



OPEN ACCESS

EDITED BY

Haifeng Hou,
Shandong First Medical University, China

REVIEWED BY

Marijana Vujkovic,
University of Pennsylvania, United States
Xiuqing Guo,
Cedars Sinai Medical Center, United States

*CORRESPONDENCE

Kei Hang Katie Chan
✉ kkhchan@cityu.edu.hk

SPECIALTY SECTION

This article was submitted to
Systems Endocrinology,
a section of the journal
Frontiers in Endocrinology

RECEIVED 21 September 2022

ACCEPTED 09 March 2023

PUBLISHED 22 March 2023

CITATION

Tan Y, He Q and Chan KHK (2023)
Identification of shared genetic
architecture between non-alcoholic
fatty liver disease and type 2 diabetes:
A genome-wide analysis.
Front. Endocrinol. 14:1050049.
doi: 10.3389/fendo.2023.1050049

COPYRIGHT

© 2023 Tan, He and Chan. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of shared genetic architecture between non-alcoholic fatty liver disease and type 2 diabetes: A genome-wide analysis

Yajing Tan¹, Qian He¹ and Kei Hang Katie Chan^{1,2,3*}

¹Department of Biomedical Sciences, City University of Hong Kong, Hong Kong, Hong Kong SAR, China, ²Department of Electrical Engineering, City University of Hong Kong, Hong Kong, Hong Kong SAR, China, ³Department of Epidemiology, Center for Global Cardiometabolic Health, Brown University, Providence, RI, United States

Background: The incidence of complications of non-alcoholic fatty liver disease (NAFLD) and type 2 diabetes (T2D) has been increasing.

Method: In order to identify the shared genetic architecture of the two disease phenotypes of NAFLD and T2D, a European population-based GWAS summary and a cross-trait meta-analysis was used to identify significant shared genes for NAFLD and T2D. The enrichment of shared genes was then determined through the use of functional enrichment analysis to investigate the relationship between genes and phenotypes. Additionally, differential gene expression analysis was performed, significant differentially expressed genes in NAFLD and T2D were identified, genes that overlapped between those that were differentially expressed and cross-trait results were reported, and enrichment analysis was performed on the core genes that had been obtained in this way. Finally, the application of a bidirectional Mendelian randomization (MR) approach determined the causal link between NAFLD and T2D.

Result: A total of 115 genes were discovered to be shared between NAFLD and T2D in the GWAS analysis. The enrichment analysis of these genes showed that some were involved in the processes such as the decomposition and metabolism of lipids, phospholipids, and glycerophospholipids. Additionally, through the use of differential gene expression analysis, 15 core genes were confirmed to be linked to both T2D and NAFLD. They were correlated with carcinoma cells and inflammation. Furthermore, the bidirectional MR identified a positive causal relationship between NAFLD and T2D.

Conclusion: Our study determined the genetic structure shared between NAFLD and T2D, offering a new reference for the genetic pathogenesis and mechanism of NAFLD and T2D comorbidities.

KEYWORDS

non-alcoholic fatty liver disease, type 2 diabetes, GWAS, differential gene expression, shared genetics, mendelian randomization

1 Introduction

Non-alcoholic fatty liver disease (NAFLD) is a clinicopathological syndrome that is characterized by hepatic parenchymal steatosis and fat storage in the absence of a history of binge drinking. In simple terms, NAFLD is usually benign, but if linked with an unhealthy lifestyle, obesity, and other metabolic syndromes, it may develop from simple fat accumulation to non-alcoholic steatohepatitis (NASH), liver fibrosis and cirrhosis, and, in rare cases, liver cancer (1). It is a complex disease that results from the interaction of environmental and genetic factors, so it has multiple pathogenic factors such as insulin resistance (IR), lipid metabolism disorders, oxidative stress, and cytokine effects (2). In addition to being considered a manifestation of IR in the liver, NAFLD often coexists with metabolic syndromes such as obesity, type 2 diabetes (T2D), and hyperlipidemia (3). Among them, T2D is a typical endocrine and metabolic disease that is affected by multiple pathogenic factors, and its incidence is increasing rapidly worldwide. T2D, the most prevalent type of diabetes, is characterized by hyperglycemia, IR, and lipid metabolism disorder as the pathological basis (4). In recent years, it has been found that NAFLD and T2D show similar pathological characteristics and often coexist as common diseases that seriously endanger public health. On one hand, T2D can lead to dysfunction of glycolipid metabolism in the body through the development of factors such as IR, chronic inflammation and oxidative stress, which results in NAFLD and further liver damage and worsens the prognosis of NAFLD (5). On the other hand, through fat deposition, inflammation, endoplasmic reticulum stress, and oxidative stress, NAFLD can also exacerbate hepatic IR and promote metabolic abnormalities including hyperglycemia, creating the ideal environment for the development of T2D (6).

Related studies have shown that NAFLD and T2D interact with each other, and that there is a complex two-way relationship between the two that can accelerate deterioration. Targher et al. (7) found that NAFLD was ubiquitous in patients with T2D (7). Similarly, Jarvis et al. (8), in a meta-analysis of population-based cohort studies, found that the occurrence of T2D was associated with a more than two-fold increase in the risk of severe liver disease events among those at risk of or diagnosed with NAFLD (8). This finding was the same as that found by Mantovani et al. (9) in a study of the impact of NAFLD on the risk of development of T2D (9). Also, Pinero et al. (10) reported that the global incidence of NASH had reached 3% to 5%. NASH occurs in 20% to 30% of patients with T2D and obesity, and NAFLD occurs in 69% to 87% of those patients (10). Hence, the incidence of NAFLD combined with T2D is higher than that of NAFLD alone or T2D alone (5). Thus, the identification of the shared genetic architecture of NAFLD and T2D has important implications for the prevention and treatment of these diseases.

In recent years, the development of next-generation sequencing and high-throughput genotyping arrays has led to the GWAS and exome-wide association studies (EWAS), which are methods for the identification of genetic factors for many complex diseases (11). The

use of GWAS, which is larger-scale than EWAS, has led to the identification of many polymorphisms and genetic variants that are associated with NAFLD and T2D and the investigation of new therapeutic targets. In addition, differentially expressed genes (DEGs) are key to learn about gene activity. They have now become one of the most important tools for the discovery of biomarkers (12). This method can be used to find genes that exhibit notable variations in expression, to analyze statistically the findings to pinpoint particular genes that are associated with those conditions, and then to analyze the biological importance of those particular genes. More importantly, DEGs can complement the knowledge of important target tissues and cell types that the GWAS approach lacks in disease pathogenesis, thus realizing the transformation of relevant gene loci into mechanisms. Hence it can be seen that the integration of GWAS summary statistics and gene expression data can identify disease-related tissues and cell types without bias, increase the credibility of the analysis results, and provide a sufficient basis to explain the pathogenesis.

Current genetic studies that target NAFLD and T2D require the discovery of more significant genetic association signals to support and translate the research through various novel analytical methods into biological and potentially therapeutic knowledge. Therefore, this study first conducted a comprehensive genetic analysis through the use of GWAS to identify susceptibility genes for NAFLD combined with T2D. The core shared genes were further screened as this information was combined with the results of DEG analysis. Subsequently, functional annotation analysis was performed to identify the underlying biological pathways of these core, shared genes. At the same time, a two-sample MR analysis was conducted to explore the causal relationship between NAFLD and T2D. The above analysis provided a robust theoretical basis for the study of NAFLD and T2D complications and new ideas and opportunities for the further development of prevention and treatment strategies.

2 Methods

2.1 Data summary

To identify genetic variants in NAFLD combined with T2D, the GWAS summary statistics for this study were obtained from the US National Human Genome Research Institute GWAS catalog (<https://www.ebi.ac.uk/gwas/>), including NAFLD (1,483 cases and 17,781 controls) and T2D (4,040 cases and 113,735 controls). (13, 14). For the DEG analysis, the organism was *Homo sapiens*, and the experiment type was expression profiling by array, which set the screening conditions of the dataset. The related gene expression datasets GSE48452, GSE25724, GSE17470, and GSE20966 were downloaded from the Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/>), which included human liver biopsy (18 NASH, 14 NAF, 27 obese, 14 controls), human islets, and NASH liver biopsy (6 case and 4 controls) and beta-cells from pancreatic tissue (n=10) (15–18). All data sources can be found in Table 1.

TABLE 1 Data source information summary.

Phenotype	Data Source	Population	Cases' size	Controls' size	Total' size	Number (SNPs)	PubMed ID	Download Link
NAFLD	GWAS Catalog	European	1,483	17,781	19,264	6,797,908	32298765	https://www.ebi.ac.uk/gwas/studies/GCST90011885
T2D	GWAS Catalog	European	4,040	113,735	117,775	8,404,432	26961502	https://www.ebi.ac.uk/gwas/studies/GCST006801
NAFLD (Discovery set)	GEO database (GSE48452, GPL11532)	Germany	32	41	73	–	23931760	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi
NAFLD (Validation set)	GEO database (GSE17470, GPL2895)	USA	7	4	11	–	20221393	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi
T2D (Discovery set)	GEO database (GSE25724, GPL96)	Italy	6	7	13	–	21127054	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi
T2D (Validation set)	GEO database (GSE20966, GPL1352)	USA	10	10	20	–	20644627	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi

2.2 Study-level quality control

The “GWASInspector” R package was used to conduct harmonized quality control (QC) on the GWAS statistics of NAFLD and T2D phenotypes to ensure that false positive signals were eliminated and that low-quality data did not obscure actual signals (19, 20). NAFLD and T2D GWAS summary data were performed separately in the QC. The reference dataset was the 1000 genomes project reference panel (21), the specific genome build version was GRCh37, and we included the relevant information from the European population. The QC involved: the deletion of variants that contained missing fundamental values or were duplicated; the deletion of monomorphic variants; the checking of the consistency of allele frequencies with reference datasets; the alignment of alleles with reference datasets, and the comparison of those alleles to ensure that the resulting allele frequencies were correct; the removal of unverifiable mismatches and multi-allelic variants, and; the setting of a threshold $\text{plot_cut-off}_p = 0.01$ to exclude low-significance SNPs.

2.3 Cross-trait meta-analysis

Cross-trait meta-analysis was performed using the CPASSOC software package. CPASSOC is a method for studying cross-phenotypic (CP) associations by using summary statistics from GWAS of multiple phenotypes. It combines effect estimates and standard errors of GWAS summary statistics to test the hypothesis of an association between SNPs and traits (22). Cross-phenotype associations increase statistical power when the traits analyzed share common variants or common genetic pathways, which are often associated with pleiotropy (23). CPASSOC includes two tests, SHom and SHet. In this study, R v.4.1.3 was used to perform the SHet test considering the effect of trait heterogeneity, which can increase

the power when the genetic effect size of different traits is different (24). At this time, the gamma distribution parameters are estimated by setting $N = 1E4$ and calling the EstimateGamma function. Due to the many hypothesis tests that may be applied in GWAS studies, the threshold is strictly controlled to minimize the number of false positives reported. Currently, the most significant threshold is generally recognized as $p < 5 \times 10^{-8}$, and this threshold also applies to CPASSOC (24, 25). Then, hg19 was used as the reference genome, and the refGene database was used to annotate the SNPs that reached the threshold of significance level using the ANNOVAR software (<http://www.openbioinformatics.org/annovar/>). Finally, the shared genes of NAFLD combined with T2D were obtained.

2.4 Enrichment analysis

In this study, SNPs that showed significant variation in meta-analysis and the genes from which they came were used for functional enrichment analysis to explore the potential biological function of shared susceptibility genes between NAFLD and T2D. The online tool Metaspace (<https://metaspace.org/gp/index.html#/main/step1>) was used to analyze comprehensively these susceptibility genes. Metaspace integrates more than 40 gene function annotation databases such as Gene Ontology (GO) and DisGeNET, providing multiple functional and diversified visualization methods such as gene enrichment analysis and protein interaction network analysis, which can be used for easy exploration and analysis of gene function (26). The GO enrichment analysis of candidate genes was focused on the use of the “clusterProfiler” package of R v.4.1.3 (<https://www.r-project.org/>). In addition, the online platform TissueEnrich (<https://tissueenrich.gdcb.iastate.edu/>) was used as a calculating input-gene centralized organization-specific enrichment tool to complete the tissue-specific expression analysis (27).

2.5 Differential gene expression analysis and enrichment

To confirm which of the chosen genes were the core shared genes in NAFLD and T2D, four GEO datasets were selected for DEG analysis. The specific dataset information is shown in Table 1. Of the four, GSE48452 and GSE25724 were used as discovery sets, while GSE17470 and GSE20966 were used as validation sets to verify the validity and disease association of the identified genes. In addition, each dataset was divided into two groups of samples, with NAFLD patients or T2D patients as the experimental group and healthy people as the control group. GEO data were processed through the application of the online analysis tool GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>) to identify DEGs. The visualization of overlapping genes in GEO datasets is realized through the online platform jvenn (<http://www.bioinformatics.com.cn/static/others/jvenn/example.html>) (28). After the discovery and validation sets were merged, the final DEGs were screened through use of *adjust.P.Value*, which is applied to adjust the *p*-value for multiple tests to control the false discovery rate (FDR). The FDR is calculated as expected rate \times (false positive/(false positive + true positive)) (29). The value of *adj.p* was set at <0.05 to screen out the DEGs of NAFLD and T2D. Then, the genes that overlapped with those found through the GWAS were identified as the shared core genes of NAFLD and T2D. Differential gene expression analysis is to identify shared genes through differential expression analysis between two phenotypes and to find overlapping genes with cross-trait analysis based on the GWAS summary statistics. Using multiple analytical methods to explore the reproducibility of our results between the two phenotypes can make the findings more reliable and robust. Next, the enrichment analysis described in section 2.4 was carried out for the genes that were found to overlap in the GWAS and DEG analyses.

2.6 Mendelian randomization analysis

The QC-processed GWAS data were used in the MR analysis. The potential causal effect between T2D and NAFLD was explored through the use of a bidirectional MR analysis, in which the two traits were evaluated alternately as exposure and outcome, and independent SNPs that were closely related to exposure and outcome traits were used as instrumental variables. Among them, the screening of exposure was essential. The parameters $p=5 \times 10^{-8}$ specified the *p*-value of the SNP in the exposure; that is, only SNPs with *p*-values of $<5 \times 10^{-8}$ were extracted (30). The NbDistribution simulation calculation was set to 1000 and the *p*-value threshold for judging whether the SNP was an outlier was set to 0.05 before the MR analysis was performed. Then, the calculation of MR pleiotropy residual sum and outlier (MR-PRESSO) was performed to identify the existence of the outliers (31). Once outliers were located, they were eliminated, and subsequent MR analysis was performed. MR and sensitivity analyses were performed through the use of the inverse variance weighted (IVW) method (32) with multiplicative random effects, supplemented by MR Egger (33, 34), weighted median (33), simple mode, and weighted mode methods (35). It is important to note that horizontal pleiotropy is a potential

confounding factor in MR analysis; i.e., instrumental variable SNPs influence the outcome through a non-causal pathway, which may affect the measurement of the relationship between traits. To examine the impact of pleiotropy on the results of the MR analysis, MR-PRESSO was also used to test for horizontal pleiotropy for multiple instrumental variables. In addition, heterogeneity statistics and leave-one-out analyses were included in the MR analysis. Heterogeneity statistics mainly test the differences between individual SNPs, and leave-one-out analysis mainly tests the stability of MR results. The “TwoSampleMR” and “MRPRESSO” packages were used for MR analysis in R v.4.1.1.3.

3 Results

3.1 Study-level QC

QC was performed through the use of “GWASInspector”. 100% of NAFLD GWAS summary data (6,797,908 SNPs) and 99.7% of T2D GWAS summary data (8,380,746 SNPs) passed the QC procedure (Table 2). SNPs that passed the QC were included in the subsequent cross-trait meta-analysis and MR analysis.

3.2 Cross-trait meta-analysis

In total, CPASSOC identified 241 SNPs that were significantly associated ($p < 5 \times 10^{-8}$) between NAFLD and T2D (Supplementary Table 1). In the results of the cross-trait meta-analysis, the SNP with the most significant *p*-value is rs73233361 ($p = 6.78 \times 10^{-11}$), which is located on chromosome 12. Most of the remaining SNPs are located on chromosome 6, chromosome 2, and chromosome 3. 115 genes were obtained by ANNOVAR annotation (Supplementary Table 1).

3.3 Enrichment analysis

The DisGeNET enrichment analysis revealed that 115 shared genes were enriched in physical activity measurement, substance-related disorders, lean body mass, smoking behaviors, substance abuse problem, etc. (Figure 1A). The relevant results that were identified based on DisGeNET enrichment analysis are listed in Supplementary Table 2. GO enrichment analysis (Supplementary Table 3; Supplementary Figure 1) showed that the shared genes of NAFLD and T2D were enriched in the biological processes of

TABLE 2 The number of SNPs after QC processing.

	NAFLD	T2D
Input variant count	6,797,908	8,404,432
Missing crucial variable	0	2
Duplicated variants	0	12,367
Monomorphic variants	0	0
Output variant count	6,797,908	8,380,746

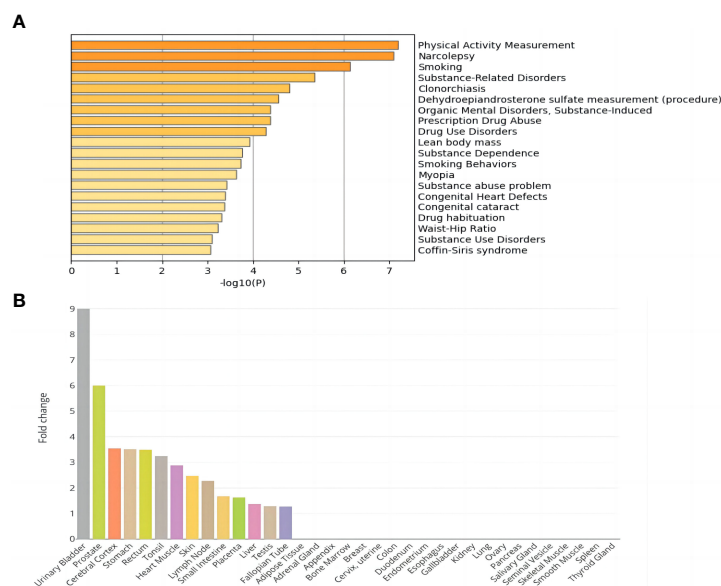


FIGURE 1
Enrichment analysis of shared genes in NAFLD and T2D. (A) DisGeNET enrichment analysis results. (B) Tissue enrichment results.

processes of glycerophospholipids, phospholipids, lipid decomposition, glycerolipid metabolism, and they also participated in the enrichment in the molecular function of activity of various enzymes. Among them, the most were associated with sensory system development, a total of 7 genes (*FASLG*, *ISL1*, *TULP1*, *TBC1D32*, *ATP8A2*, *MAX*, and *ADAMTS18*). In addition, tissue enrichment analysis showed that NAFLD and T2D shared genes were mainly enriched in 14 tissues: the urinary bladder, prostate, cerebral cortex, stomach, rectum, tonsil, heart muscle, skin, lymph node, small intestine, placenta, liver, testis, and fallopian tube (Figure 1B). Among these tissues, the liver is closely related to the pathogenesis of NAFLD and T2D.

3.4 Differential gene expression analysis and enrichment

The DEG analysis results for each dataset are shown in Supplementary Figures 2–5. The discovery sets GSE48452 and GSE25724 contained 11,795 shared genes for NAFLD and T2D (Supplementary Figures 6A, B), whereas the validation sets GSE17470 and GSE20966 contained 15,557 shared genes for NAFLD and T2D (Supplementary Figures 6C, D). A combination of all discovery and validation sets yielded 9711 DEGs (Supplementary Figure 6E). Subsequently, 5545 DEGs shared by NAFLD and T2D were screened by adj.P (Supplementary Table 4). Consideration of these genes with the candidate genes that had been obtained through the GWAS produced fifteen core genes that were shared by NAFLD and T2D, namely *DNAJB9*, *VPS53*, *SCGN*, *CMAS*, *RGS6*, *FASLG*, *ABHD10*, *ATRN*, *PLA2G2F*, *ITIH2*, *ROBO1*, *SGCG*, *SH3GL2*, *CNR1*, and *FOXN3* (Table 3).

DEG analysis of the above core genes (Supplementary Table 5) showed that seven genes were upregulated ($\log\text{FoldChange}>0$) and eight genes were downregulated ($\log\text{FoldChange}<0$) in disease. These fifteen core genes were subjected to enrichment analysis, and DisGeNET enrichment analysis revealed that they were enriched in carcinoma cells and inflammation (Figure 2A). Relevant findings from the DisGeNET enrichment analysis are provided in Supplementary Table 6.

GO enrichment analysis showed significant enrichment of several biological processes including regulation of endopeptidase and peptidase activity, lipid catabolic process, fatty acid metabolic process, response to lipopolysaccharide, and positive regulation of proteolysis; cellular components of the distal axon, endoplasmic reticulum lumen, and glutamatergic synapse; and molecular functions such as carboxylic ester hydrolase activity (Supplementary Table 7; Supplementary Figure 7). In addition, tissue enrichment analysis showed that NAFLD and T2D core shared genes were enriched in the urinary bladder, stomach, rectum, tonsil, heart muscle, lymph node, skeletal muscle, liver, skin, cerebral cortex, and testis (Figure 2B). The above enrichment results supported the earlier finding that the fifteen core shared genes, *DNAJB9*, *VPS53*, *SCGN*, *CMAS*, *RGS6*, *FASLG*, *ABHD10*, *ATRN*, *PLA2G2F*, *ITIH2*, *ROBO1*, *SGCG*, *SH3GL2*, *CNR1*, and *FOXN3*, were closely related to NAFLD and T2D.

3.5 Mendelian randomization analysis

No outliers were detected after processing with the “MRPRESSO” R package. The results of MR analysis in T2D and NAFLD are listed in Table 4. Among the results, regardless of

TABLE 3 Core genes after GWAS analysis and differential gene expression analysis combined.

	Gene	Cytogenetic region	Description	Remark
1	VPS53	17p13.3	VPS53 subunit of GARP complex	Not novel gene
2	SCGN	6p22.2	secretagoin, EF-hand calcium binding protein	Not novel gene
3	RGS6	14q24.2	regulator of G protein signaling 6	Not novel gene
4	SGCG	13q12.12	sarcoglycan gamma	Not novel gene
5	FOXP3	14q32.11	forkhead box N3	Not novel gene
6	DNAJB9	7q31.1	DnaJ heat shock protein family (Hsp40) member B9	Novel gene
7	CMAS	12p12.1	cytidine monophosphate N-acetylneuraminic acid synthetase	Novel gene
8	FASLG	1q24.3	Fas ligand	Novel gene
9	ABHD10	3q13.2	abhydrolase domain containing 10, depalmitoylase	Novel gene
10	ATRNL1	20p13	attractin	Novel gene
11	PLA2G2F	1p36.12	phospholipase A2 group IIF	Novel gene
12	ITIH2	10p14	inter-alpha-trypsin inhibitor heavy chain 2	Novel gene
13	ROBO1	3p12.3	roundabout guidance receptor 1	Novel gene
14	SH3GL2	9p22.2	SH3 domain containing GRB2 like 2, endophilin A1	Novel gene
15	CNR1	6q15	cannabinoid receptor 1	Novel gene

whether NAFLD or T2D was used as exposure or outcome, the *p*-value obtained by the IVW method was less than 0.05, indicating a causal relationship between T2D and NAFLD; the related *beta*-value was more than zero, which indicated that the causal relationship between T2D and NAFLD was positive; this meant that increasing exposure (T2D) increased the risk of the outcome (NAFLD). From the scatter plot of the MR results (Figure 3), it can be seen that the IVW method yielded the most significant results among the five methods that were used for MR analysis. The plot

also demonstrates the positive relationship between T2D and NAFLD, as did the forest plot (Supplementary Figure 8).

The statistical results of heterogeneity show that there was no heterogeneity between the instrumental variable SNPs (*Q*-*p*val >0.05), which can be confirmed from the funnel plot (Supplementary Figure 9). The results of the pleiotropy test show that there was no statistical difference (*p*>0.05), which indicates that there was no horizontal pleiotropic effect. Through leave-one-out analysis (Supplementary Table 8; Supplementary Figure 10), it can

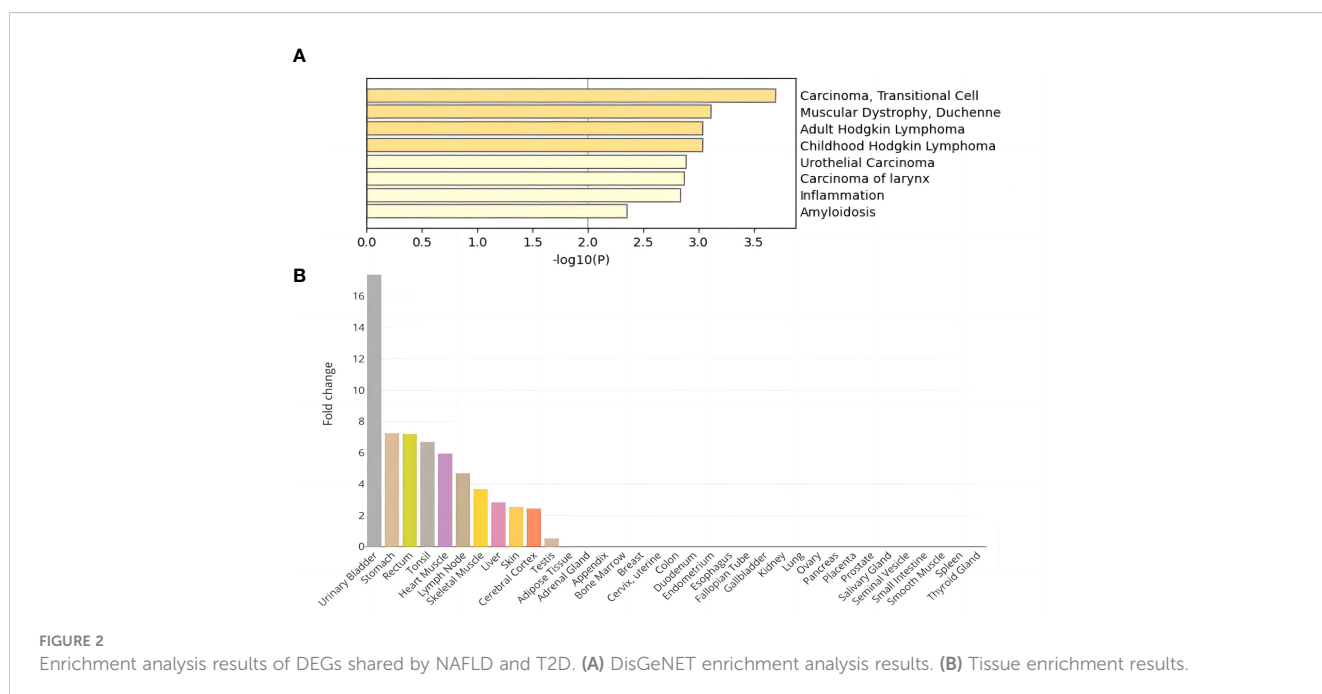


FIGURE 2 Enrichment analysis results of DEGs shared by NAFLD and T2D. (A) DisGeNET enrichment analysis results. (B) Tissue enrichment results.

TABLE 4 Results of two-sample MR analysis of NAFLD and T2D.

Outcome	Exposure	Method	nsnp	b(exposure/outcome)	se	pval	OR (95%CI)
NAFLD	T2D	MR Egger	62	0.01109	0.00652	0.09406	1.01115 (0.99832-1.02415)
		Weighted median	62	0.00278	0.00184	0.13017	1.00278 (0.99918-1.00639)
		Inverse variance weighted	62	0.00343	0.00129	0.00801	1.00344 (1.00090-1.00599)
		Simple mode	62	-0.00016	0.00461	0.97296	0.99984 (0.99084-1.00892)
		Weighted mode	62	-0.00016	0.00359	0.96526	0.99984 (0.99283-1.00690)
T2D	NAFLD	MR Egger	217	0.00344	0.00573	0.54896	1.00344 (0.99224 -1.01477)
		Weighted median	217	0.01771	0.00389	5.36E-06	1.01787 (1.01013-1.02566)
		Inverse variance weighted	217	0.02414	0.00302	1.24E-15	1.02443 (1.01839-1.03051)
		Simple mode	217	0.02673	0.01887	0.15799	1.02709 (0.98980-1.06578)
		Weighted mode	217	0.02673	0.01909	0.16297	1.02709 (0.98936 -1.06625)

be seen that no matter which SNP was removed, it would not have a fundamental impact on the results. So the MR results are robust.

4 Discussion

This study used GWAS summary data for 6,797,908 NAFLD and 8,404,432 T2D from European populations to determine the shared genetic architecture of these two phenotypes. A cross-trait meta-analysis identified 115 shared genes, and subsequent DEG analysis identified fifteen core shared genes: *DNAJB9*, *VPS53*, *SCGN*, *CMAS*, *RGS6*, *FASLG*, *ABHD10*, *ATRN*, *PLA2G2F*, *ITIH2*, *ROBO1*, *SGCG*, *SH3GL2*, *CNR1*, and *FOXN3*.

The liver is a vital organ that regulates glucose and lipid metabolism, and hepatic fat deposition is a critical factor in the pathogenesis of NAFLD and T2D (36). The twin-cycle hypothesis based on T2D explains that a gradual increase in the level of fat in

the liver can lead to IR, which weakens the ability of insulin to suppress hepatic glucose production. This leads to an aggravation of hepatic gluconeogenesis and rises in blood sugar levels (37). The excess glucose is used to synthesize triglycerides, which results in increased levels of liver fat and reduced capacity to use glucose. These processes create a vicious circle between the liver and pancreas (38). At the same time, hepatic triglyceride synthesis is increased in NAFLD patients. When the level of free fatty acids (FFAs) produced by lipoprotein lipase exceeds the lipid storage capacity of adipose tissue, β -cells will take up many fatty acids and store them as triglycerides. This damages the β cells and causes IR (39). This may eventually promote the progression of liver damage to HCC.

Relevant studies to date have shown that the mechanism of action of the mechanisms mentioned above has become a tool for the conduct of research in clinical practice. Previous studies have suggested some potential links between these mechanisms and the

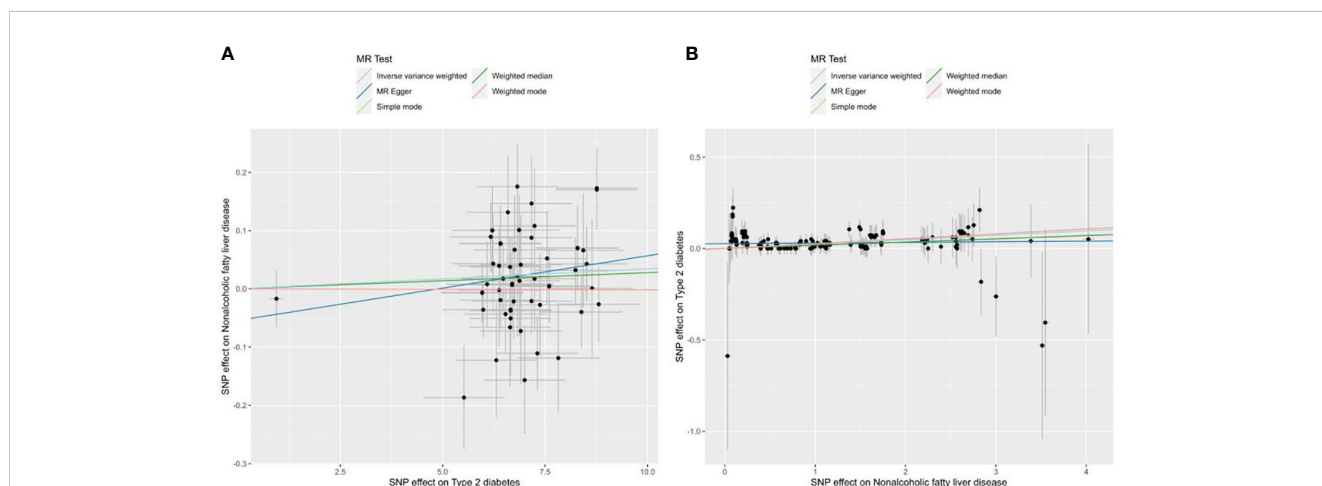


FIGURE 3 Scatter plots of the MR analysis. The light blue line shows the result of the IVW method, which has the most significant impact. The dark blue line shows the result of the MR Egger method; the light green is the result of the simple model method; the dark green is the result of the weighted median mode; and the red line represents the result of the weighted mode method. (A) Scatter plot of T2D as exposure and NAFLD as outcome. (B) Scatter plot of NAFLD as exposure and T2D as outcome.

identified, core shared genes. Forkhead box N3 (*FOXN3*), an important member of the FOX transcription factor family, is an important tumor suppressor gene that plays a crucial role in several cancers such as liver cancer, lung cancer, and colon cancer (40). The *FOXN3* gene locus is associated with fasting blood glucose levels. Hepatic *FOXN3* increases fasting blood glucose by inhibiting hepatic glucose utilization while also regulating the expression of amino acid transporters and catabolic enzymes (41, 42). Studies have shown that *FOXN3* suppresses the mRNA and protein expression of *E2F5* by inhibiting the promoter activity of potential oncogene *E2F5*, thereby inhibiting the proliferation of HCC cells *in vitro* and *in vivo* (43). Another tumor suppressor, regulator of G protein signaling 6 (*RGS6*), is upregulated in the liver of NAFLD patients, forms a complex with ATM in the liver, promotes ATM phosphorylation, and drives hepatic steatosis (44, 45). A study confirmed that hepatic *RGS6* increases oxidative stress and inflammation, which drive lipid deposition, fibrosis, and nonalcoholic fatty liver disease (46). In contrast, *RGS6* deficiency effectively ameliorated fat deposition, attenuated alcohol-dependent liver injury, and enhanced liver regeneration (47).

Other genes that may play a role in NAFLD and T2D include *SGCG*, a single-pass transmembrane glycoprotein implicated in the pathogenesis of obesity and T2D in humans (48). It has beneficial effects on glucose homeostasis, and elevated levels in diabetic patients may be compensatory for IR (49). Furthermore, *SCGN* is highly enriched in pancreatic β -cells and has pronounced effects on lipolysis and lipogenesis (50). It also regulates insulin expression and secretion, which is downregulated in type 2 diabetes (51, 52). Studies have shown that the *SCGN*-insulin interaction can stabilize insulin, enhance the hypoglycemic activity of insulin *in vivo*, and reduce hepatic steatosis and cholesterol metabolism disorders (51). In addition, the homologous gene *HCCS1* of *VPS53* also has a strong anti-tumor effect on liver cancer cells (53, 54).

Through a combination of the results of this study with the known mechanisms of action of NAFLD and T2D and related research findings, it can be shown that essential pathways affecting NAFLD and T2D include catabolism of lipids such as fatty acids, glycerides, and phospholipids. These biological processes affect lipid levels in tissues and hence affect hepatic fat accumulation and IR. Further research on this aspect of our findings should be considered.

In conclusion, the determination of triglyceride, FFA, and cholesterol levels can assist in the clinical observation of the dynamic changes in liver fat levels and IR and is of great significance in the prediction of comorbidities. At the same time, through the continuous deepening of genetic research, the development of targeted drugs to regulate the level of liver fat and the regulation of liver fat content is expected to become a key and effective treatment method for comorbidities. In addition, once the relevant mechanism of action is identified, specific gene therapy for NAFLD and T2D is expected to be realized.

One limitation of this study was that the shared genes were all screened from the results of GWAS studies in European populations, so other populations were not considered. Few replicated validation studies of the susceptibility loci associated with NAFLD and T2D have been conducted in other populations.

Genetic and environmental factors influence the genetic backgrounds of populations and result in variations in allele frequencies, which affect illness incidence rates and the findings of GWAS analyses of susceptibility genes. Therefore it is uncertain whether the susceptibility genes identified in this study exist in other populations. However, the results of this study can provide a reference for research on NAFLD combined with T2D in other populations. GWAS research involves not only different populations but also different genders and different ages, and this richness of the data should be exploited for further exploration.

It is essential to note that this study cannot avoid the shortcomings of GWAS itself, such as the fact that the study is focused on the loci that achieve the significance threshold for genome-wide association, even though these loci account only partially for the complicated heredity of the disease (55). GWAS studies often overlook signals of mild or moderate association and ignore the effects of other variants such as gene deletions, copy number variations, etc. These neglected factors may involve underlying biological mechanisms that ultimately lead to the occurrence of disease. While NAFLD and T2D are complex diseases in which genetic and environmental factors interact, the pathogenesis is often caused by mutations or abnormalities of multiple genes, and each gene may play a part in a specific pathway but its role cannot explain the whole mechanism. Therefore, the study design can be effectively improved to make up for these issues with GWAS and the complexity of the disease. For example, the candidate gene method is used to find low-frequency variants, or the data from multiple studies can be combined in a meta-analysis to increase the sample size, and rare variants with substantial genetic effects can be found in this way (56, 57).

The strength of this study was that it involved the first comprehensive use of GWAS and DEG analysis to identify shared genes for NAFLD and T2D. During gene screening, strict thresholds were used to ensure the accuracy of the results, and significant shared genes were discovered efficiently. The study reconfirmed the association of the unveiled core genes, *VPS53*, *SCGN*, *RGS6*, *SGCG*, and *FOXN3*, with NAFLD and T2D, which had been reported in previous studies. The core genes *DNAJB9*, *CMAS*, *FASLG*, *ABHD10*, *ATRN*, *PLA2G2F*, *ITIH2*, *ROBO1*, *SH3GL2*, and *CNR1* were found to be related to NAFLD and T2D for the first time, and this provides a new research target for the precise treatment of NAFLD and T2D comorbidities.

5 Conclusion

In summary, this study found a causal relationship between NAFLD and T2D, which will be beneficial for the elucidation of the pathogenesis of NAFLD and T2D comorbidities. Fifteen core genes, *DNAJB9*, *VPS53*, *SCGN*, *CMAS*, *RGS6*, *FASLG*, *ABHD10*, *ATRN*, *PLA2G2F*, *ITIH2*, *ROBO1*, *SGCG*, *SH3GL2*, *CNR1*, and *FOXN3*, were identified as shared between NAFLD and T2D. This finding provided new ideas for the genetic study of NAFLD combined with T2D. Further gene expression verification and functional mechanism research should be carried out on these candidate

genes in the future to explore the specific biological mechanisms of NAFLD and T2D comorbidities and to provide new drug-targeting sites for the prevention and treatment of comorbidities.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Author contributions

YT participated in data analysis, drafting, writing, and revising the paper. QH participated in the revision of the manuscript and carried out a strict review of the manuscript. KC conceived, designed, coordinated the study, and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the City University of Hong Kong new research initiatives/infrastructure support from central funding (APRC). (grant no. 9610401)

Acknowledgments

In the study, GWAS analysis was performed using the GWAS Catalog database resource of the National Human Genome

Research Institute. Furthermore, the differential gene expression analysis used the GEO database resource of the National Center for Biotechnology Information. We thank Anstee QM, Wood AR, Jochen H, Baker SS, Bugliani M, Marselli L, and other researchers for their GWAS summary statistics and GEO data. We also thank all participants who provided study samples. We thank the strong support of the Department of Biomedical Sciences, City University of Hong Kong.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fendo.2023.1050049/full#supplementary-material>

References

- Hardwick RN, Fisher CD, Canet MJ, Lake AD, Cherrington NJ. Diversity in antioxidant response enzymes in progressive stages of human nonalcoholic fatty liver disease. *Drug Metab Dispos* (2010) 38:2293–301. doi: 10.1124/dmd.110.035006
- Bellan M, Colletta C, Barbaglia MN, Salmi L, Clerici R, Mallela VR, et al. Severity of nonalcoholic fatty liver disease in type 2 diabetes mellitus: Relationship between nongenetic factors and PNPLA3/HSD17B13 polymorphisms. *Diabetes Metab J* (2019) 43:700–10. doi: 10.4093/dmj.2018.0201
- Marchesini G, Brizi M, Bianchi G, Tomassetti S, Bugianesi E, Lenzi M, et al. Nonalcoholic fatty liver disease: A feature of the metabolic syndrome. *Diabetes* (2001) 50:1844–50. doi: 10.2337/diabetes.50.8.1844
- Pan Q, Lin S, Li Y, Liu L, Li X, Gao X, et al. A novel GLP-1 and FGF21 dual agonist has therapeutic potential for diabetes and non-alcoholic steatohepatitis. *EBioMedicine* (2021) 63:103202. doi: 10.1016/j.ebiom.2020.103202
- Shimizu M, Suzuki K, Kato K, Jojima T, Iijima T, Murohisa T, et al. Evaluation of the effects of dapagliflozin, a sodium-glucose co-transporter-2 inhibitor, on hepatic steatosis and fibrosis using transient elastography in patients with type 2 diabetes and non-alcoholic fatty liver disease. *Diabetes Obes Metab* (2019) 21:285–92. doi: 10.1111/dom.13520
- Mu W, Cheng XF, Liu Y, Lv QZ, Liu GL, Zhang JG, et al. Potential nexus of non-alcoholic fatty liver disease and type 2 diabetes mellitus: Insulin resistance between hepatic and peripheral tissues. *Front Pharmacol* (2018) 9:1566. doi: 10.3389/fphar.2018.01566
- Targher G, Bertolini L, Padovani R, Rodella S, Tessari R, Zenari L, et al. Prevalence of nonalcoholic fatty liver disease and its association with cardiovascular disease among type 2 diabetic patients. *Diabetes Care* (2007) 30:1212–8. doi: 10.2337/dc06-2247
- Jarvis H, Craig D, Barker R, Spiers G, Stow D, Anstee QM, et al. Metabolic risk factors and incident advanced liver disease in non-alcoholic fatty liver disease (NAFLD): A systematic review and meta-analysis of population-based observational studies. *PLoS Med* (2020) 17:e1003100. doi: 10.1371/journal.pmed.1003100
- Mantovani A, Byrne CD, Bonora E, Targher G. Nonalcoholic fatty liver disease and risk of incident type 2 diabetes: A meta-analysis. *Diabetes Care* (2018) 41:372–82. doi: 10.2337/dc17-1902
- Pinero F, Pages J, Marciano S, Fernandez N, Silva J, Anders M, et al. Fatty liver disease, an emerging etiology of hepatocellular carcinoma in Argentina. *World J Hepatol* (2018) 10:41–50. doi: 10.4254/wjvh.v10.i1.41
- Du X, Deforest N, Majithia AR. Human genetics to identify therapeutic targets for NAFLD: Challenges and opportunities. *Front Endocrinol (Lausanne)* (2021) 12:777075. doi: 10.3389/fendo.2021.777075
- Haynes WA, Higdon R, Stanberry L, Collins D, Kolker E. Differential expression analysis for pathways. *PLoS Comput Biol* (2013) 9:e1002967. doi: 10.1371/journal.pcbi.1002967
- Wood AR, Tyrrell J, Beaumont R, Jones SE, Tuke MA, Ruth KS, et al. Variants in the FTO and CDKAL1 loci have recessive effects on risk of obesity and type 2 diabetes, respectively. *Diabetologia* (2016) 59:1214–21. doi: 10.1007/s00125-016-3908-5
- Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tiniakos D, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort(☆). *J Hepatol* (2020) 73:505–15. doi: 10.1016/j.jhep.2020.04.003
- Baker SS, Baker RD, Liu W, Nowak NJ, Zhu L. Role of alcohol metabolism in non-alcoholic steatohepatitis. *PLoS One* (2010) 5:e9570. doi: 10.1371/journal.pone.0009570

16. Marselli L, Thorne J, Dahiya S, Sgroi DC, Sharma A, Bonner-Weir S, et al. Gene expression profiles of beta-cell enriched tissue obtained by laser capture microdissection from subjects with type 2 diabetes. *PLoS One* (2010) 5:e11499. doi: 10.1371/journal.pone.0011499
17. Dominguez V, Raimondi C, Somanath S, Bugliani M, Loder MK, Edling CE, et al. Class II phosphoinositide 3-kinase regulates exocytosis of insulin granules in pancreatic beta cells. *J Biol Chem* (2011) 286:4216–25. doi: 10.1074/jbc.M110.200295
18. Ahrens M, Ammerpohl O, Von Schönfels W, Kolarova J, Bens S, Itzel T, et al. DNA Methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab* (2013) 18:296–302. doi: 10.1016/j.cmet.2013.07.004
19. Laurie CC, Doheny KF, Mirel DB, Pugh EW, Bierut LJ, Bhangale T, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet Epidemiol* (2010) 34:591–602. doi: 10.1002/gepi.20516
20. Ani A, van der Most PJ, Snieder H, Vaez A, Nolte IM. GWASinspector: Comprehensive quality control of genome-wide association study results. *Bioinformatics* (2021) 37:129–30. doi: 10.1093/bioinformatics/btaa1084
21. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature* (2015) 526:68–74. doi: 10.1038/nature15393
22. Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am J Hum Genet* (2015) 96:21–36. doi: 10.1016/j.ajhg.2014.11.011
23. Park H, Li X, Song YE, He KY, Zhu X. Multivariate analysis of anthropometric traits using summary statistics of genome-wide association studies from GIANT consortium. *PLoS One* (2016) 11:e0163912. doi: 10.1371/journal.pone.0163912
24. Li X, Zhu X. Cross-phenotype association analysis using summary statistics from GWAS. *Methods Mol Biol* (2017) 1666:455–67. doi: 10.1007/978-1-4939-7274-6_22
25. Trepo E, Valenti L. Update on NAFLD genetics: From new variants to the clinic. *J Hepatol* (2020) 72:1196–209. doi: 10.1016/j.jhep.2020.02.020
26. Alexandrov T, Ovchinnikova K, Palmer A, Kovalev V, Tarasov A, Stuart L, et al. METASPACE: A community-populated knowledge base of spatial metabolomes in health and disease. *BioRxiv* (2019) 539478. doi: 10.1101/539478
27. Jain A, Tuteja G. TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* (2019) 35:1966–7. doi: 10.1093/bioinformatics/bty890
28. Bardou P, Mariette J, Escudière F, Djemiel C, Klopp C. Jvenn: an interactive Venn diagram viewer. *BMC Bioinf* (2014) 15:1–7. doi: 10.1186/1471-2105-15-293
29. Jafari M, Ansari-Pour N. Why, when and how to adjust your p values? *Cell J (Yakhteh)* (2019) 20:604–7. doi: 10.22074/cellj.2019.5992
30. Xiang K, Hu Y-Q, HE Y-S, Wang P, Tao S-S, Chen Y, et al. Non-causal effect of circulating vitamin d levels on the risk of rheumatoid arthritis: A two-sample mendelian randomization study. (2021). doi: 10.21203/rs.3.rs-479402/v1
31. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet* (2018) 50:693–8. doi: 10.1038/s41588-018-0099-7
32. Mounier N, Kutalik Z. Bias correction for inverse variance weighting mendelian randomization. *BioRxiv* (2021). doi: 10.1101/2021.03.26.437168
33. Bowden J, Smith GD, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol* (2016) 40:304–14. doi: 10.1002/gepi.21965
34. Burgess S, Thompson SG. Interpreting findings from mendelian randomization using the MR-egger method. *Eur J Epidemiol* (2017) 32:377–89. doi: 10.1007/s10654-017-0255-x
35. Hartwig FP, Davey SG, Bowden J. Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol* (2017) 46:1985–98. doi: 10.1093/ije/dyx102
36. Dongiovanni P, Stender S, Pietrelli A, Mancina RM, Cespiati A, Petta S, et al. Causal relationship of hepatic fat with liver damage and insulin resistance in nonalcoholic fatty liver. *J Intern Med* (2018) 283:356–70. doi: 10.1111/joim.12719
37. Taylor R. Type 2 diabetes and remission: practical management guided by pathophysiology. *J Internal Med* (2021) 289:754–70. doi: 10.1111/joim.13214
38. Taylor R. Putting insulin resistance into context by dietary reversal of type 2 diabetes. *J R Coll Physicians Edinburgh* (2017) 47:168–71. doi: 10.4997/JRCP.2017.216
39. Birkenfeld AL, Shulman GI. Nonalcoholic fatty liver disease, hepatic insulin resistance, and type 2 diabetes. *Hepatology* (2014) 59:713–23. doi: 10.1002/hep.26672
40. Lin Z, Chen M, Wan Y, Lei L, Ruan H. miR-574-5p targets FOXN3 to regulate the invasion of nasopharyngeal carcinoma cells via wnt/beta-catenin pathway. *Technol Cancer Res Treat* (2020) 19:1533033820971659. doi: 10.1177/1533033820971659
41. Erickson ML, Karanth S, Ravussin E, Schlegel A. FOXN3 hyperglycemic risk allele and insulin sensitivity in humans. *BMJ Open Diabetes Res Care* (2019) 7:e000688. doi: 10.1136/bmjdr-2019-000688
42. Karanth S, Chaurasia B, Bowman FM, Tippetts TS, Holland WL, Summers SA, et al. FOXN3 controls liver glucose metabolism by regulating gluconeogenic substrate selection. *Physiol Rep* (2019) 7:e14238. doi: 10.14814/phyz.2.14238
43. Sun J, Li H, Huo Q, Cui M, Ge C, Zhao F, et al. The transcription factor FOXN3 inhibits cell proliferation by downregulating E2F5 expression in hepatocellular carcinoma cells. *Oncotarget* (2016) 7:43534. doi: 10.18632/oncotarget.9780
44. Mahata T, Sengar AS, Basak M, Das K, Pramanick A, Verma SK, et al. Hepatic regulator of G protein signaling 6 (RGS6) drives non-alcoholic fatty liver disease by promoting oxidative stress and ATM-dependent cell death. *Redox Biol* (2021) 46:102105. doi: 10.1016/j.redox.2021.102105
45. Wang Z, Chen J, Wang S, Sun Z, Lei Z, Zhang HT, et al. RGS6 suppresses TGF-beta-induced epithelial-mesenchymal transition in non-small cell lung cancers via a novel mechanism dependent on its interaction with SMAD4. *Cell Death Dis* (2022) 13:656. doi: 10.1038/s41419-022-05093-0
46. Dao W, Xiao Z, Yang W, Luo X, Xia H, Lu Z. RGS6 drives spinal cord injury by inhibiting AMPK pathway in mice. *Dis Markers* (2022) 2022:4535652. doi: 10.1155/2022/4535652
47. Stewart A, Maity B, Anderegg SP, Allamargot C, Yang J, Fisher RA. Regulator of G protein signaling 6 is a critical mediator of both reward-related behavioral and pathological responses to alcohol. *Proc Natl Acad Sci U.S.A.* (2015) 112:E786–95. doi: 10.1073/pnas.1418795112
48. Israeli D, Cosette J, Corre G, Amor F, Poupiot J, Stockholm D, et al. An AAV-SGCG dose-response study in a gamma-sarcoglycanopathy mouse model in the context of mechanical stress. *Mol Ther Methods Clin Dev* (2019) 13:494–502. doi: 10.1016/j.jomtm.2019.04.007
49. Kuhn T, Kaiser K, Lebek S, Altenhofen D, Knebel B, Herwig R, et al. Comparative genomic analyses of multiple backcross mouse populations suggest SGCG as a novel potential obesity-modifier gene. *Hum Mol Genet* (2022) 31:4019–33. doi: 10.1093/hmg/ddac150
50. Chin YX, Mi Y, Cao WX, LiM PE, Xue CH, Tang QJ. A pilot study on anti-obesity mechanisms of kappaphycus alvarezii: The role of native kappa-carrageenan and the leftover sans-carrageenan fraction. *Nutrients* (2019) 11. doi: 10.3390/nu11051133
51. Sharma AK, Khandelwal R, Chadalawada S, Ram NS, Raj TA, Kumar MJM, et al. SCGN administration prevents insulin resistance and diabetic complications in high-fat diet fed animals. *BioRxiv* (2017), 189324. doi: 10.1101/189324
52. Liu Z, Tan S, Zhou L, Chen L, Liu M, Wang W, et al. SCGN deficiency is a risk factor for autism spectrum disorder. *Signal Transduction Targeted Ther* (2023) 8:3. doi: 10.1038/s41392-022-01225-2
53. Perez-Victoria FJ, Schindler C, Magadan JG, Mardones GA, Delevoe C, Romao M, et al. Ang2/fat-free is a conserved subunit of the golgi-associated retrograde protein complex. *Mol Biol Cell* (2010) 21:3386–95. doi: 10.1091/mbc.e10-05-0392
54. Peng H, Zheng J, Su Q, Feng X, Peng M, Gong L, et al. VPS53 suppresses malignant properties in colorectal cancer by inducing the autophagy signaling pathway. *Oncotargets Ther* (2020) 13:10667–75. doi: 10.2147/OTT.S254823
55. Hu C, Jia W. Diabetes in China: epidemiology and genetic risk factors and their clinical utility in personalized medication. *Diabetes* (2018) 67:3–11. doi: 10.2337/dbi17-0013
56. Quan C, Zhang X-J. Research strategies for the next step of genome-wide association study. *Yi Chuan = Hereditas* (2011) 33:100–8. doi: 10.3724/SP.J.1005.2011.00100
57. Tomlinson IP, Carvajal-Carmona LG, Dobbins SE, Tenesa A, Jones AM, Howarth K, et al. Multiple common susceptibility variants near BMP pathway loci GREM1, BMP4, and BMP2 explain part of the missing heritability of colorectal cancer. *PLoS Genet* (2011) 7:e1002105. doi: 10.1371/journal.pgen.1002105