immune regulatory mechanisms potentially linked to the biomarkers CPOX and SLC35F5, the xCell computational framework was employed to quantify the proportional representation of 64 distinct immune and stromal cellular subtypes across all analyzed specimens. Within obesity-derived samples, a notable perturbation in the equilibrium between T helper 1 (Th1) and T helper 2 (Th2) lymphocyte populations was identified, a finding accompanied by a concurrent elevation in the proportional presence of mesenchymal stem cells (**Fig 5B**). Subsequent statistical analysis employing the Wilcoxon test demonstrated statistically meaningful distinctions ( $P < 0.05$ ) regarding the degree of immune cell infiltration patterns when comparing adiposity-affected specimens against non-obese control counterparts. A total of nine immune cell types exhibiting differential infiltration were subsequently determined, including adipocytes, basophils, CD8<sup>+</sup> central memory T cells, epithelial cells, megakaryocytes, mesenchymal stem cells, myocytes, as well as Th1 and Th2 cells. These findings indicate complex alterations in immune cell infiltration patterns in obese patients (**Fig 5C**). Correlation analysis among differentially infiltrating immune cells was subsequently performed. Notably, a robust co-occurrence pattern emerged between Th1 lymphocytes and mesenchymal stromal cells ( $r = 0.65$ ,  $P < 0.001$ ), suggesting coordinated immunomodulatory activity. An equally pronounced positive relationship was observed linking CD8<sup>+</sup> central memory T lymphocytes with

Th2 lymphocytes ( $r = 0.65$ ,  $P < 0.001$ ), indicative of potential cooperative adaptive immune regulation. Conversely, CD8<sup>+</sup> central memory T lymphocytes demonstrated a marked inverse relationship with basophilic granulocytes ( $r = -0.53$ ,  $P < 0.001$ ), implying possible antagonistic dynamics between these two distinct immune populations within the tumor microenvironment (**Fig 5D**). Correlation analysis between biomarkers and differentially infiltrating immune cells was further performed. CPOX expression was found to be significantly positively associated with T helper 2 cells ( $r = 0.54$ ,  $P < 0.001$ ), myocytes ( $r = 0.62$ ,  $P < 0.001$ ), and CD8<sup>+</sup> central memory T cells ( $r = 0.32$ ,  $P < 0.01$ ). In addition, SLC35F5 expression showed a significant positive correlation with myocytes ( $r = 0.33$ ,  $P < 0.01$ ) (**Fig 5E**). Significant negative correlations were observed between both biomarkers and T helper 1 cells as well as mesenchymal stem cells. Specifically, CPOX expression was inversely associated with T helper 1 cells ( $r = -0.50$ ,  $P < 0.001$ ) and mesenchymal stem cells ( $r = -0.58$ ,  $P < 0.001$ ). Similarly, SLC35F5 showed strong negative correlations with T helper 1 cells ( $r = -0.62$ ,  $P < 0.001$ ) and mesenchymal stem cells ( $r = -0.74$ ,  $P < 0.001$ ) (**Fig 5E**). Overall, both CPOX and SLC35F5 are closely associated with the immune microenvironment, suggesting their potential involvement in immune regulatory mechanisms. These insights may inform future investigations into immune-mediated disorders and the dynamics of the immune landscape in obesity.

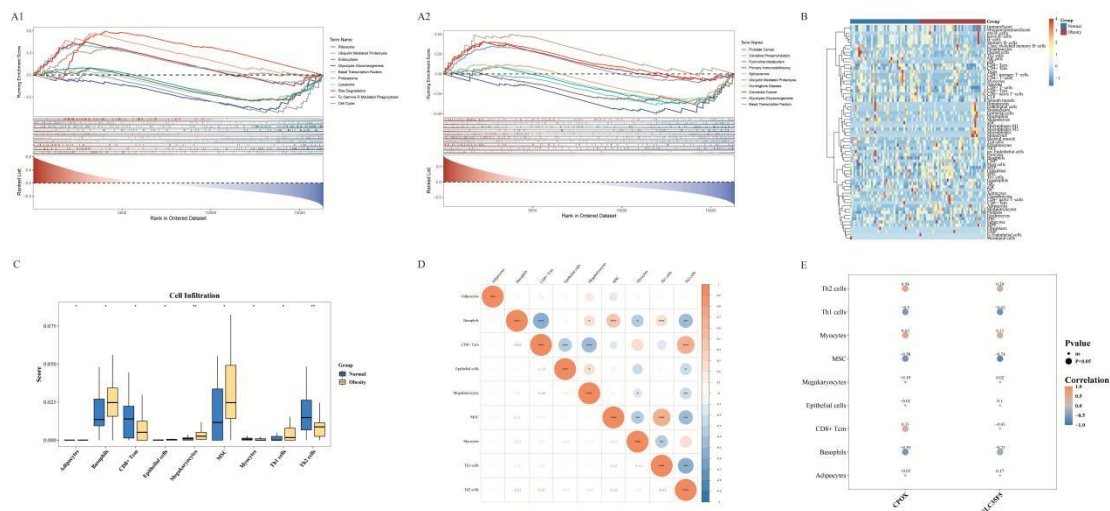


**FIGURE 5 GSEA analysis of CPOX and SLC35F5 and characterization of the**

**immune microenvironment in obesity.** (A1) GSEA visualization for CPOX. (A2) GSEA visualization for SLC35F5. (B) Stacked bar plot of 64 immune cell types. (C) Boxplot of differentially infiltrating immune cells. (D) Correlation heatmap among differential immune cells. (E) Correlation heatmap between biomarkers and immune cells. KEGG pathway data were used with permission from Kanehisa Laboratories.

### **3.6 Unsupervised clustering identifies distinct Obesity subtypes and functional analysis of subtypes**

To further dissect the heterogeneity and underlying features within the patient cohort, consensus clustering based on cumulative distribution function (CDF) analysis stratified obese patients in the training dataset into two distinct subgroups: C1 and C2 (**Fig 6A**). A comparative analysis was performed between samples from subgroups C1 and C2. Differential gene expression analysis revealed a total of 3,142 DEGs meeting the thresholds of  $|\log_2FC| > 0.25$  and  $P_{\text{adjust}} < 0.05$ , comprising 2,045 transcripts exhibiting upregulation alongside 1,097 transcripts demonstrating significant downregulation (**Supplementary Table S5**). The distribution and statistical significance of the DEGs were visualized using a volcano plot and a heatmap (**Fig 6B**). Genes demonstrating the most substantial transcriptional divergence were subsequently prioritized for closer examination, with the ten highest-ranking transcripts from both the up-regulated and down-regulated categories ordered according to the magnitude of their  $|\log_2FC|$  values being labeled directly within the graphical representation. Compared to subgroup C2, the genes GRAMD1C, SPIN4, SKIL, LIG4, SNX16, ABHD13, SLC25A40, LCORL, RP2, and ADAM9 were significantly highly expressed in subgroup C1. Conversely, the genes GYPE, RHCE, WCE, TAC3, TMEM150C, TUBB2A, SH3GL3, CT45A1, GSTM3, and RNASE3 were significantly down-regulated in subgroup C1. Enrichment analyses utilizing GO and KEGG pathway frameworks were subsequently applied to the identified DEGs. For the biological process (BP) category, DEGs were mainly enriched in Golgi vesicle transport, phospholipid biosynthetic processes, macromolecule methylation, macroautophagy, and centrosome cycle-related pathways. For CC, significant

enrichment was observed in the ubiquitin ligase complex, Golgi apparatus subcompartment, pericentriolar material, trans-Golgi network, and peroxisomal membrane. For MF, significant enrichment was observed in terms including GTPase regulatory activity, nucleoside triphosphatase regulatory activity, magnesium ion binding, ubiquitin-like protein transferase activity, and ubiquitin-like protein binding (**Fig 6C, Supplementary Table S6**). KEGG analysis showed significant enrichment in 4 main pathways: Autophagy - animal, Mineral absorption, Longevity regulating pathway, and TNF signaling pathway (**Fig 6D**).

GSEA was conducted to elucidate the biological pathway landscapes distinguishing molecular subtypes C1 and C2. The results indicated that both subtypes were significantly enriched in multiple pathways, including Asthma, NOD-like receptor signaling, Purine metabolism, RNA polymerase activity, Toll-like receptor signaling, Fatty acid metabolism, Sphingolipid metabolism, Oxidative phosphorylation, Proximal tubule bicarbonate reclamation, and Primary immunodeficiency (**Fig 6E, Supplementary Table S7**). Finally, to elucidate discrete biological mechanisms and delineate prospective therapeutic targets across the identified subgroups, pathway enrichment analysis was conducted. The results demonstrated that 5 canonical pathways exhibited the most pronounced differential activity and highest enrichment scores when comparing cluster C1 and cluster C2: kegg-dorso-ventral-axis-formation, kegg-basal-transcription-factors, kegg-pantothenate-and-coa-biosynthesis, kegg-sphingolipid-metabolism, and kegg-B-cell-receptor-signaling-pathway. Conversely, the top 5 pathways with the lowest scores and the largest differences in C1 compared to C2 were: kegg-parkinsons-disease, kegg-proximal-tubule-bicarbonate-reclamation, kegg-oxidative-phosphorylation, kegg-proteasome, and kegg-RNA-polymerase (**Fig 6F, Supplementary Table S8**).

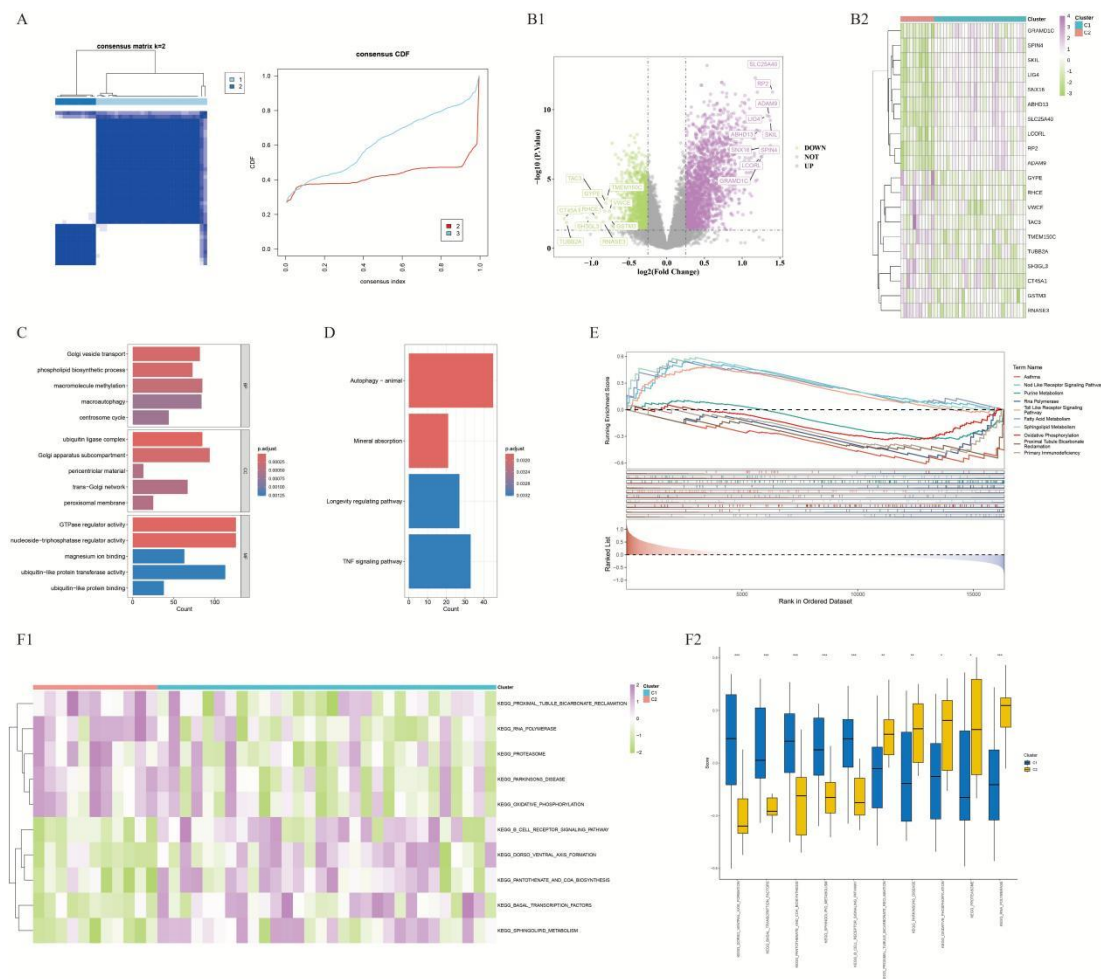


**FIGURE 6 Immune and Molecular Mechanisms in Obesity and Depression.** (A) Immune Cell Infiltration in Obesity. (B1) Differential Gene Expression Analysis (Obesity vs Control). (B2) Differential Gene Expression Analysis (Anxiety-Depression vs Control). (C) Pathway Enrichment of Key Biomarkers. (D) Correlation between Biomarkers and Immune Cells. (E) Gene Set Enrichment Analysis (GSEA) for Obesity. (F1) Immune Signaling Pathways in Obesity Subtypes. (F2) Subtype-Specific Pathway Activation. KEGG pathway data were used with permission from Kanehisa Laboratories.

### 3.7 Subtype drug prediction and molecular docking analysis of CPOX and SLC35F5 with drugs

To evaluate potential therapeutic drugs for obesity-related subtypes (C1 and C2), CMap analysis was performed. The top 5 small molecule compounds with the lowest CMap scores were selected. The results showed that the top 5 potential drugs for

subtype C1 were mazindol, acridinium, moclobemide, nilutamide, and venlafaxine. For subtype C2, the top 5 potential drugs were FK-866, celecoxib, triacin C, nifedipine, and pyrazinamide (**Fig 7A**).

To further assess the binding affinity between the biomarkers CPOX and SLC35F5 and the small molecule drugs, molecular docking analysis was performed. The results revealed that, among the candidate drugs for subtype C1, mazindol displayed the highest binding affinity to CPOX (-13.1 kcal/mol) (**Table 2**). Among the potential drugs for subtype C2, FK-866 showed the strongest binding affinity to CPOX (-12.6 kcal/mol) (**Table 3**). Subsequently, the interactions between these 2 drug-target pairs were visualized using molecular docking diagrams. The results revealed that both mazindol and FK-866 formed multiple hydrogen bond interactions with residues of CPOX (**Fig 7B**), suggesting that mazindol may be a potential drug for treating obesity subtype C1, while FK-866 may be a potential drug for treating type C2 obesity.

**Table 2 Minimum binding energy of molecular docking between biomarkers and potential drugs for subtype C1 patients.**

|             | CPOX  | SLC35F5 |
|-------------|-------|---------|
| Mazindol    | -13.1 | -9.3    |
| Acridinium  | -8.8  | -9.3    |
| Moclobemide | -10   | -7.7    |
| Nilutamide  | -12.4 | -9.2    |
| Venlafaxine | -9.6  | -7.7    |

**Table 3 Minimum binding energy of molecular docking between biomarkers and potential drugs for subtype C2 patients.**

|           | CPOX  | SLC35F5 |
|-----------|-------|---------|
| FK-866    | -12.6 | -10     |
| Celecoxib | -12.5 | -10.1   |

|              |       |      |
|--------------|-------|------|
| Triacsin-c   | -8.1  | -6   |
| Nicardipine  | -11.8 | -9.7 |
| Pyrazinamide | -7    | -5.4 |

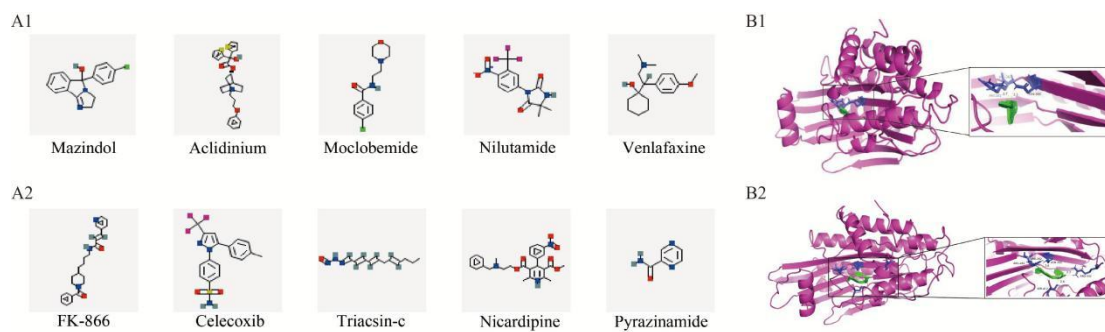


**FIGURE 7 Molecular Docking and Drug Prediction for Obesity Subtypes.** (A1) Drug Prediction for C1 Subtype. (A2) Drug Prediction for C2 Subtype. (B1) Molecular Docking of Mazindol and CPOX. (B2) Molecular Docking of FK-866 and CPOX.

### 3.8 Expression of biomarkers in clinical samples

RT-qPCR was employed to validate the differential expression patterns identified through bioinformatics analysis. The results revealed a statistically significant downregulation of both CPOX and SLC35F5 at the transcriptional level in obese subjects relative to their non-obese counterparts ( $P < 0.05$ , **Fig 8A-B**). These experimental observations align with and corroborate the computational predictions derived from the preceding bioinformatics pipeline, thereby reinforcing the biological relevance of these candidate genes in the pathophysiological context of obesity.

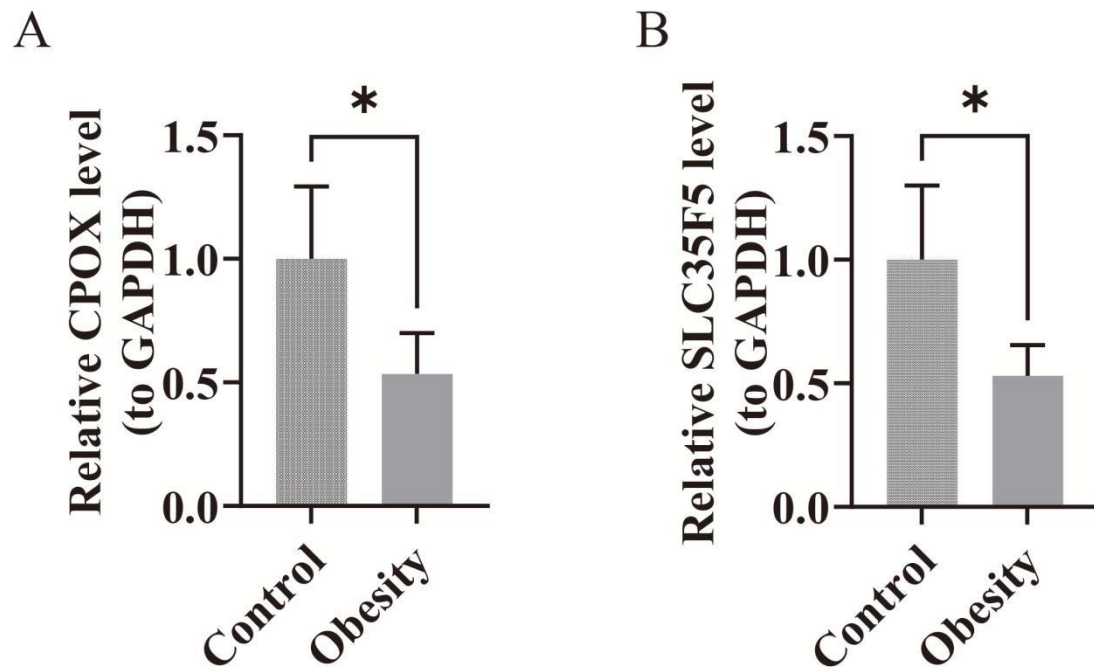


**FIGURE 8 Expression and Validation of Biomarkers in Clinical Samples.** (A) Gene Expression of CPOX and SLC35F5 in Obesity. (B) Gene Expression Validation by RT-qPCR.

### 4. Discussion

Obesity is a chronic metabolic condition marked by an abnormal accumulation of body fat[29]. Previous studies have shown that overweight or obese individuals frequently exhibit symptoms of anxiety or depression [30]. However, systematic research on the potential impact of depression and anxiety disorders in obesity remains lacking. Leveraging transcriptomic profiling in conjunction with an ensemble

of machine learning approaches, the present investigation systematically delineated a set of hub genes implicated in the pathological overlap between obesity and affective conditions, specifically depressive and anxiety-related disorders. Further validation of expression levels yielded two biomarkers (CPOX and SLC35F5). A nomogram developed using these two biomarkers exhibited excellent discriminatory power and strong calibration. The study also analyzed the biological pathways associated with the biomarkers and conducted functional analyses of different obesity subtypes, revealing the molecular associations between obesity and depression/anxiety disorders, offering new theoretical insights and a basis for early intervention and treatment strategies for obesity.

Coproporphyrinogen oxidase (CPOX) is the sixth-step enzyme in heme synthesis, located in the mitochondrial intermembrane space, catalyzing the oxidative decarboxylation of coproporphyrinogen III to protoporphyrinogen IX[31]. Germline pathogenic variants in CPOX lead to hereditary coproporphyria (HCP), which belongs to the spectrum of acute hepatic porphyrias[32]. Studies have found that adipose tissue possesses a transcriptional program for heme synthesis, suggesting its potential association with lipid metabolism[33]. In this study, CPOX was significantly downregulated in obesity samples, indicating reduced heme synthesis, which may lead to impaired mitochondrial fission and fatty acid oxidation, thereby exacerbating energy metabolism disorders and oxidative stress[34,35]. On the other hand, insufficient heme and abnormal accumulation of intermediate metabolites can interfere with the synthesis and function of monoamine neurotransmitters (such as dopamine and serotonin) [36], increasing an individual's susceptibility to anxiety and depression, a mechanism similar to the psychiatric symptoms commonly observed in patients with acute hepatic porphyria[37]. Therefore, the downregulation of CPOX in obesity not only reflects metabolic abnormalities but may also link obesity to the occurrence of depression and anxiety through inflammatory and neurotransmitter pathways, suggesting its potential as a biomarker for the comorbidity of obesity and psychiatric disorders.

Solute Carrier Family 35 Member F5 (SLC35F5) belongs to the solute carrier family

(SLC superfamily), specifically, it is a member of the SLC35 subfamily, class F. Proteins in the SLC35 family are generally localized to the intracellular membrane system (such as the Golgi apparatus and endoplasmic reticulum), primarily involved in the transmembrane transport of nucleotides or related substrates, playing roles in glycosylation modification, cellular metabolism, and signal transduction[38–40]. In obesity samples, SLC35F5 expression is significantly downregulated, suggesting its potential involvement in obesity-related metabolic reprogramming and endoplasmic reticulum homeostasis imbalance. Insufficient nucleotide supply or impaired transport can lead to compromised glycosylation processes, affecting the function of membrane receptors, transporters, and secretory proteins, thereby exacerbating insulin resistance and chronic inflammation[41,42]. Abnormal glycosylation, on the other hand, may disrupt neurotransmitter receptors and synaptic plasticity, thereby affecting signaling pathways such as dopamine and serotonin. This alteration is associated with an increased susceptibility to depression and anxiety disorders[41]. In summary, the significant downregulation of SLC35F5 not only reflects metabolic and inflammatory abnormalities in obesity but also suggests its potential as a molecular marker linking obesity with depression/anxiety, providing new insights for understanding metabolic-psychiatric comorbidities. Subsequent PCR analysis confirmed the downregulation of CPOX and SLC35F5 expression in the obesity group, consistent with the bioinformatics findings and further supporting the reliability of our results.

The GSEA analysis revealed that CPOX and SLC35F5 were co-enriched in pathways such as kegg-ubiquitin-mediated-proteolysis, kegg-huntingtons-disease, kegg-glycolysis-gluconeogenesis, kegg-proteasome, and kegg-toll-like-receptor-signaling-pathway. Among these, the Toll-like receptor (TLR) signaling pathway serves as a key hub connecting metabolic disorders and mood disorders. Obesity is accompanied by intestinal endotoxemia (LPS) and elevated levels of various endogenous DAMPs, which preferentially activate TLR4/TLR2. This activation triggers the MyD88/TRIF→NF-κB/IRF3 pathway, leading to the release of inflammatory factors such as TNF-α, IL-6, and IL-1β, driving M1 macrophage infiltration in adipose tissue, impairing insulin signaling, and inducing

systemic low-grade inflammation[43,44]. Peripheral inflammation affects the central nervous system through increased blood-brain barrier permeability, vagal nerve reflexes, and the entry of cytokines/chemokines[45]. This activates microglial TLR4 ( $\pm$  TLR2/3)-NLRP3 inflammasome and neuroinflammation[46], impacting the HPA axis, tryptophan-kynurenine (IDO/TDO), and neurotransmitter pathways such as dopamine/5-HT, manifesting as depression/anxiety-like behaviors and anhedonia[47,48]. These results indicate that CPOX and SLC35F5 may modulate the development of obesity-related depression and anxiety via this pathway.

Additional analysis of immune cell infiltration differences between obese and control samples revealed 9 immune cell types with significant variations, including adipocytes, basophils, CD8<sup>+</sup> central memory T cells, and epithelial cells, among others. Among these, adipocytes are not only energy reservoirs but also active endocrine and immune cells[49]. During obesity, adipocyte hypertrophy and mild hypoxia trigger inflammation and insulin resistance, shifting their secretory profile from "metabolically favorable" to "pro-inflammatory." [50]. The lipotoxicity and metabolic stress caused by visceral fat and ectopic fat (in the liver/muscles) exacerbate emotional symptoms[51]. Additionally, immune infiltration results showed an imbalance of Th1/Th2 cells and an increase in mesenchymal stem cells (MSCs) in the obese group, consistent with chronic low-grade inflammation[52]. The pro-inflammatory Th1 phenotype (e.g., IFN- $\gamma$  secretion) may promote adipose tissue fibrosis by activating macrophages[53], while the protective role of Th2 cells (IL-4/IL-13) is weakened[54], collectively leading to metabolic dysregulation. A positive correlation was observed between CPOX and Th2, suggesting that CPOX may influence Th2 differentiation and metabolic programming through porphyrin/heme metabolism, thereby regulating the inflammatory cascade[55,56]. The positive correlation between SLC35F5 and myocyte infiltration may reflect abnormal adipose-muscle metabolic interactions[57,58]. In summary, these results indicate that the adipocyte-immune-neural axis could serve as a potential target for intervention in obesity.

GSVA analysis showed that subtype C1 was primarily enriched in pathways like the

KEGG B-cell receptor signaling pathway. This pathway regulates B cell activation, antibody production, and cytokine release through the CD79A/B-SYK-BTK-BLNK-PLCG2-PI3K/AKT-NF- $\kappa$ B/MAPK signaling cascade [59]. In obesity, B cells in adipose tissue increase and shift toward a pro-inflammatory phenotype, producing IgG/IL-6/TNF- $\alpha$ , etc. Coupled with elevated BAFF levels and impaired Breg (IL-10) function, this promotes low-grade inflammation and insulin resistance[59,60]. Peripheral B cell activation and antibodies/cytokines affect the central nervous system (e.g., microglial activation, HPA axis, and monoamine pathway regulation) via the gut-immune-brain and inflammation-brain axes[61,62], suggesting an association with depression/anxiety phenotypes. Meanwhile, BCR-TLR co-signaling and mucosal IgA-microbiota alterations further couple metabolic and emotional pathways[63]. The C2 subtype was found to be enriched in pathways such as KEGG-Oxidative-phosphorylation, suggesting mitochondrial dysfunction and impaired protein degradation, both of which are strongly associated with insulin resistance.

Molecular docking showed that among the potential drugs for C1 subtype patients, Mazindol had the strongest binding affinity to CPOX (-13.1 kcal/mol), and among the potential drugs for C2 subtype patients, FK-866 had the strongest binding affinity to CPOX. Mazindol is a sympathomimetic appetite suppressant that primarily reduces appetite and body weight by inhibiting norepinephrine/dopamine transporters (possibly with weak 5-HT effects) [64], suggesting it may alleviate anxiety and reduce appetite by modulating neurotransmitters.

This study identified biomarkers associated with obesity, depression, and anxiety disorders, namely CPOX and SLC35F5. Pathways related to these biomarkers include glycolysis/gluconeogenesis, ubiquitin-mediated proteolysis, Toll-like receptor signaling pathway. These results offer important insights for further investigation into depression and anxiety disorders in relation to obesity. However, the study's sample size was limited, and the biomarkers identified need further extensive clinical validation. The clinical application of bioinformatics analysis results necessitates data support from larger sample sizes. Additionally, it indicates that we will continue to

monitor the roles of these mechanisms.

### **Author contributions**

Xiaochen Ren and Ziqi Zhao contributed equally to this work and share first authorship. For the two authors with the same name and affiliation, Bin Wang (eighth author) refers to the first-listed Bin Wang, and Bin Wang (ninth and corresponding author) refers to the second-listed Bin Wang.

Conceptualization, X.R., Z.Z., Q.Z. and Bin Wang (ninth and corresponding author); methodology, X.R., Z.Z., N.X. and Bin Wang (ninth and corresponding author); software, Z.Z., Q.Y. and Y.L.; validation, X.R., Z.Z., N.X., Q.Y. and Y.L.; formal analysis, X.R., Z.Z., Q.Y. and Y.L.; investigation, X.R., Z.Z. and N.X.; resources, N.X., Y.M., Q.Z. and Bin Wang (ninth and corresponding author); data curation, X.R., Z.Z., N.X., Q.Y. and Y.L.; writing-original draft preparation, X.R. and Z.Z.; writing-review and editing, X.R., Z.Z., Y.M., Q.Z., Bin Wang (eighth author) and Bin Wang (ninth and corresponding author); visualization, X.R., Z.Z. and Bin Wang (eighth author); supervision, Q.Z. and Bin Wang (ninth and corresponding author); project administration, Q.Z. and Bin Wang (ninth and corresponding author); funding acquisition, Q.Z. and Bin Wang (ninth and corresponding author). All authors have read and agreed to the submitted version of the manuscript.

### **Ethical statement**

This study was granted by the IRB of the First Teaching Hospital of Tianjin University of TCM, with the ethics approval number TYLL2025[Z] Clinical Ethics Review No. 056.

### **Funding**

This work was funded by the Research Topics on Chronic Disease Management (2025) organized by the National Health Commission's Capacity Building and Continuing Education Center (No. GWJJMB202510024008), the National Construction Unit of Traditional Chinese Medicine Dominant Specialties (No. 90 [2024] of the Medical Administration Department, National Administration of Traditional Chinese Medicine), and the Science & Technology Development Fund of Tianjin Education Commission for Higher Education (No. 2022ZD043).

## Competing interests

The authors declare no competing interests.

## Data Availability

The public transcriptome datasets analysed in this study are available in the Gene Expression Omnibus (GEO) database under accession numbers GSE109597, GSE69039 and GSE98793 (<https://www.ncbi.nlm.nih.gov/geo/>). Additional publicly available resources used in this study include the Human Protein Atlas (<https://www.proteinatlas.org/>), MSigDB (<https://www.gsea-msigdb.org/gsea/msigdb>), CMap (<https://clue.io/>), PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) and the Protein Data Bank (<https://www.rcsb.org/>). The de-identified RT-qPCR data generated during the current study are not publicly available because of ethical and privacy considerations related to human participants, but are available from the corresponding author upon reasonable request. The source data supporting the figures and tables are provided in the Supplementary Information files. The source data supporting the supplementary tables are provided in Supplementary Data.

## References

1. Moreno-Fresneda, P. *et al.* Lower frequency of anti-citrullinated protein antibodies among early arthritis patients with high body mass index. *Clin Exp Rheumatol* **38**, 1155–1160 (2020).
2. Perdomo, C. M., Avilés-Olmos, I., Dicker, D. & Frühbeck, G. Towards an adiposity-related disease framework for the diagnosis and management of obesities. *Rev Endocr Metab Disord* **24**, 795–807 (2023).
3. King, S. E. & Skinner, M. K. Epigenetic Transgenerational Inheritance of Obesity Susceptibility. *Trends Endocrinol Metab* **31**, 478–494 (2020).
4. Pledger, S. L. & Ahmadizar, F. Gene-environment interactions and the effect on obesity risk in low and middle-income countries: a scoping review. *Front*

*Endocrinol (Lausanne)* **14**, 1230445 (2023).

5. Liu, Z. *et al.* A comprehensive approach to lifestyle intervention based on a calorie-restricted diet ameliorates liver fat in overweight/obese patients with NAFLD: a multicenter randomized controlled trial in China. *Nutr J* **23**, 64 (2024).
6. Cena, H., Chiovato, L. & Nappi, R. E. Obesity, Polycystic Ovary Syndrome, and Infertility: A New Avenue for GLP-1 Receptor Agonists. *J Clin Endocrinol Metab* **105**, e2695-2709 (2020).
7. Monroe, S. M. & Harkness, K. L. Major Depression and Its Recurrences: Life Course Matters. *Annu Rev Clin Psychol* **18**, 329–357 (2022).
8. Penninx, B. W., Pine, D. S., Holmes, E. A. & Reif, A. Anxiety disorders. *Lancet* **397**, 914–927 (2021).
9. Chellappa, S. L. & Aeschbach, D. Sleep and anxiety: From mechanisms to interventions. *Sleep Med Rev* **61**, 101583 (2022).
10. Milano, W. *et al.* Depression and obesity: Analysis of common biomarkers. *Diseases* **8**, 23 (2020).
11. Fulton, S., Décarie-Spain, L., Fioramonti, X., Guiard, B. & Nakajima, S. The menace of obesity to depression and anxiety prevalence. *Trends Endocrinol Metab* **33**, 18–35 (2022).
12. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
13. Wang, J. *et al.* PROS1 shapes the immune-suppressive tumor microenvironment and predicts poor prognosis in glioma. *Front. Immunol.* **13**, 1052692 (2022).

14. Zhang, X. *et al.* Single-cell RNA and transcriptome sequencing profiles identify immune-associated key genes in the development of diabetic kidney disease. *Front. Immunol.* **14**, 1030198 (2023).
15. Zheng, Y. *et al.* Ferroptosis and autophagy-related genes in the pathogenesis of ischemic cardiomyopathy. *Front. Cardiovasc. Med.* **9**, 906753 (2022).
16. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).
17. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res* **51**, D587–D592 (2023).
18. Lo, H.-J., Tsai, C.-H. & Huang, T.-W. Apoptosis-associated genetic mechanisms in the transition from rheumatoid arthritis to osteoporosis: A bioinformatics and functional analysis approach. *APL Bioeng* **8**, 046107 (2024).
19. Engebretsen, S. & Bohlin, J. Statistical predictions with glmnet. *Clin. Epigenet.* **11**, 123 (2019).
20. Van Essen, D. C. Cortical cartography and caret software. *Neuroimage* **62**, 757–764 (2012).
21. Chen, M. *et al.* Application research of visible near-infrared spectroscopy technology for detecting intracerebral hematoma. *World Neurosurg* **180**, e422–e428 (2023).
22. Robin, X. *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).

23. Li, J. PsyCh journal. Editorial. *Psych J* **4**, 1–2 (2015).
24. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
25. Aran, D., Hu, Z. & Butte, A. J. xCell: Digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).
26. Liu, E.-H. *et al.* Association between serum vitamin E and bacterial vaginitis in women: A cross-sectional study. *BMC Womens Health* **24**, 316 (2024).
27. Hu, X., Ni, S., Zhao, K., Qian, J. & Duan, Y. Bioinformatics-led discovery of osteoarthritis biomarkers and inflammatory infiltrates. *Front. Immunol.* **13**, 871008 (2022).
28. Ji, L. *et al.* Identification of bioactive compounds and potential mechanisms of scutellariae radix-coptidis rhizoma in the treatment of atherosclerosis by integrating network pharmacology and experimental validation. *Biomed Pharmacother* **165**, 115210 (2023).
29. Zou, Y. & Pitchumoni, C. S. Obesity, obesities and gastrointestinal cancers. *Dis Mon* **69**, 101592 (2023).
30. Dakanalis, A. *et al.* The association of emotional eating with overweight/obesity, depression, anxiety/stress, and dietary patterns: A review of the current clinical evidence. *Nutrients* **15**, 1173 (2023).
31. Honda, M. *et al.* Coproporphyrinogen oxidase deficiency causes primary adrenal insufficiency and 46,XY DSD. *J Clin Endocrinol Metab* **111**, e83–e91 (2025).
32. Wang, B. & Bissell, D. M. Hereditary coproporphyria. in (eds Adam, M. P. *et al.*)

(University of Washington, Seattle, Seattle (WA), 1993).

33. Hamada, S. *et al.* Five-aminolevulinic acid (5-ALA) induces heme oxygenase-1 and ameliorates palmitic acid-induced endoplasmic reticulum stress in renal tubules. *Int J Mol Sci* **24**, 10151 (2023).
34. Chanséaume, E. & Morio, B. Potential mechanisms of muscle mitochondrial dysfunction in aging and obesity and cellular consequences. *Int J Mol Sci* **10**, 306–324 (2009).
35. Li, H., Ren, J., Li, Y., Wu, Q. & Wei, J. Oxidative stress: The nexus of obesity and cognitive dysfunction in diabetes. *Front Endocrinol (Lausanne)* **14**, 1134025 (2023).
36. Soladogun, A. S. & Zhang, L. The neural palette of heme: Altered heme homeostasis underlies defective neurotransmission, increased oxidative stress, and disease pathogenesis. *Antioxidants (Basel)* **13**, 1441 (2024).
37. Dickey, A. K., Leaf, R. K. & Balwani, M. Update on the porphyrias. *Annu. Rev. Med.* **75**, 321–335 (2024).
38. Maszczak-Seneczko, D., Wiktor, M., Skurska, E., Wiertelak, W. & Olczak, M. Delivery of nucleotide sugars to the mammalian golgi: A very well (un)explained story. *Int J Mol Sci* **23**, 8648 (2022).
39. Song, Z. Roles of the nucleotide sugar transporters (SLC35 family) in health and disease. *Mol. Aspects Med.* **34**, 590–600 (2013).
40. Van den Bossche, F. *et al.* Residence of the nucleotide sugar transporter family members SLC35F1 and SLC35F6 in the endosomal/lysosomal pathway. *Int J Mol*

- Sci* **25**, 6718 (2024).
41. Chen, J. *et al.* Hepatic glycogenesis antagonizes lipogenesis by blocking S1P via UDPG. *Science* **383**, eadi3332 (2024).
  42. Liu, K., Jin, X., Zhang, X., Lian, H. & Ye, J. The mechanisms of nucleotide actions in insulin resistance. *J Genet Genomics* **49**, 299–307 (2022).
  43. Clemente-Postigo, M. *et al.* Metabolic endotoxemia promotes adipose dysfunction and inflammation in human obesity. *Am J Physiol Endocrinol Metab* **316**, E319–E332 (2019).
  44. Mohammad, S. & Thiemermann, C. Role of metabolic endotoxemia in systemic inflammation and potential interventions. *Front. Immunol.* **11**, 594150 (2020).
  45. Huang, X., Hussain, B. & Chang, J. Peripheral inflammation and blood-brain barrier disruption: Effects and mechanisms. *CNS Neurosci Ther* **27**, 36–47 (2021).
  46. Alfonso-Loeches, S., Ureña-Peralta, J., Morillo-Bargues, M. J., Gómez-Pinedo, U. & Guerri, C. Ethanol-induced TLR4/NLRP3 neuroinflammatory response in microglial cells promotes leukocyte infiltration across the BBB. *Neurochem. Res.* **41**, 193–209 (2016).
  47. Correia, A. S. & Vale, N. Tryptophan metabolism in depression: A narrative review with a focus on serotonin and kynurenine pathways. *Int J Mol Sci* **23**, 8493 (2022).
  48. Mingoti, M. E. D., Bertollo, A. G., de Oliveira, T. & Ignácio, Z. M. Stress and kynurenine-inflammation pathway in major depressive disorder. *Adv. Exp. Med. Biol.* **1411**, 163–190 (2023).

49. Amatya, S. *et al.* Adipocyte glucocorticoid receptor inhibits immune regulatory genes to maintain immune cell homeostasis in adipose tissue. *Endocrinology* **164**, bqad143 (2023).
50. Uchiyama, T., Ota, H., Ohbayashi, C. & Takasawa, S. Effects of intermittent hypoxia on cytokine expression involved in insulin resistance. *Int J Mol Sci* **22**, 12898 (2021).
51. Fourman, L. T. & Grinspoon, S. K. Approach to the patient with lipodystrophy. *J Clin Endocrinol Metab* **107**, 1714–1726 (2022).
52. Valentine, Y. & Nikolajczyk, B. S. T cells in obesity-associated inflammation: The devil is in the details. *Immunol. Rev.* **324**, 25–41 (2024).
53. Friedline, R. H. *et al.* IFN $\gamma$ -IL12 axis regulates intercellular crosstalk in metabolic dysfunction-associated steatotic liver disease. *Nat. Commun.* **15**, 5506 (2024).
54. Hartung, F. & Esser-von Bieren, J. Trained immunity in type 2 immune responses. *Mucosal Immunol.* **15**, 1158–1169 (2022).
55. Belot, A., Puy, H., Hamza, I. & Bonkovsky, H. L. Update on heme biosynthesis, tissue-specific regulation, heme transport, relation to iron metabolism and cellular energy. *Liver Int* **44**, 2235–2250 (2024).
56. Wilfahrt, D. & Delgoffe, G. M. Metabolic waypoints during T cell differentiation. *Nat. Immunol.* **25**, 206–217 (2024).
57. Fang, P. *et al.* Adipose-muscle crosstalk in age-related metabolic disorders: The emerging roles of adipo-myokines. *Ageing Res. Rev.* **84**, 101829 (2023).
58. Kubrak, O. *et al.* LGR signaling mediates muscle-adipose tissue crosstalk and

- protects against diet-induced insulin resistance. *Nat. Commun.* **15**, 6126 (2024).
59. Schmid, V. K. & Hobeika, E. B cell receptor signaling and associated pathways in the pathogenesis of chronic lymphocytic leukemia. *Front. Oncol.* **14**, 1339620 (2024).
  60. Wei, L., Zhang, L., Zhao, D., Ma, Y. & Yang, L. The characteristic activity of regulatory B cells during the occurrence and development of insulin resistance. *Endocrine* **87**, 562–568 (2025).
  61. El Aidy, S., Dinan, T. G. & Cryan, J. F. Immune modulation of the brain-gut-microbe axis. *Front. Microbiol.* **5**, 146 (2014).
  62. Westfall, S. *et al.* Chronic stress-induced depression and anxiety priming modulated by gut-brain-axis immunity. *Front. Immunol.* **12**, 670500 (2021).
  63. Rivera, C. E. *et al.* Intrinsic B cell TLR-BCR linked coengagement induces class-switched, hypermutated, neutralizing antibody responses in absence of T cells. *Sci. Adv.* **9**, eade8928 (2023).
  64. Tanaka, Y., Maeda, N., Koseki, M. & Maeda, K. Changes in body weight in severely obese patients treated with the anorexiant mazindol. *J Clin Med* **13**, 1860 (2024).

## Supplementary Files

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

- [SupplementaryData.xlsx](#)
- [SupplementaryInformation.pdf](#)