


RESEARCH

Open Access



Identification of metabolism related biomarkers in obesity based on adipose bioinformatics and machine learning

Yanping Wang^{1,2†} , Honglin Wang^{3†}, Xingrui Yu⁴, Qinan Wu², Xinlu Lv², Xuelian Zhou⁵, Yong Chen^{5*} and Shan Geng^{1,2,6*}

Abstract

Background Obesity has emerged as a growing global public health concern over recent decades. Obesity prevalence exhibits substantial global variation, ranging from less than 5% in regions like China, Japan, and Africa to rates exceeding 75% in urban areas of Samoa.

Aim To examine the involvement of metabolism-related genes.

Methods Gene expression datasets GSE110729 and GSE205668 were accessed from the GEO database. DEGs between obese and lean groups were identified through DESeq2. Metabolism-related genes and pathways were detected using enrichment analysis, WGCNA, Random Forest, and XGBoost. The identified signature genes were validated by real-time quantitative PCR (qRT-PCR) in mouse models.

Results A total of 389 genes exhibiting differential expression were discovered, showing significant enrichment in metabolic pathways, particularly in the propanoate metabolism pathway. The orange4 module, which exhibited the highest correlation with propanoate metabolism, was identified using Weighted Correlation Network Analysis (WGCNA). By integrating the DEGs, WGCNA results, and machine learning methods, the identification of two metabolism-related genes, Storkhead Box 1 (STOX1), NACHT and WD repeat domain-containing protein 2 (NWD2) was achieved. These signature genes successfully distinguished between obese and lean individuals. qRT-PCR analysis confirmed the downregulation of STOX1 and NWD2 in mouse models of obesity.

Conclusion This study has analyzed the available GEO dataset in order to identify novel factors associated with obesity metabolism and found that STOX1 and NWD2 may serve as diagnostic biomarkers.

Keywords Bioinformatics analysis, Machine learning, Metabolism, Differentially expressed genes (DEGs, Biomarkers)

[†]Yanping Wang and Honglin Wang have contributed equally to this work.

*Correspondence:

Yong Chen

chenyongdz@21cn.com

Shan Geng

gengshan@cqmu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

In recent decades, economic development has led to obesity becoming a major global public health issue. Obesity prevalence exhibits significant global variation, ranging from less than 5% in regions like China, Japan, and Africa to rates exceeding 75% in urban areas of Samoa [1]. Whether in developing or developed countries, obesity can affect the overall health of the population. Moreover, obesity is closely associated with a variety of non-communicable diseases, such as cardiovascular diseases, diabetes, musculoskeletal disorders, and specific types of cancer [2]. Obesity is a multifactorial condition, with increased risk tied to environmental, socioeconomic, and demographic factors. Furthermore, links have been found between being overweight or obese and various factors, including age, gender, socioeconomic status, and whether one resides in urban or rural areas [3].

Obesity classification commonly depends on the body mass index (BMI) calculation. This formula requires dividing an individual's weight (in kilograms) by their height squared (in meters), resulting in a value measured in kg/m^2 [4]. Based on a systematic review of prevalence data, individuals with a $\text{BMI} \geq 30 \text{ kg}/\text{m}^2$ are commonly considered obese. Estimates within this classification indicate that the occurrence of metabolically healthy obesity (MHO) spans from 10 to 51% [5]. Furthermore, research conducted by Phillips and colleagues demonstrated that the prevalence of metabolic health in obese individuals was between 6.8 and 36.6%, while the prevalence of metabolically unhealthy subjects in non-obese individuals ranged from 21.8 to 87% [6]. Additionally, obesity is a major factor contributing to hypertension, with metabolic abnormalities closely associated with the severity of this condition and the risk of target organ damage. Disruptions in body composition and the presence of visceral obesity significantly influence metabolic risk factors. By rearranging sentence structures, using synonyms, and adding slight variations, the core message remains intact while reducing the risk of plagiarism [7]. Obesity is closely linked with various cardiac metabolic disorders such as type 2 diabetes (T2D), dyslipidemia, hypertension, and coronary artery disease. However, the underlying mechanisms differentiating obesity from these common cardiac metabolic complications remain incompletely understood. In addition to current research, genome-wide association studies can reveal new genetic factors and pathways [8, 9].

The main goal of this research is to clarify how genes related to metabolism are involved in obesity-linked metabolic diseases. Furthermore, we seek to evaluate their expression in cellular models.

Materials and methods

Data collection

The GSE110729 and GSE205668 datasets were derived from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). GSE110729 expression profile data included 28 patients with 15 in the lean group and 13 in the obese group and served as a training data set. Meanwhile, the expression profile data of GSE205668 included 61 patients, with 35 in the lean group and 26 in the obese group, and served as the validation dataset. The GSE205668 dataset is a bulk RNA-Seq analysis of adipose samples collected during routine surgeries as part of the Leipzig Childhood Study in Germany, while GSE110729 is a bulk RNA-Seq study of adult subjects in the United States.

Differential analysis

The "DESeq2" package in R was utilized to perform differential analysis on the GSE110729 dataset. Genes with differential expression (DEGs) were identified using screening criteria of $\text{padj} \leq 0.05$ and $|\log\text{FC}| > 1$.

GO and KEGG pathway enrichment analyses

Gene Ontology (GO) enrichment analysis is a widely used bioinformatics technique for extracting in-depth insights from extensive genomic datasets, such as Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). Additionally, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, another common approach for understanding biological processes and functions, was executed. For the KEGG enrichment analysis, the "clusterProfiler" package was employed, whereas the "Metascape" database was used for the GO enrichment analysis.

WGCNA

Weighted Correlation Network Analysis (WGCNA) is a genomics research method that facilitates the discovery of gene clusters with high relatedness. This is accomplished by constructing a coexpression network using the WGCNA-R package, with a focus on the top 5000 genes exhibiting the highest variance. This network facilitates web-based gene screening to identify potential biomarkers or therapeutic targets. Modules of genes are then identified through hierarchical clustering, and gene expression patterns are used to build a weighted gene network. Genes are categorized based on their expression patterns, grouping those with similar patterns into modules. This process divides tens of thousands of genes into multiple modules based on their expression patterns,

utilizing the correlation and correlation coefficient as key measures.

GSVA

We utilized Gene Set Variation Analysis (GSVA) to evaluate the activity of biological pathways in our gene expression dataset. GSVA is an unsupervised, non-parametric method that calculates pathway scores for each sample. This approach enables a comprehensive and data-driven exploration of changes at the pathway level, facilitating the discovery of biologically significant insights within our dataset.

At first, the gene sets related to metabolism were obtained from single-sample gene set enrichment analysis (GSEA) | MSigDB (gsea-msigdb.org). Following this, the GSVA package in R was utilized, focusing on the ssGSEA technique for computing metabolic scores. Additionally, the limma package was employed for conducting differential analysis to pinpoint significant KEGG pathways. Metabolism points for major KEGG pathways were used as patient-specific inputs. The WGCNA network was then constructed using mRNA expression data to identify module genes most involved in metabolism, and the specific molecular mechanism was further studied.

Identify metabolism-associated signature genes of obesity by machine learning

eXtreme Gradient Boosting (XGBoost) and Random Forest (RF) were applied to identify signature genes linked to metabolism in obesity. XGBoost, short for Extreme Gradient Boosting, is a powerful and versatile machine learning algorithm widely employed in various fields due to its exceptional predictive performance. Originally introduced as an ensemble learning technique, XGBoost builds upon decision trees and aims to minimize prediction errors by iteratively adding weak learners to the model, thus boosting its overall accuracy. This algorithm has proven particularly effective in addressing challenges posed by high-dimensional datasets and complex relationships among variables. Our experiment involved the utilization of the XGBoost algorithm through the "caret" package in R software for both feature selection and classification purposes. The Random Forest (RF) algorithm functions as an ensemble method that merges various decision trees to produce a single decision based on the combined outcomes of different classifiers. Each tree in the forest is constructed utilizing the bootstrap technique, which involves selecting diverse samples from the original dataset and training them with a randomly selected function using the bagging mechanism. Subsequently, decisions made by numerous individual trees are aggregated through a voting process, with the class

receiving the most votes being designated as the prediction. In this instance, we applied the RF algorithm to forecast RF using the cellular senescence-associated signature genes within the "randomForest" package in R software.

In conclusion, The machine learning methods mentioned above identified specific genes that were deemed to be characteristic of obesity in terms of metabolism, specifically known as the signature genes of obesity. The expression levels of these genes were assessed in both the training set (GSE110729) and the testing set (GSE205668). To assess the predictive precision of these characteristic genes, we generated an ROC (Receiver Operating Characteristic) curve utilizing the "pROC" package within the R software.

The curve showing the ROC demonstrates the correlation between sensitivity, which is the true positive rate, and specificity, calculated as 1 minus the false positive rate. The X-axis, representing 1-specificity (false positive rate), approaches zero as accuracy increases. On the other hand, the Y-axis, denoting sensitivity (true positive rate), exhibits higher accuracy as it increases.

Animals

Thirteen-week-old male C57BL/6J (WT), ob/ob, and DB/DB mice were sourced from Shulaibao (Wuhan) Biotechnology Co., Ltd. These mice were housed in cages with regulated temperatures and a 12-h light–dark cycle, with unrestricted access to water and a standard diet (chow diet, CD) for a duration of 4 weeks. In a separate treatment, six-week-old male C57BL/6J mice were fed a high-fat diet (HFD, D12492, Research Diet, New Brunswick, NJ, USA) for a period of 12 weeks. Every mouse participating in this research maintained a C57BL/6J genetic background. Euthanasia was performed via intraperitoneal administration of ketamine (100 mg/kg) and xylazine (10 mg/kg), followed by cervical dislocation. Subsequent to euthanasia, white adipose tissue was harvested for further experimental processing and analysis. The Animal Experimental Ethics Committee of Chongqing Medical University approved all the experimental procedures, which were conducted in compliance with applicable guidelines and regulations.

Real-time quantitative PCR (qRT-PCR)

Total RNA extraction was accomplished utilizing the RNeasy mini kit (Qiagen, Germany) according to the given protocols, after which cDNA synthesis was conducted using Qiagen's quantitative reverse transcription kit pursuant to the manufacturer's guidelines. For the quantitative analysis of target genes, the FastStart Universal SYBR Green Master kit was employed, and DNA amplification was executed with the LightCycler

480 system. To determine the relative abundance of target gene mRNA, the delta-delta Ct ($\Delta\Delta Ct$) method was applied, utilizing an internal control for comparison purposes. The PCR was processed under the following conditions: an initial denaturation at 95 °C for 5 min, succeeded by 40 cycles of denaturation at 95 °C for 10 s, annealing at 60 °C for 60 s, and a final extension at 72 °C for 30 s.

Data and statistical analysis

For statistical processing, GraphPad Prism 6.01 (GraphPad Software Inc., San Diego, CA, USA) was used. Data is represented as mean \pm standard error of mean (mean \pm SEM). Comparative analysis of quantitative values was performed using a two-factor analysis of variance. Tukey's honest significant difference (HSD) test was implemented for subsequent pairwise evaluations [10].

Results

Identification of DEGs related to obesity

Data from the GSE110729 dataset was retrieved from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/gds/>). Principal Component Analysis (PCA) was utilized to assess dataset variance, employing the "FactoMineR" package. Visualization of the findings was carried out using the "ggplot2" package (Fig. 1C). Differentially expressed genes (DEGs) were pinpointed by categorizing the 28 samples into 13 obese and 15 lean samples. DEGs analysis was performed using the "DESeq2" package, with screening criteria set as $\text{padj} < 0.05$, $|\log FC| > 1$, and p-values arranged in ascending order. A total of 40 genes, comprising 20 up-regulated genes and 20 down-regulated genes, were identified as the most significant. Hierarchical clustering analysis was performed using all DEGs, revealing distinct expression differences between the two groups, as depicted in

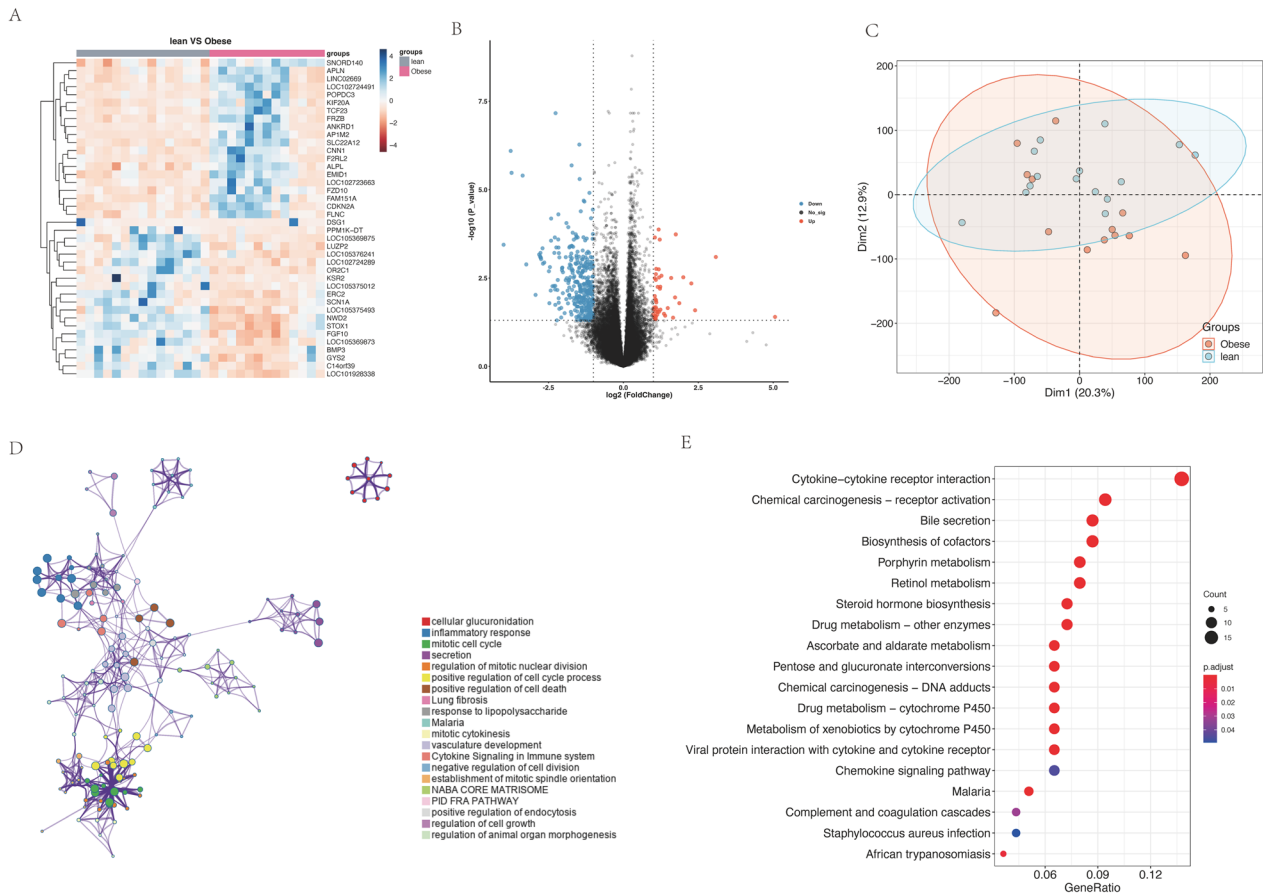


Fig. 1 Identification results of differentially expressed genes (DEGs) and functional enrichment analysis. **A** Heatmap of 40 DEGs identified using the "DESeq2" package. Samples in the differentiating gene sets (GSE110729) are displayed by columns, while genes are represented by rows. Gray squares indicate lean samples, and red squares indicate obese samples. DEGs: differentially expressed genes. **B** Volcano plot of DEGs, with red color indicating high expression and blue color indicating low expression. **C** Principal component analysis (PCA) was used for quality control, with each point representing a sample. Blue points represent lean samples, and red points represent obese samples. **D** GO enrichment analysis of DEGs performed by Metascape. **E** KEGG enrichment analysis of DEGs performed by clusterProfiler

the heatmap (Fig. 1A). In addition, through visualization using “ggplot2”, it can be found in the volcano plot that the differential genes in the chip are mostly downregulated genes (Fig. 1B).

GO and KEGG enrichment analysis of DEGs

To investigate the biological characteristics of the DEGs, enrichment analysis was performed using the Metascape database [11]. BP suggests that DEGs are enriched in processes such as “cellular glucuronidation,” “uronic acid metabolic process,” and “glucuronate metabolic process.” This observation implies that metabolic reactions might be crucial for obese populations. Regarding CC, the terms “spindle,” “collagen-containing extracellular matrix,” and “condensed chromosome” showed significant enrichment. Hence, we hypothesize that DEGs primarily influence chromosomal activities. Additionally, the MF terms associated with DEGs included “glucuronosyltransferase activity,” “cytokine activity,” and “receptor ligand activity” (Fig. 1D). To investigate the underlying roles of DEGs, an analysis of KEGG pathway enrichment was performed utilizing the software package “clusterProfiler”. The findings indicated that DEGs were notably enriched in the pathways of “Porphyrin metabolism,” “Ascorbate and aldarate metabolism,” and “Interconversions of Pentose and glucuronate” (Fig. 1E).

GSVA reveals the potential function of differential genes

We obtained pathway scores for each sample through GSVA analysis. Thereafter, differential analysis was performed using the limma package, setting the threshold at P value < 0.01 , to identify differential metabolic pathways. The top differential GSVA terms related to metabolism were “propanoate metabolism,” “porphyrin and chlorophyll metabolism” and “galactose metabolism” (Additional file 1: Fig. S1A). The volcano plot clearly demonstrate the downregulation of genes associated with metabolic pathways such as “propanoate metabolism” in obese individuals compared to lean individuals. Conversely, genes involved in “porphyrin and chlorophyll metabolism” and “galactose metabolism” exhibit upregulation in obese individuals when compared to their lean counterparts (Additional file 1: Fig. S1B).

WGCNA analysis of gene expression profiles in obese and lean

WGCNA facilitates the detection of disease-related modules characterized by coordinated expression patterns, thereby significantly enhancing the identification of central genes. To construct a gene co-expression network, clustering analysis was performed using the GSE110729 dataset, involving a total of 28 samples for the construction of a hierarchical clustering tree. The soft threshold power of 24 was selected using the pickSoft-Threshold function available in the WGCNA package for the GSE110729 dataset (Fig. 2A). We utilized the innovative mixed cropping technique to combine modules with high similarity in feature genes, resulting in the discovery of 10 unique gene modules represented by various colors. We employed the plotEigengeneNetworks function from the WGCNA package to visualize and analyze the consensus eigengene networks. The eigengene adjacency heatmap effectively illustrates the correlation structure among modules, uncovering a significant level of interconnectivity among specific eigengenes, such as ME_{maroon} and ME_{mediumpurple4}, as well as ME_{darkorange2} and ME_{brown2} (Fig. 2B). Notably, the gray module consisted of genes that could not be classified (Fig. 2C). The heatmap illustrates that the orangered4 module exhibited the strongest correlation with propanoate metabolism among all three metabolic pathways (Fig. 2D). To evaluate the potential function and mechanism of propanoate metabolism, we picked the orangered4 module that exhibited the highest correlation with propanoate metabolism, which was validated as a hub module. These genes from the hub orangered4 module was analyzed by the “clusterProfiler” package for GO and KEGG enrichment analysis.

GO and KEGG enrichment analysis of metabolism related genes

BP suggests that genes associated with metabolic processes are highly present in “carboxylic acid breakdown,” “organic acid breakdown,” and “cellular amino acid breakdown”. Regarding CC, the phrases “Golgi apparatus cisterna,” “cis-Golgi network,” and “Golgi apparatus cisterna membrane” showed significant enrichment. As for MF, differentially expressed genes related to metabolism exhibited terms such as “vitamin binding,” “ligase

(See figure on next page.)

Fig. 2 Weighted Gene Coexpression Network Analysis (WGCNA) in the GSE110729 cohort. **A** Scale independence and average connectivity in a metaqueue of differentially metabolized genes. **B** Hierarchical clustering dendrogram of module eigengenes with color labels. **C** Gene dendrogram and modules before and after merging in the GSE110729 cohort. **D** Correlation analysis of merged modules with metabolic pathways in the GSE110729 cohort. **E** The intersection of genes related to obesity metabolism obtained from WGCNA with the set of differentially expressed genes (DEGs) in the GSE110729 cohort

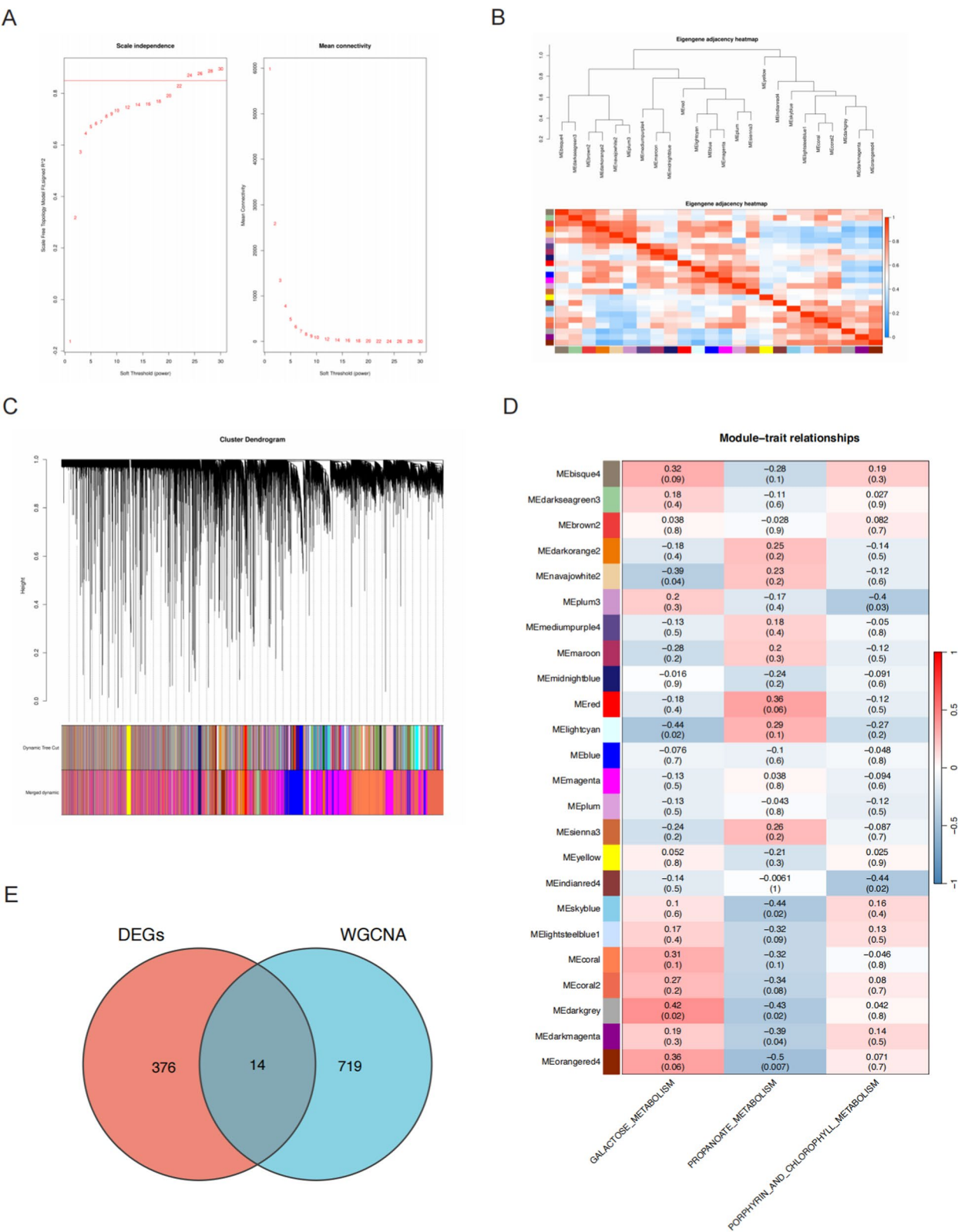


Fig. 2 (See legend on previous page.)

activity", and "pyridoxal phosphate binding" (Additional file 1: Fig. S2A). We utilized the "clusterProfiler" package to perform KEGG pathway enrichment analysis. The analysis identified that the differentially expressed genes were significantly enriched in the following pathways: "Valine, leucine, and isoleucine degradation," "Propanoate metabolism," and "Glyoxylate and dicarboxylate metabolism." (Additional file 1: Fig. S2B).

Construction and evaluation of machine learning models

As depicted in the Venn diagram (Fig. 2E), the metabolism-related module obtained from WGCNA, referred to as orangered4, was intersected with the set of DEGs. Consequently, we identified 14 metabolism-related DEGs. Subsequently, two machine learning models were constructed: Random Forest and XGBoost. For the XGBoost model, gene importance was evaluated using variable importance values. Genes with values greater than or equal to the median were selected (Additional file 1: Fig. S3A). Similarly, the Random Forest model evaluated gene importance using MeanDecreaseGini values, selecting genes with values ≥ 1 (Additional file 1:

Fig. S3B). Consequently, two key genes were identified among the 14 metabolism-related DEGs using the two machine learning models (Additional file 1: Fig. S3C). We evaluated the diagnostic performance of the model we built and the two characteristic genes by analyzing the ROC curve. In the training dataset, the model demonstrated an area under the curve (AUC) of 1 (95% confidence interval 1–1), whereas in the validation dataset, the AUC was 0.799 (95% CI 0.727–1) (Fig. 3A, D). Specifically, for the gene Storkhead Box 1 (STOX1), the AUC was 0.877 (95% CI 0.727–1) in the training set and 0.791 (95% CI 0.666–0.917) in the validation set. For the gene NACHT and WD repeat domain-containing protein 2 (NWD2), the AUC was 0.908 (95%CI 0.779–1) in the training set and 0.803 (95%CI=0.681–0.925) in the validation set (Fig. 3B, E). The expression differences of obesity signature genes between lean and obese individuals are illustrated in Fig. 3C, F. In obese individuals, NWD2 and STOX1 were discovered to be notably reduced in comparison to their lean counterparts ($P < 0.05$).

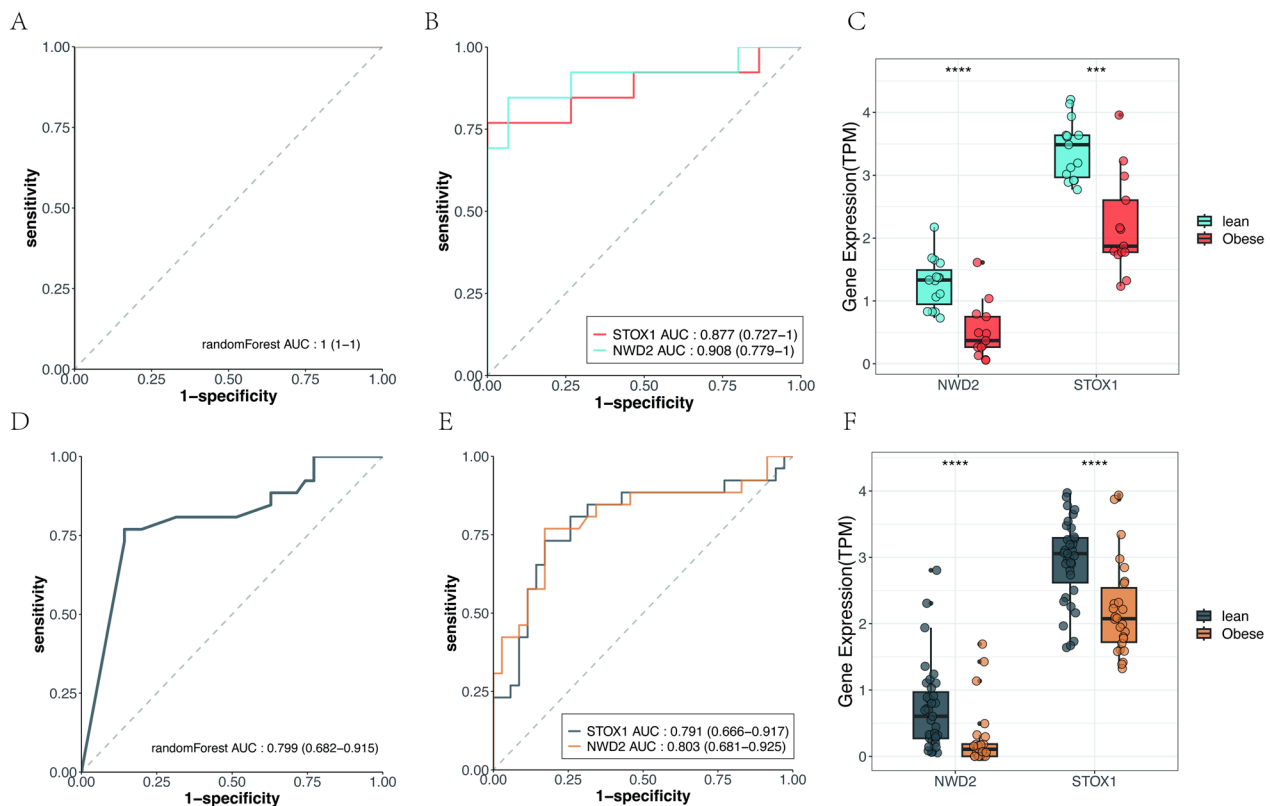


Fig. 3 Expression difference and ROC curve of obesity signature genes in the GSE110729 (training set) and GSE205668 (valid set). The ROC curves were used to evaluate the diagnostic efficacy of the Random Forest model in the GSE110729 (A) and GSE205668 (D). The ROC curves of the two obesity signature genes in the GSE110729 (B) and GSE205668 (E). Expression difference of obesity signature genes among lean and obese in the GSE110729 (C) and GSE205668 (F)

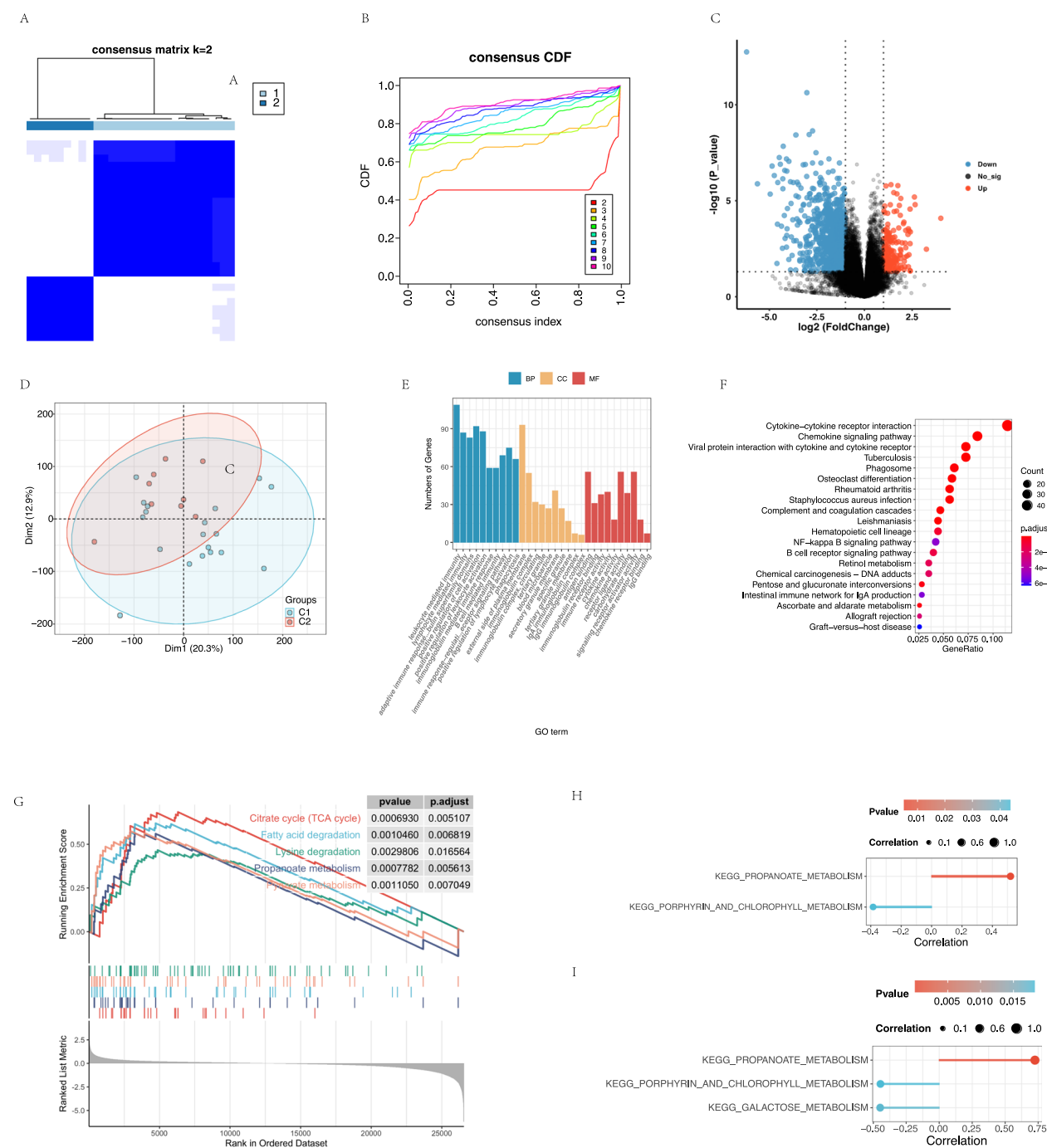


Fig. 4 Identification and expression analysis of obesity metabolism related gene clusters. Enrichment analysis of subtypes including C1 and C2. Correlation analysis of NWD2 and STOX1 with metabolic pathways. **A** Uniform clustering matrix at K=2. **B** Representative CDF curve. **C** PCA between C1 and C2 gene clusters. **D** Expression of C1 and C2 gene clusters. **E** GO analysis. **F** KEGG analysis. **G** GSEA. **H** Correlation analysis of NWD2 with metabolic pathways. **I** Correlation analysis of STOX1 with metabolic pathways

Distinct metabolic subtypes in obesity

To investigate the predictive potential of metabolic genes in 28 obesity and slim individuals, we clustered the data unattended using the R-package "ConsensusClusterplus".

We successfully identified two different subtypes and the cluster stability turned out to be optimal for k=2 (Fig. 4A, B). It is worth noting that there were notable variances in the levels of metabolic-related compounds

expressed among the pair of clusters (Fig. 4C). We further validated our findings using principal component analysis and observed that the expression levels of 28 metabolic-related molecules can clearly distinguish the two clusters (Fig. 4D). Moreover, the outcomes of the GO, KEGG analyses and GSEA demonstrated a strong correlation between the subtypes identified by consensus cluster analysis and the metabolic pathway (Fig. 4E–G). The lollipop plot illustrates a strong correlation between the two key genes, NWD2 and STOX1, and metabolic pathways (Fig. 4H, I).

Expression of STOX1 and NWD2 in mice

To investigate the potential relationship between the STOX1 and NWD2 genes and metabolic processes, we analyzed their expression levels in the white adipose tissue of wild-type (WT), ob/ob, DB/DB, and high-fat diet (HFD) mice. Using qRT-PCR, we discovered a significant differential expression, with STOX1 and NWD2 mRNA levels being lower in obese mice as opposed to lean ones ($p < 0.05$). These results indicate that the STOX1 and NWD2 genes could potentially act as biomarkers for metabolism associated with obesity (Fig. 5).

Discussion

Adipokines are signal molecules secreted by fatty tissue that play a crucial role in various physiological functions, including energy balance, metabolism, inflammation and immune function. Adipokines play a crucial role in maintaining metabolic homeostasis and regulating various

physiological functions. Further research on adipokines and their interactions with different tissues and organs may provide valuable insights into the pathogenesis of obesity and associated diseases, as well as potential therapeutic targets for managing these conditions.

For instance, loci near IRS1 have been found to harbor alleles associated with favorable cardiac metabolic risk characteristics, particularly increased body fat. Several genes related to insulin signaling (ADCY5, CCCDC9, MTOR, RAC1), energy expenditure (IGF2BP2), and inflammation (SH2B3, ADCY9) may serve as therapeutic targets for mitigating cardiac metabolic risks linked with obesity [12].

The current research utilized the DESeq2 software to conduct a comparative analysis of genes, aiming to discover differentially expressed genes (DEGs) within a sample population consisting of both obese and lean subjects. Following this, a comprehensive enrichment analysis was carried out to explore the biological pathways linked to these DEGs. The analysis revealed that the DEGs were primarily enriched in “cellular glucuronization” and “glucuronic acid metabolism processes”. Machine learning techniques were then employed to identify STOX1 and NWD2 genes that are functionally related to metabolism. Through consensus clustering based on the expression of STOX1 and NWD2 genes, two distinct molecular subtypes were identified. The accuracy of the model constructed using the random forest algorithm was evaluated using ROC curves, and the expression of these genes in different cells was

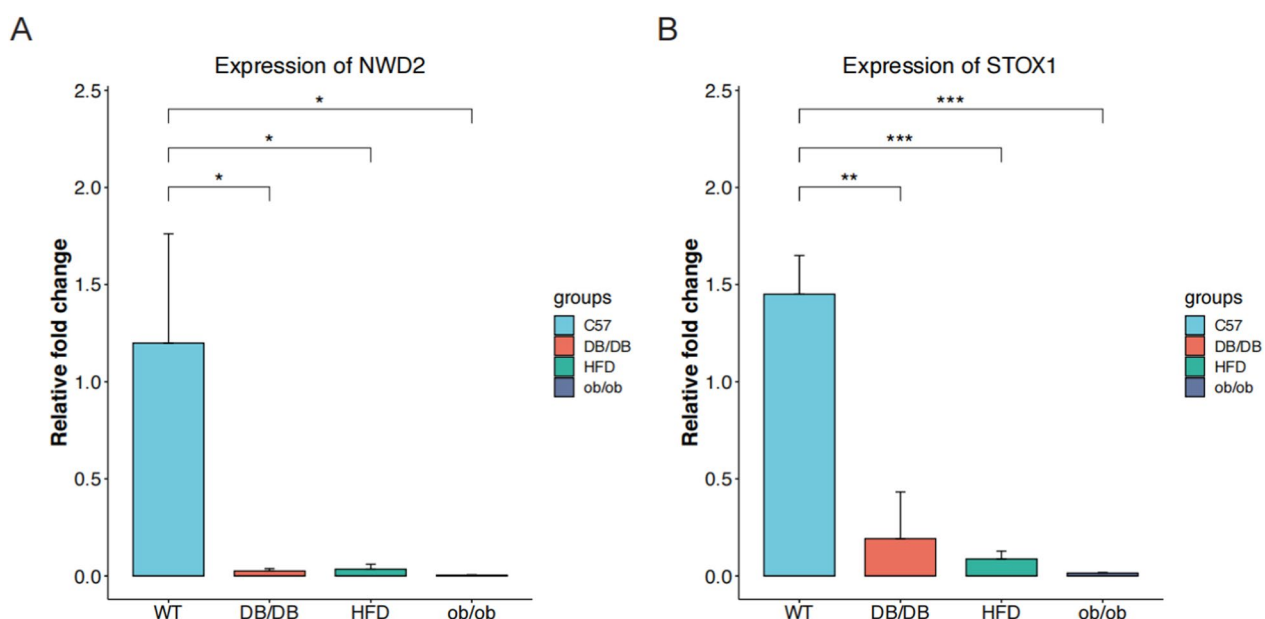


Fig. 5 Expression of STOX1 and NWD2 in white adipose tissue of mice. **A** NWD2 **B** STOX1 groups: c57, DB/DB, HFD, ob/ob

verified using qRT-PCR. Interestingly, the expression level of these genes was found to be relatively low in obese individuals, implying that STOX1 and NWD2 may have potential as valuable biomarkers for obesity-related diseases. Furthermore, the bar plot clearly demonstrates a strong correlation between the two key genes and metabolic pathways. GSEA was performed on the DEGs identified among the subtypes clustered based on the NWD2 and STOX1 genes using consensus clustering, revealing a significant association with metabolic pathways.

STOX1 is an angle encryption factor that shares structures and functional comparisons with disguise transfer factor [13]. Initially STOX1 was described as consisting of six isomers, namely A, B, C, D, E and F, which were identified by selective displacement. Among these isomers STOX1A and STOX1B have been significantly studied. STOX1A represents the most complete isomer possessing a DNA binding domain and an activator domain in which STOX1B shares only the former [14]. Functional investigations of STOX1 have largely focused on its involvement in various biological processes [15], such as the cell cycle [16], early development [17], oxidative stress regulation [18] etc. Moreover, STOX1 also has significant implications in various diseases. The STOX1 gene is responsible for encoding a cytoplasmic protein predominantly associated with fetal development and maternal blood pressure regulation [19]. Yi Xu et al. demonstrated that STOX1 overexpression modulates genes involved in hypoxia, redox balance, carbon monoxide, and energy metabolism, thereby playing a significant role in pulmonary artery remodeling. Additionally, STOX1 has been shown to promote mitotic entry and proliferation of inner ear epithelial cells while inhibiting cerebellar granule neurogenesis and the synthesis of neural tube cell tumors [20].

NWD2 is a nodular-like receptor (NLR) that possesses an N-terminal motif resembling the β -solenoid folding pattern. This repeating unit, known as NWD2, can undergo a structural transformation from its prion protein-forming region to adopt β -solenoid folds, thereby activating Het-s pore formation proteins [21]. NWD2, located on the fifth position of the chromosome, has been shown to possess a unique role in signal transduction, particularly in cholinergic neurons within the habenular nucleus [22]. Further investigations have indicated that oligomeric NWD2 triggers the transformation of HET-s prions and disrupts the nucleating HET-s's aggregation ability through mutations affecting its helix folding mechanism [21]. Additionally, in fungi, the expression of NWD2 redirects HET-s towards the cell's periphery,

activating the pore-forming protein and providing evidence for NWD2's signaling interaction in primitive cells [23]. Nonetheless, the contribution of STOX1 and NWD2 to obesity metabolism remains inadequately investigated.

Our analysis reveals significant associations between the genes STOX1 and NWD2 and the KEGG pathways KEGG_PROANOATE_METABOLISM and KEGG_PORPHYRIN_AND_CHLOROPHYLL_METABOLISM. These findings suggest novel potential connections between these genes and adipose tissue biology, specifically within the context of short-chain fatty acid metabolism and heme processing. The association with propanoate metabolism pathway points to a possible involvement of STOX1 and NWD2 in short-chain fatty acid (SCFA) metabolism. SCFAs, particularly propionate, have been shown to play crucial roles in energy homeostasis and adipose tissue function. They act as signaling molecules, influencing appetite regulation, insulin sensitivity, and inflammation in adipose tissue [24]. The link between STOX1 and NWD2 and propanoate metabolism suggests these genes may modulate SCFA production or utilization in adipose tissue, potentially impacting obesity development.

Concurrently, the association with the porphyrin and chlorophyll metabolism pathway implies a potential role in heme processing. Recent studies have highlighted the importance of heme metabolism in adipose tissue function and insulin sensitivity [25]. Dysregulation of heme homeostasis has been linked to adipose tissue dysfunction and insulin resistance, key factors in obesity pathogenesis [26]. The connection of STOX1 and NWD2 to this pathway suggests they may influence adipose tissue function through modulation of heme metabolism.

Recent studies have established a close association between related genes and obesity metabolism. Our research findings indicate that STOX1 and NWD2 hold significant diagnostic value in predicting obesity-related metabolic diseases. Our analysis of the training set (GSE110729, AUC=1) and validation set (GSE205668, AUC=0.799) demonstrates their potential as excellent diagnostic biomarkers. Therefore, STOX1 and NWD2, due to their close relationship with cellular metabolism, may serve as potential biomarkers for obesity metabolism. Furthermore, we seek to determine the association between these genes and the biological pathways governing the initiation of obesity, as well as their impact on metabolic characteristics.

Although age-related differences in gene expression and metabolic regulation may influence the roles of STOX1 and NWD2 in propanoate and porphyrin

metabolism pathways, both genes are associated with KEGG_PROANOATE_METABOLISM and KEGG_PORPHYRIN_AND_CHLOROPHYLL_METABOLISM across datasets of different ages. Furthermore, these genes exhibit differential expression between obese and normal samples, a finding validated in animal models. The GSE205668 dataset comprises bulk RNASeq analysis of adipose samples from the Leipzig Childhood Study in Germany, while GSE110729 involves deep bulk RNA sequencing of 12 obese and 15 lean adults from the Karolinska Institutet in Sweden. We selected GSE110729 as the training set and GSE205668 for external validation, focusing on their compatibility and data quality. We acknowledge that the ethnic backgrounds of the participants is a crucial factor of diseases like obesity. However, these datasets primarily offer geographical location information and lack detailed data on the ethnicity of the participants. Consequently, our study could not directly address the ethnicity of the cohort. We recognize this as a limitation in our research. Moving forward, we plan to engage in research involving more ethnically diverse cohorts to enhance the universality of our findings. Besides, we acknowledge that the generalizability of these findings to other cohorts may be limited by the demographic and ethnic composition of the datasets we used. While our findings offer valuable insights into the genes associated with obesity, we are planning to conduct further studies using more diverse and larger cohorts. This future research will aim to validate and extend the applicability of these genes across various populations, providing a more comprehensive understanding of their role in obesity. Last but not least, it is essential to note that further examination of a larger sample size is warranted to validate these findings and unravel their underlying mechanisms. It is essential to note that further examination of a larger sample size is warranted to validate these findings and unravel their underlying mechanisms.

Conclusion

This preliminary study aims to examine the involvement of metabolism-related genes, STOX1 and NWD2, in obesity-related metabolic disorders. The findings of this report contribute significantly to the existing knowledge and suggest that STOX1 and NWD2 might serve as promising biomarkers in the obesity population. Future research should employ updated datasets for validation, utilize immunohistochemistry and Western blot techniques to assess protein expression, and investigate the clinical utility of these genes by employing metabolic inhibitors.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-024-05615-8>.

Additional file 1.

Acknowledgements

We acknowledge GEO database for providing platform for uploading their datasets with meaningful information, and thank all animals involved in this study who gave their lives for the advancement of human health. This study was supported by the grants from the Chongqing Dazu District Science and Technology Development project (DZKJ2024JSYJ-KWXM1008), 2023 Municipal Education Commission Science (KJQN202300460), Postdoctoral Science Foundation of China (No. 2020M683263), Natural Science Foundation Project of Chongqing, Chongqing Science and Technology Commission (cstc2021jcyj-msxmX0353).

Author contributions

YW, XY and QW researched and analyzed the data. HW contributed to statistical analysis. XL and XZ wrote and edited the manuscript. YC and SG were the guarantors of this work and, as such, had full access to all the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Data availability

All data generated during the study are included in the article and Additional file 1.

Declarations

Ethics approval and consent to participate

All experimental procedures were approved by the Animal Experimental Ethics Committee of Chongqing Medical University (approval number: 2023017).

Competing interests

No potential conflicts of interest relevant to this article were reported.

Author details

¹Department of Laboratory, The Affiliated Dazu Hospital of Chongqing Medical University, No. 1073 South Erhuan Road, Tangxiang Street, Dazu District, Chongqing 402360, China. ²Department of Endocrinology, The Affiliated Dazu Hospital of Chongqing Medical University, Chongqing 402360, China. ³Department of Orthopedic Surgery, The Affiliated Dazu Hospital of Chongqing Medical University, Chongqing 402360, China. ⁴Institute of Information, Xiamen University, Xiamen, China. ⁵The Affiliated Dazu Hospital of Chongqing Medical University, No. 1073 South Erhuan Road, Tangxiang Street, Dazu District, Chongqing 402360, China. ⁶State Key Laboratory of Ultrasound in Medicine and Engineering, College of Biomedical Engineering, Chongqing Medical University, Chongqing, China.

Received: 10 May 2024 Accepted: 18 August 2024

Published online: 31 October 2024

References

- Shukla A, Kumar K, Singh A. Association between obesity and selected morbidities: a study of BRICS countries. *PLoS ONE*. 2014;9:e94433. <https://doi.org/10.1371/journal.pone.0094433>.
- Piché ME, Tchernof A, Després JP. Obesity phenotypes, diabetes, and cardiovascular diseases. *Circ Res*. 2020;126:1477–500. <https://doi.org/10.1161/circresaha.120.316101>.
- Puoane T, Steyn K, Bradshaw D, Laubscher R, Fourie J, Lambert V, Mbananga N. Obesity in South Africa: the South African demographic and health survey. *Obes Res*. 2002;10:1038–48. <https://doi.org/10.1038/oby.2002.141>.
- Aung K, Lorenzo C, Hinojosa MA, Haffner SM. Risk of developing diabetes and cardiovascular disease in metabolically unhealthy normal-weight

- and metabolically healthy obese individuals. *J Clin Endocrinol Metab.* 2014;99:462–8. <https://doi.org/10.1210/jc.2013-2832>.
5. Maksimovic MZ, Vlajinac HD, Radak DJ, Marinkovic JM, Jorga JB. Prevalence of the metabolic syndrome in patients with carotid disease according to NHLBI/AHA and IDF criteria: a cross-sectional study. *BMC Cardiovasc Disord.* 2012;12:2. <https://doi.org/10.1186/1471-2261-12-2>.
 6. Phillips CM, Dillon C, Harrington JM, McCarthy VJ, Kearney PM, Fitzgerald AP, Perry IJ. Defining metabolically healthy obesity: role of dietary and lifestyle factors. *PLoS ONE.* 2013;8:e76188. <https://doi.org/10.1371/journal.pone.0076188>.
 7. Litwin M, Kulaga Z. Obesity, metabolic syndrome, and primary hypertension. *Pediatr Nephrol (Berlin, Germany).* 2021;36:825–37. <https://doi.org/10.1007/s00467-020-04579-3>.
 8. Ghaben AL, Scherer PE. Adipogenesis and metabolic health. *Nat Rev Mol Cell Biol.* 2019;20:242–58. <https://doi.org/10.1038/s41580-018-0093-z>.
 9. Smith GI, Mittendorfer B, Klein S. Metabolically healthy obesity: facts and fantasies. *J Clin Investig.* 2019;129:3978–89. <https://doi.org/10.1172/jci129186>.
 10. Tukey JW. Comparing individual means in the analysis of variance. *Biometrics.* 1949;5:99–114.
 11. Yue T, Chen S, Zhu J, Guo S, Huang Z, Wang P, Zuo S, Liu Y. The aging-related risk signature in colorectal cancer. *Aging (Albany NY).* 2021;13:7330–49. <https://doi.org/10.18632/aging.202589>.
 12. Huang LO, Rauch A, Mazzaferro E, Preuss M, Carobbio S, Bayrak CS, Chami N, Wang Z, Schick UM, Yang N, Itan Y, Vidal-Puig A, den Hoed M, Mandrup S, Kilpeläinen TO, Loos RJF. Genome-wide discovery of genetic loci that uncouple excess adiposity from its comorbidities. *Nat Metab.* 2021;3:228–43. <https://doi.org/10.1038/s42255-021-00346-2>.
 13. Parchem JG, Kanasaki K, Lee SB, Kanasaki M, Yang JL, Xu Y, Earl KM, Keuls RA, Gattone VH, Kalluri R. STOX1 deficiency is associated with renin-mediated gestational hypertension and placental defects. *JCI Insight.* 2021. <https://doi.org/10.1172/jci.insight.141588>.
 14. Jin FQ, Jin L, Wang YL. Downregulation of STOX1 is a novel prognostic biomarker for glioma patients. *Open Life Sci.* 2021;16:1164–74. <https://doi.org/10.1515/biol-2021-0119>.
 15. Abel D, Abdul-Hamid O, Dijk M, Oudejans CB. Transcription factor STOX1A promotes mitotic entry by binding to the CCNB1 promoter. *PLoS ONE.* 2012;7:e29769. <https://doi.org/10.1371/journal.pone.0029769>.
 16. Liu Y, Sun Z, Zhu J, Xiao B, Dong J, Li X. LncRNA-TCONS_00034812 in cell proliferation and apoptosis of pulmonary artery smooth muscle cells and its mechanism. *J Cell Physiol.* 2018;233:4801–14. <https://doi.org/10.1002/jcp.26279>.
 17. Jones RL, Stoikos C, Findlay JK, Salamonsen LA. TGF-beta superfamily expression and actions in the endometrium and placenta. *Reproduction (Cambridge, England).* 2006;132:217–32. <https://doi.org/10.1530/rep.1.01076>.
 18. Thulluru HK, Park C, Dufort D, Kleiverda G, Oudejans C, van Dijk M. Maternal Nodal inversely affects NODAL and STOX1 expression in the fetal placenta. *Front Genet.* 2013;4:170. <https://doi.org/10.3389/fgene.2013.00170>.
 19. Berends AL, Bertoli-Avella AM, de Groot CJ, van Duijn CM, Oostra BA, Steegers EA. STOX1 gene in pre-eclampsia and intrauterine growth restriction. *BJOG Int J Obstet Gynaecol.* 2007;114:1163–7. <https://doi.org/10.1111/j.1471-0528.2007.01414.x>.
 20. Nie X, Zhang K, Wang L, Ou G, Zhu H, Gao WQ. Transcription factor STOX1 regulates proliferation of inner ear epithelial cells via the AKT pathway. *Cell Prolif.* 2015;48:209–20. <https://doi.org/10.1111/cpr.12174>.
 21. Daskalov A, Habenstein B, Martinez D, Debets AJ, Sabaté R, Loquet A, Saupe SJ. Signal transduction by a fungal NOD-like receptor based on propagation of a prion amyloid fold. *PLoS Biol.* 2015;13:e1002059. <https://doi.org/10.1371/journal.pbio.1002059>.
 22. Yamada S, Furukawa R, Sakakibara SI. Identification and expression profile of novel STAND gene Nwd2 in the mouse central nervous system. *Gene Expression Patterns GEP.* 2022;46: 119284. <https://doi.org/10.1016/j.gep.2022.119284>.
 23. Cai X, Xu H, Chen ZJ. Prion-like polymerization in immunity and inflammation. *Cold Spring Harb Perspect Biol.* 2017. <https://doi.org/10.1101/cshperspect.a023580>.
 24. Canfora EE, Meex RCR, Venema K, Blaak EE. Gut microbial metabolites in obesity, NAFLD and T2DM. *Nat Rev Endocrinol.* 2019;15:261–73. <https://doi.org/10.1038/s41574-019-0156-z>.
 25. Chambers ES, Preston T, Frost G, Morrison DJ. Role of gut microbiota-generated short-chain fatty acids in metabolic and cardiovascular health. *Curr Nutr Rep.* 2018;7:198–206. <https://doi.org/10.1007/s13668-018-0248-8>.
 26. Obi CD, Bhuiyan T, Dailey HA, Medlock AE. Ferrochelatase: mapping the intersection of iron and porphyrin metabolism in the mitochondria. *Front Cell Dev Biol.* 2022;10:894591. <https://doi.org/10.3389/fcell.2022.894591>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.