ORIGINAL RESEARCH

# Multi-Omics Analysis and Validation of Cell Senescence-Related Genes Associated with Non-Alcoholic Fatty Liver Disease

Jianhua Gong[1–3,]*, Zhijie Qin[3,]*, Yihao Xiao[3,]*, Jixue Li[3], Qing Wang[3], Liping Lei[4], Jiangfa Li[3]

[1]Department of Hepatobiliary and Pancreatic Surgery, The First College of Clinical Medical Science, China Three Gorges University, Yichang, Hubei, 443003, People's Republic of China; [2]Department of Hepatobiliary and Pancreatic Surgery, Yichang Central People's Hospital, Yichang, Hubei, 443003, People's Republic of China; [3]Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital of Guilin Medical University, Guilin, Guangxi, 541001, People's Republic of China; [4]Department of geriatric medicine, The First Affiliated Hospital of Guilin Medical University, Guilin, Guangxi, 541001, People's Republic of China

*These authors contributed equally to this work

Correspondence: Jianhua Gong, Department of Hepatobiliary and Pancreatic Surgery, The First College of Clinical Medical Science, China Three Gorges University, Yichang Central People's Hospital, No. 183 Yiling Ave, Yichang, Hubei, 443003, People's Republic of China, Tel +86-13886703242, Email 38891488@qq.com; Jiangfa Li, Department of Hepatobiliary and Pancreatic Surgery, The First Affiliated Hospital of Guilin Medical University, Guilin, Guangxi, 541001, People's Republic of China, Tel +86-18907839510, Email 247546160@qq.com

**Objective:** To assess causal links between senescence-related genes and non-alcoholic fatty liver disease (NAFLD) using summary-data Mendelian randomization (SMR) and colocalization analyses.

**Methods:** Our study examined the relationship between senescence and NAFLD by integrating DNA methylation, gene expression, and protein quantitative trait loci (mQTL, eQTL, and pQTL) data. Summary statistics for NAFLD were sourced from a previous study (discovery) and the FinnGen database (replication), with additional cohorts for nonalcoholic steatohepatitis and liver fibrosis. Genetic variants near senescence-related genes were used as instrumental variables to assess causal relationships. Colocalization analysis was performed to confirm shared causal variants and liver-specific eQTL data were used for validation. Furthermore, we validated findings using cell and mouse models of NAFLD. Cell models were treated with oleic acid, and NAFLD mice were induced using a high-fat diet.

**Results:** We identified 40 mQTLs, 9 eQTLs, and 3 pQTLs significantly linked to NAFLD in the discovery cohort. Multi-omics data highlighted three genes—*S100A6, ENDOG*, and *TP53I3*—as potential causal contributors. Notably, S100A6 was confirmed at both the methylation sites (cg24155129 and cg01910639) and gene expression levels, with methylation at these CpG sites significant regulating its expression. Liver-specific validation revealed that *ENDOG* expression was negatively associated with NAFLD, consistent with findings in blood. Finally, differential expression of all three genes was confirmed in cell models, with S100A6 and *ENDOG* further validated in a mouse model.

**Conclusion:** Our findings suggest that *S100A6, ENDOG*, and *TP53I3* are associated with NAFLD, providing insights for further research into the disease's underlying etiology.

**Keywords:** cellular senescence, non-alcoholic fatty liver disease, causal inference, Mendelian randomization, colocalization

## Introduction

Nonalcoholic fatty liver disease (NAFLD) is a widespread chronic liver condition, affecting an estimated global prevalence of 37.8%, which has significantly increased from 25.5% around 2005.[1] The terminology for this condition has evolved to metabolic dysfunction-associated steatotic liver disease (MASLD), which more accurately reflects its metabolic basis.[2] However, we continue to use NAFLD in this manuscript for consistency with historical GWAS datasets. NAFLD encompasses a spectrum of liver disorders, ranging from simple steatosis to nonalcoholic steatohepatitis (NASH), which can potentially progress to advanced stages like fibrosis, cirrhosis, and hepatocellular carcinoma.[3] The disease is linked to a high risk of liver-related morbidity and metabolic syndromes, imposing a substantial burden on

healthcare systems.[3] Despite advancements in its treatments, the exact causes of NAFLD are not completely understood and are likely influenced by a complex interplay of genetic and environmental factors, such as lifestyle choices, dietary habits, and exposure to certain medications or toxins.[4]

Cell senescence is a state of irreversible cell-cycle arrest that occurs in response to various stressors, such as DNA damage, oxidative stress, and telomere shortening.[5] It is characterized by a distinct secretory phenotype known as the senescence-associated secretory phenotype (SASP), which involves the secretion of pro-inflammatory cytokines, chemokines, and matrix metalloproteinases.[6] In the context of NAFLD, cellular senescence is thought to play a role in the transition from simple steatosis to NASH, and potentially to more advanced stages such as fibrosis and cirrhosis.[7] The SASP can create a pro-inflammatory and profibrotic microenvironment, which may contribute to the progression of liver disease.[8] Additionally, senescent hepatocytes and hepatic stellate cells may directly influence the development of liver cancer through the secretion of factors that promote cell proliferation and invasion.[9,10] However, whether senescence is a marker or a potential mediator of NAFLD progression remains unclear. Therefore, a comprehensive analysis of senescence-related genes in NAFLD using a robust method is necessary to determine whether senescence is a cause or consequence of NAFLD.

Mendelian randomization (MR) offers an alternative to conduct causality assumptions that cannot be readily obtained from conventional observational studies.[11] By utilizing randomly allocated genetic variants as instrumental variables (IVs), MR investigates the causal connections between two factors, thereby mitigating confounding bias and reverse causality.[12,13] Summary-data-based Mendelian randomization (SMR) utilizes independent genome-wide association study (GWAS) summary statistics and quantitative trait locus (QTL) data to identify causal genes from GWAS results.[14] Unlike traditional MR analysis, SMR combines multi-omics data including genetic, epigenetic, proteomic evidence to improve the accuracy and reliability of causal inference. Using this approach, potential causal associations between senescence-related genes and NAFLD were identified, followed by a heterogeneity in independent instruments (HEIDI) test.[15]

Here, an SMR analysis was executed to investigate the potential associations of senescence-related genes methylation, expression, and protein abundance with the risk of NAFLD.

## Methods

### Study Design

Figure 1 summarized the overall study design. The current SMR analysis was based on publicly available datasets obtained from previous studies and the FinnGen. In this study, IVs for senescence-related genes extracted at the methylation, gene expression and protein abundance levels. Subsequent SMR analysis was conducted for NAFLD, NASH or liver cirrhosis at these levels. To strengthen the causal inference, colocalization analysis was conducted. Through the integration of results obtained from SMR analysis at these levels, we identified causal candidate genes or proteins. The reporting of MR analysis adhered to the Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) guidelines.[16]

### Data Sources

GWAS summary statistics for NAFLD was obtained from publicly available databases. The primary discovery dataset (GCST90275041), which comprised 6,623 cases and 26,318 controls of the European ancestry,[17] was supplemented with validation from three independent cohorts: NAFLD (2,568 cases and 409,613 controls) and NASH (175 cases and 412,006 controls) cohorts from FinnGen, and cohort of liver cirrhosis in NAFLD (1,106 cases and 8,571 controls).[18] The definition of diseases is based on the International Classification of Diseases, 9th and 10th Revision (ICD-9 and ICD-10). The detailed information for each phenotypic outcome data was provided in Supplementary Table 1. There is no overlap in samples between the discovery and validation cohorts. This study utilized summary statistics from public GWAS studies, for which ethic approvement has been obtained. Consequently, no further ethical approval was necessary.

949 senescence-related genes were extracted from the CellAge (https://genomics.senescence.info/cells/) database (Build 3) using the keyword "cell senescence". QTLs can uncover the relationships between SNPs and variations in DNA
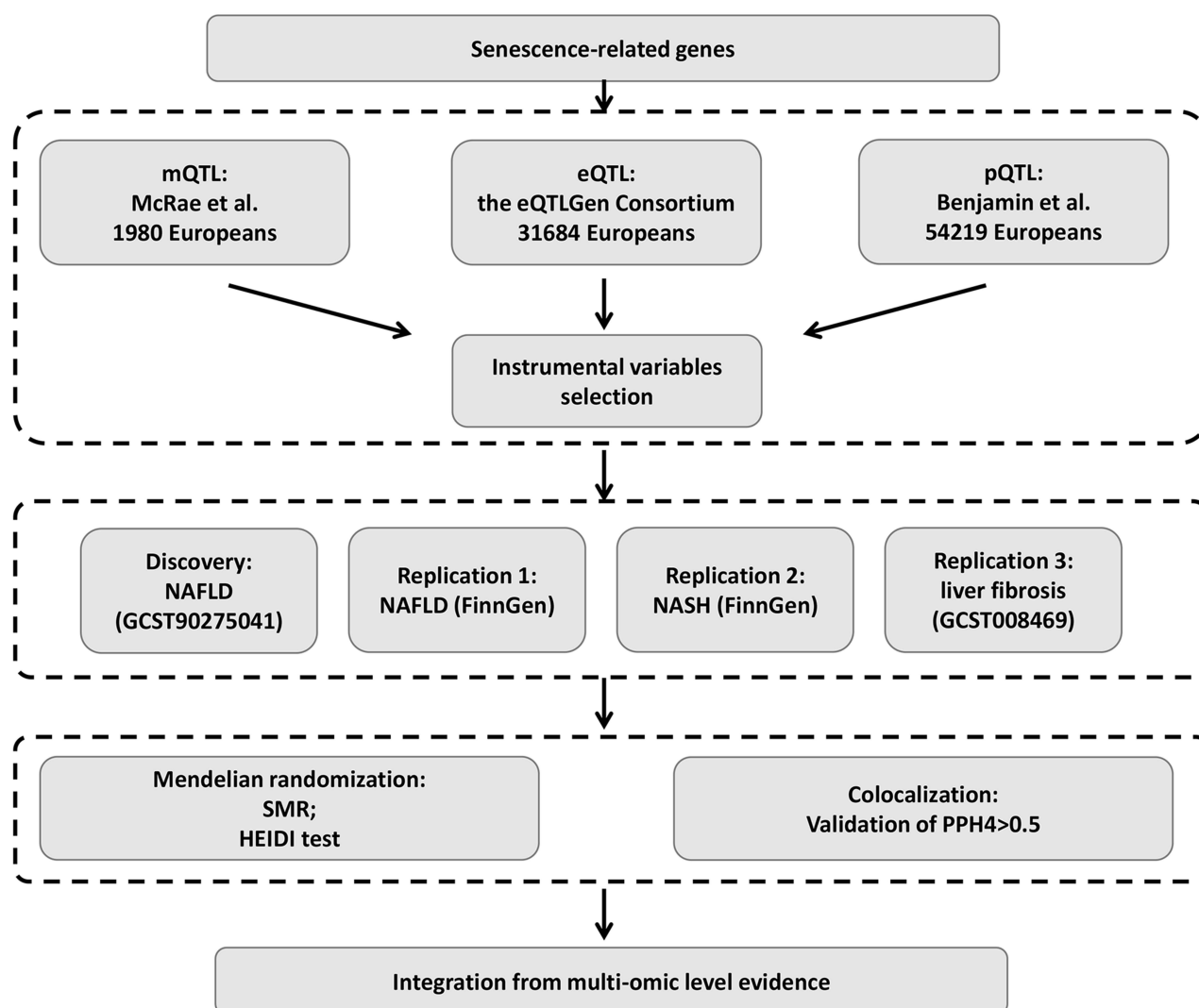
**Figure 1** Overall study design of the MR analysis. A flow chart depicts how the SMR analysis was conducted in this study.

methylation, gene expression, and protein abundance. Blood eQTL summary statistics were obtained from eQTLGen, encompassing genetic data of blood gene expression in 31,684 individuals from 37 datasets.[18] Blood mQTL summary data were generated from a meta-analysis of two European cohorts: the Brisbane Systems Genetics Study (n = 614) and the Lothian Birth Cohorts (n = 1366).[15] Data on genetic associations with circulating protein levels were sourced from a protein quantitative trait loci (pQTL) investigation involving 54219 individuals.[17]

## Summary-Data-Based MR Analysis

SMR was employed to assess the association of senescence-related genes methylation, expression, and protein abundance with the risk of NAFLD. Leveraging top associated cis-QTLs, SMR achieved enhanced statistical power compared to conventional MR analysis, particularly in scenarios with large sample sizes and independent datasets for exposure and outcome. Cis-QTLs were selected based on a ±1000 kb window around the gene of interest and a significance threshold of $5.0×10^{-8}$.[19] SNPs with allele frequency differences exceeding 0.2 between datasets were excluded. Thresholds for pQTL, mQTL, and eQTL were set at 0.05. The original version of SMR only uses the lead cis-QTL variant as IV, and it has since been extended to SMR-multi to accommodate the potential presence of multiple cis-xQTL causal variants.[15]

In addition to exploring the causal associations between QTLs and NAFLD, the study further investigated the causal relationships between mQTL as the exposure and eQTL as the outcome. The key findings linking mQTL and eQTL with

**8823**

NAFLD are highlighted as signals of particular interest between mQTL and eQTL. Additionally, this study extends to the causal connections between eQTL and pQTL, with a focus on key genes from the mQTL-eQTL association and significant findings from NAFLD GWAS analysis associated with pQTL.

To differentiate between pleiotropy and linkage, we employed the HEIDI test, with P-HEIDI <0.05 indicating potential pleiotropy and leading to exclusion from the analysis. Associations meeting the criteria (p SMR < 0.05, multi-SNP-based P-value < 0.05 and P-HEIDI > 0.05) were considered for colocalization analysis in mQTL, eQTL and pQTL datasets.

## Colocalization Analysis

We conducted colocalization analyses using the R package "coloc" to identify shared causal variants between NAFLD and the mQTLs, eQTLs, or pQTLs of senescence-related genes. In these analyses, five different posterior probabilities are reported, corresponding to the following hypotheses: H0 (no causal variants for either trait), H1 (a causal variant for gene expression only), H2 (a causal variant for disease risk only), H3 (distinct causal variants for two traits), and H4 (the same shared causal variant for both traits).[20] When GWAS signals and QTLs are found to colocalize, it suggests that the GWAS locus may influence the complex trait or disease phenotype by modulating gene expression or splicing.[21,22] For colocalization analysis, all SNPs within 1000 kb upstream and downstream of each top cis-QTL were retrieved to determine the posterior probability of H4 (PPH4). A PPH4 > 0.5 was used as the cut-off, indicating strong evidence of colocalization between GWAS and QTL associations.[23]

## Cell Culture and Treatments

The human liver-7702 (HL-7702) cell line was obtained from the Cell Bank of the Type Culture Collection of the Chinese Academy of Sciences (Shanghai, China). The complete culture medium for the HL-7702 cells consisted of DMEM/F-12 (1:1) (Gibco, 11330–032) with 89 mL ITS liquid medium (Sigma, I3146), 1 mL dexamethasone (Sigma, D4902-100mg), and 10 mL FBS (Gibco). The cells were cultured at 37°C in a 5% CO2 incubator. Once the cells reached 60–70% confluence, they were divided into two groups (n=3): (1) Control group (treated with normal saline for 24 hours) and (2) NAFLD group (treated with 1 mM oleic acid (OA; Sigma, USA) for 24 hours). Cell conditions were assessed using Oil Red O staining.

## Creation of NAFLD Mouse Model and Histological Process

Six 8-week-old male, C57BL/6 WT mice, were utilized in this experiment. In the experimental group, male C57BL/6 mice were given a diet high in fat, sugar, and cholesterol, along with a high-sugar solution (23.1g/Ld fructose and 18.9g/Ld glucose) and a weekly low dose (0.2 ul /g) of carbon tetrachloride (dissolved in olive oil) administered intraperitoneally. After 16 weeks, NAFLD/NASH mouse models were established. In the control group, male C57 BL/6 mice were given a standard maintenance diet and a weekly intraperitoneal injection of the same dose of olive oil as the experimental group.

After 16 weeks, all mice were euthanized, and blood was drawn from the inferior vena cava using a 1 mL syringe and centrifuged at 3000 rpm for 15 minutes. The supernatant was collected to obtain mouse plasma. Serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels were measured using an automatic biochemical analyzer (ANTECH Diagnostics, Los Angeles, CA, USA). Liver tissue samples were also collected from the mice. A portion of each liver sample was immediately frozen in liquid nitrogen in an EP tube. The remaining tissue was fixed in formalin, embedded in paraffin, and stained with hematoxylin and eosin (HE). A section of the freshly frozen liver tissue, 8 μm thick, was stained with Masson. All animal experiments received approval from the Institutional Animal Care and Use Committee of Guilin Medical University (GLMC-IACUC-20241090). All animal experiments strictly adhered to the National Standards for Laboratory Animal Welfare issued by the Chinese government (GB/T 35892–2018) and the Guide for the Care and Use of Laboratory Animals (National Research Council, 8th Edition, 2011).

## Quantitative Reverse Transcription-Polymerase Chain Reaction (qRT-PCR)

The frozen liver tissue was weighed, lysed, homogenized, and mixed with anhydrous ethanol. RNA was extracted from both mouse liver tissues and the HL-7702 cell line using TRIzol reagent (VAZYME, China). After extraction and elution through an RNA binding column, purified total RNA samples were obtained. cDNA was synthesized using the first strand cDNA synthesis kit. SYBR Green qRT-PCR premix was used for quantitative PCR, with gene expression levels normalized to GAPDH. RNA reverse transcription was performed with the PrimeScript™ RT Reagent Kit (VAZYME, China), and qRT-PCR was conducted using an FX Connect system (VAZYME, China) and SYBR® Green Supermix (VAZYME, China). qRT-PCR was performed in triplicate, with primer details provided in Supplementary Table 2.

## Statistical Analysis

All statistical analyses were performed using R (v4.3.0). The R package "ggplot2" and "ggrepel" was used for Manhattan plot generation, and "forestplot" for forest plot generation. The code for SMRLocusPlot and SMREffectPlot was sourced from Zhu et al.[14]

# Results

## Senescence-Related Genes Methylation and NAFLD

Results for causal effects of senescence-related genes methylation on NAFLD were visualized in Figure 2A (See full results in Supplementary Table 3). A total of 143 methylation loci (58 genes) passed the screening criteria (P-SMR < 0.05, multi-SNP-based P-value < 0.05 and P-HEIDI > 0.05). Of the identified signals, 40 near 13 unique genes were found to have strong colocalization evidence support (PPH4 >0.5) including *ENDOG* (cg13630871), *S100A6* (cg24155129, cg01910639) and *TP5313* (cg14273083). Specifically, *ENDOG* methylation at cg13630871 (OR = 1.02, 95% CI = 1–1.04) was linked to an increased risk of NAFLD. Conversely, certain methylation loci exhibited divergent association with NAFLD, such as *S100A6*, with cg24155129 (OR = 0.94, 95% CI = 0.9–0.98) linked to a decreased incidence of NAFLD and cg01910639 demonstrating the opposite (OR = 1.03, 95% CI = 1.01–1.06). The colocalization for representative methylation loci and NAFLD was visualized in Figure 2B. Among these identified CpG sites, the association for *CD34* (cg15031826), *PPARG* (cg04632671), *FOXP1* (cg06175008), *TACC3* (cg10756475), *FGFR3* (cg07041428, cg25342568, cg01464969, cg14661159, cg14101193, cg07458712) were replicated in the NAFLD replication cohort (FinnGen). The detailed associations in the NAFLD, NASH and liver cirrhosis replication cohorts were provided in Supplementary Tables 4–6.
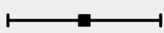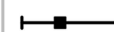
## Senescence-Related Genes Expression and NAFLD

Causal effects of senescence-related genes expression on NAFLD were presented in Figure 3A (See full results in Supplementary Table 7). A total of 16 genes were found to be associated with NAFLD (P-SMR < 0.05, multi-SNP-based P-value < 0.05 and P-HEIDI > 0.05), in which *S100A6, DTL, DNMT3A, ATG7, THRB, EGR2, FOXO1* and *CHEK2* were positively associated with NAFLD incidence. Specifically, *S100A6* (OR = 1.11, 95% CI = 1.04–1.19) was a potential risk factor for NAFLD and *ENDOG* (OR = 0.99, 95% CI = 0.97–1) exhibited the opposite. Among the loci corresponding to these genes, colocalization between representative genes and NAFLD was visualized (PPH4 > 0.5) (Figure 3B and C). Among the identified genes, none of them were replicated in the NAFLD cohort, NASH cohort and liver cirrhosis cohort (Supplementary Tables 8–10).

## Senescence-Related Protein Abundance and NAFLD

Causal effects of senescence-related protein abundance on NAFLD were presented in Figure 4A (See full results in Supplementary Table 11). In total, 6 proteins were found to be associated with NAFLD at the criteria (P-SMR < 0.05, multi-SNP-based P-value < 0.05 and P-HEIDI > 0.01), in which EIF2AK3, TIGAR and ING1 were positively associated with NAFLD incidence. Specifically, ING1 (OR = 1.16, 95% CI = 1.02–1.31) was a potential risk factor for NAFLD. Colocalization analysis between representative proteins and NAFLD were visualized (PPH4 > 0.5) Figure 4B and C.
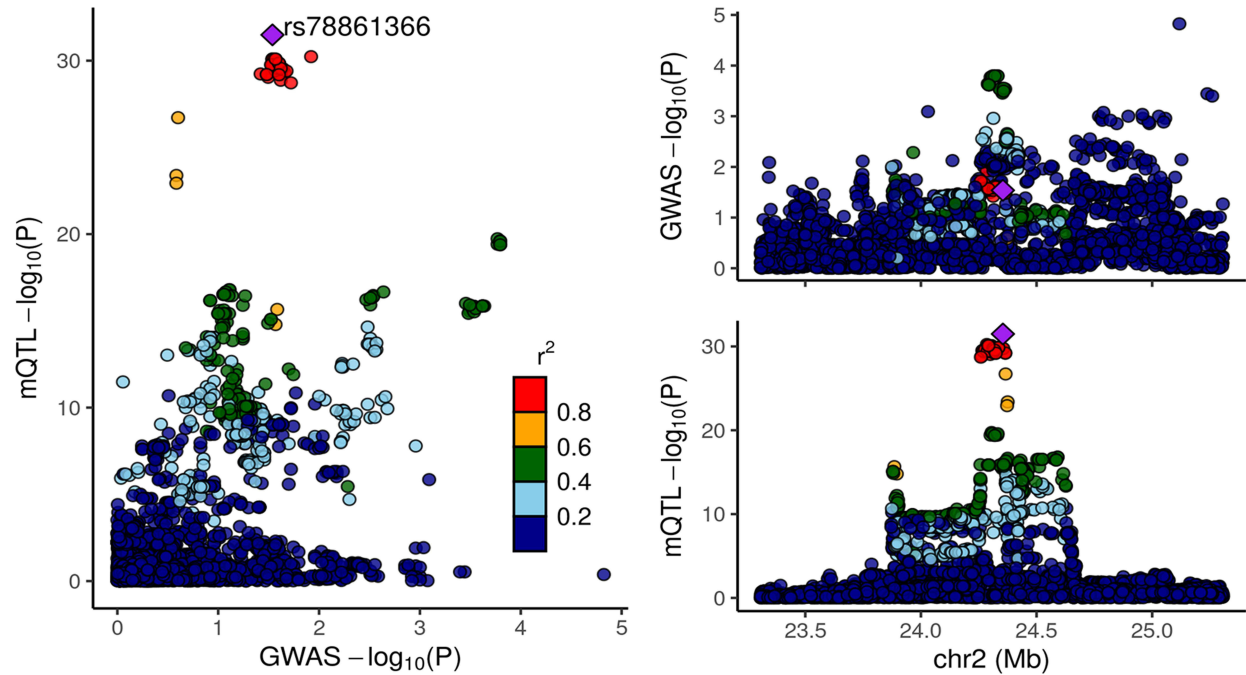
**Figure 2** SMR analyses of the causal effects of senescence-related genes mQTL on NAFLD. (**A**). Forest plot depicting the association between representative gene methylation and NAFLD. *Indicated causal associations supported by colocalization evidence. (**B**) Locus comparison plots between a representative gene (*TP53I3*) methylation loci and NAFLD. The scatter plot compares -$\log_{10}(p)$ values from GWAS (x-axis) and mQTL (y-axis) analyses. Each point represents a SNP, with color indicating linkage disequilibrium with the lead SNP (highlighted in purple).

Among the identified proteins, only TIGAR was associated with NAFLD in the replication cohort (FinnGen) (Supplementary Tables 12–14).

## Tissue-Specific Validation

We further explored the causal associations between gene expression and NAFLD in the liver tissues. The expression of *ENDOG* in the liver tissues was negatively associated with NAFLD (OR = 0.98, 95% CI = 0.97–1), which was consistent with the protective role suggested in the SMR analysis. The detailed information regarding the association between identified genes with NAFLD in the liver tissues was provided in Supplementary Table 15.
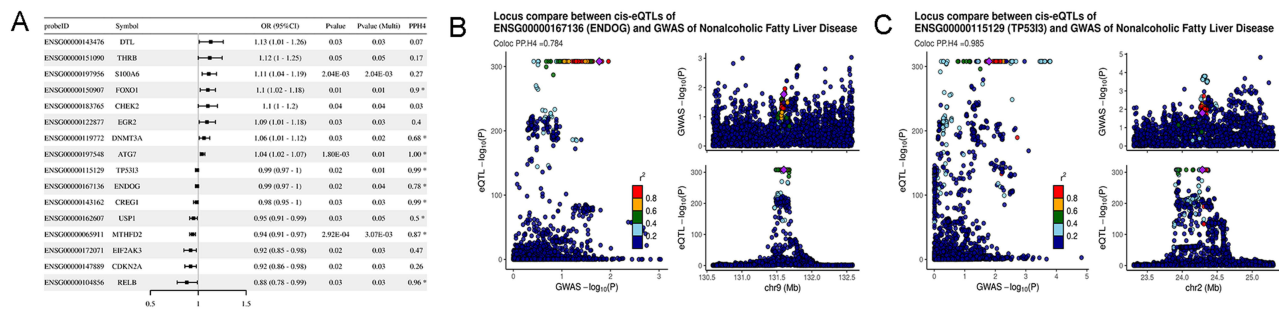
**Figure 3** SMR analyses of the causal effects of senescence-related genes eQTL on NAFLD. (**A**) Forest plot depicting the association between representative gene expressions and NAFLD. *Indicated causal associations supported by colocalization evidence. Locus comparison plots between (**B**) *ENDOG* and (**C**) *TP53I3* expression and NAFLD. The scatter plot compares -log₁₀(p) values from GWAS (x-axis) and eQTL (y-axis) analyses. Each point represents a SNP, with color indicating linkage disequilibrium with the lead SNP (highlighted in purple).
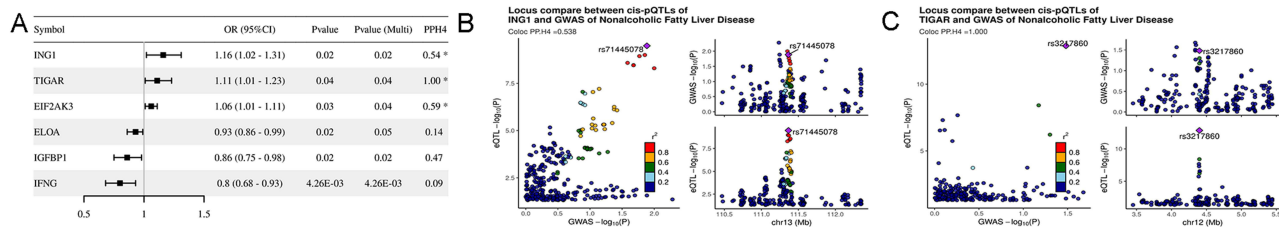


**Figure 4** SMR analyses of the causal effects of senescence-related protein abundance on NAFLD. (**A**) Forest plot depicting the association between representative protein abundance and NAFLD. *Indicated causal associations supported by colocalization evidence. Locus comparison plots between the level of (**B**) ING1 and (**C**) TIGAR and NAFLD. The scatter plot compares -log₁₀(p) values from GWAS (x-axis) and pQTL (y-axis) analyses. Each point represents a SNP, with color indicating linkage disequilibrium with the lead SNP (highlighted in purple).

## Multi-Omics Data Integration

By integrating blood mQTL and eQTL data, we performed SMR with the methylation loci of the common genes in mQTL-GWAS and eQTL-GWAS results as the exposure and the expressions of these genes as the outcome. At a stringent criteria (P-SMR < 0.05, multi-SNP-based P-value < 0.05 and P-HEIDI > 0.05), *S100A6* methylation at cg24155129 (OR = 0.6, 95% CI = 0.49–0.73) and cg01910639 (OR = 1.35, 95% CI = 1.24–1.47) were associated with a decreased and increased expression of *S100A6* respectively (Table 1). The detailed integrated associations were provided in Supplementary Table 16.

We did not identify common proteins between intersecting genes between mQTL and eQTL, and pQTL-GWAS results. Therefore, no SMR analysis was performed with the eQTL as the exposure and the pQTL as the outcome.

Integrating the multi-omics level evidence, we found that *S100A6* may be causally associated with NAFLD. In particular, the methylation site cg01910639 showed a positive correlation with NAFLD risk and positively regulated *S100A6* gene expression, which was positively associated with NAFLD risk. Additionally, cg24155129, which was also negatively correlated with NAFLD risk, negatively regulated *S100A6* expression. Therefore, we propose that the higher

**Table 1** Causal Effects of the Senescence-Related Gene Methylation on Gene Expression

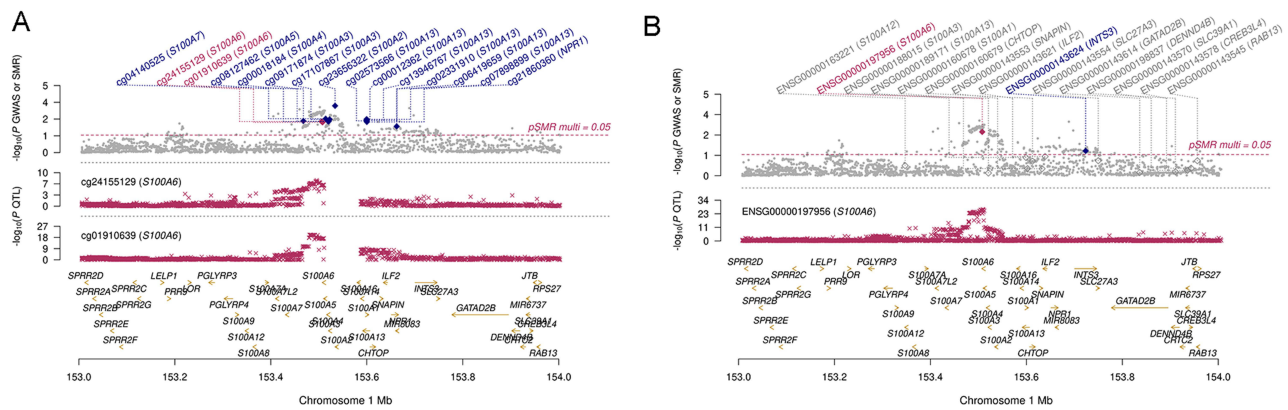| Expo ID | Outco Gene | p SMR | p SMR multi | p HEIDI | OR SMR (95% CI) |
|---|---|---|---|---|---|
| cg24155129 | S100A6 | 8.34E-07 | 8.34E-07 | 0.115 | 0.6 (0.49–0.73) |
| cg01910639 | S100A6 | 6.02E-12 | 4.85E-10 | 0.100 | 1.35 (1.24–1.47) |
| cg14273083 | TP53I3 | 0.006 | 0.006 | 1.38E-15 | 1.06 (1.02–1.11) |
| cg13630871 | ENDOG | 2.87E-22 | 2.87E-22 | 2.22E-05 | 0.3 (0.24–0.39) |

**Figure 5** Locus plots showing (**A**) *S100A6* methylation and (**B**) *S100A6*, their locations within the chromosome (lower panel). The Y-axis indicated the negative log of the p-values instrumental in deeming this locus significant in the SMR analysis.
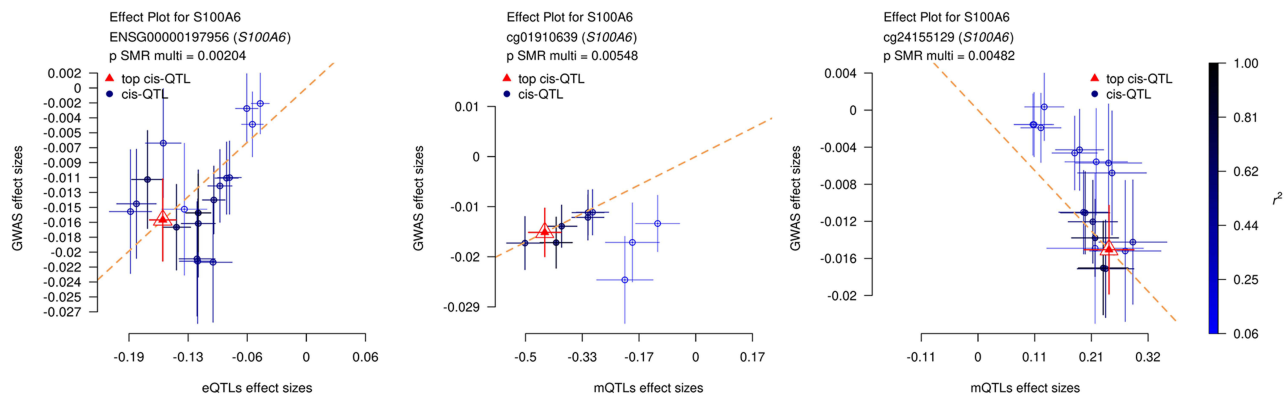


**Figure 6** SMR effect plots for (**A**) *S100A6*, (**B**) methylation site cg01910639 and cg24155129, and their associations with NAFLD. *cis*-QTLs were marked by blue dots, while top *cis*-QTLs were highlighted in red triangles.

methylation levels at cg20552903 and lower methylation levels at cg24155129 upregulates *S100A6* gene expression, leading to an increased risk of NAFLD.

To visualize the results of our SMR analysis, we created locus plots for *S100A6* methylation, expression and NAFLD (Figure 5A and B). Furthermore, we also provided the effect plots confirming the effects between *S100A6* methylation and expression and NAFLD (Figure 6).

## Validation of Candidate Genes in Mouse and Cell Models of NAFLD

To validate the findings from the analysis above, we conducted experiments using both mouse and cell models of NAFLD. We assessed the expression levels of *S100A6, ENDOG* and *TP53I3* in cell cultures (normal and steatotic). Oil Red O staining revealed substantial lipid accumulation in the NAFLD group cells, marked by an increased number of fat droplets (Figure 7A). qRT-PCR analysis of mRNA levels showed a significant rise in the expression of *S100A6* and *TP53I3*, and lower expression of *ENDOG* in the NAFLD group compared to the control group (Figures 7B).

In the animal model, we specifically focused on *S100A6* due to multi-omics evidence suggesting that methylation at cg20552903 and/or cg24155129 might regulate its expression. H&E and Masson staining suggested hepatic steatosis in the NAFLD group (Figure 7C). AST and ALT levels were significantly higher in the NAFLD group than in the control group (Figures 7D), indicating successful establishment of the NAFLD model. qRT-PCR measurements revealed that the expression levels of *S100A6* and ENDOG were significantly different in the NAFLD group compared to the control group (Figures 7E), suggesting their potential regulatory role in NAFLD development.
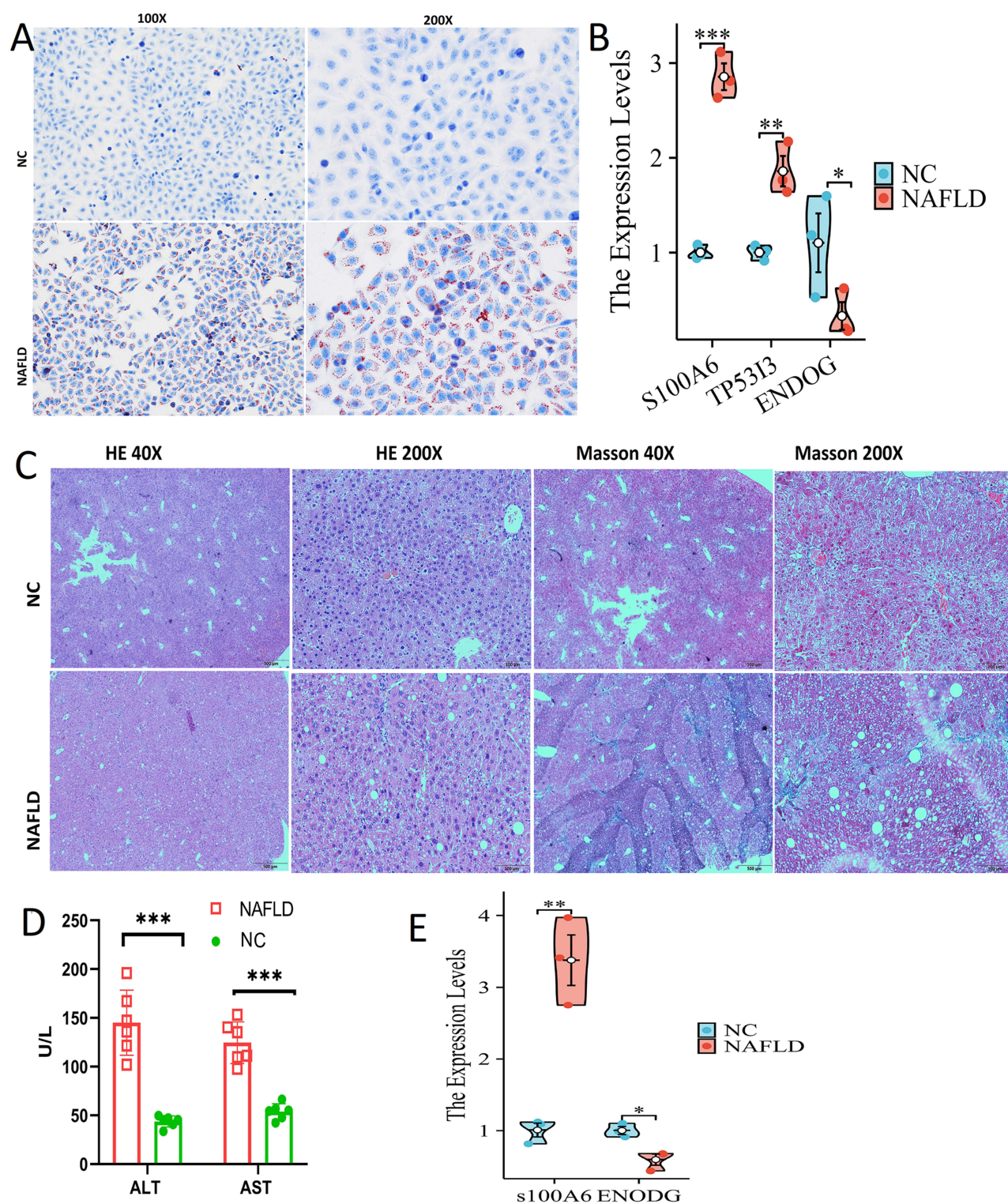
**Figure 7** Expression of the Key Genes in a Cell and Mouse NAFLD Model. The NAFLD mouse model was generated in C57BL/6J mice. Pair-fed mice were used as controls. Serum and liver tissues were collected on the 16 weeks for further analysis. (**A**) Oil Red O staining. (**B**) The relative mRNA expression of *S100A6, ENDOG* and *TP53I3* in cell NAFLD model was verified by qRT–PCR. (**C**) HE and Masson staining. (**D**) Serum alanine aminotransferase (ALT) and aspartate aminotransferase (AST) levels. (**E**) The relative mRNA expression of the S100A6 in mouse NAFLD model was verified by qRT–PCR. N = 4 in mouse model and N = 3 in Cell model, *p < 0.05, **p < 0.01, ***p < 0.001.

# Discussion

In this study, we systematically investigated the causal relationships between the methylation, gene expression and protein abundance of senscence-related genes and NAFLD using a multi-omics approach and SMR analysis. We chose to use the term NAFLD, given the ongoing transition to MASLD terminology, to maintain consistency with historical datasets and clinical contexts. Integrated multi-omics evidence from blood mQTL and eQTL SMR analysis revealed 3 genes (S100A6, *ENDOG* and *TP53I3*) as potential causal genes associated with NAFLD. And we further confirmed these findings by validation in mouse and cell models of NAFLD.

At the mQTL and eQTL levels, *S100A6* was found to be a potential risk factor for NAFLD. *S100A6*, also referred to as calcyclin, encodes a protein belonging to the S100 family and is integral to the regulation of cellular senescence. This gene has been shown to have an inhibitory effect on senescence-like changes in various cell types.[24] Its deficiency has been shown to induce morphological and biochemical features that are characteristic of cellular senescence.[25] Recently, there has been an ongoing research into the role of *S100A6* in NAFLD. A recent study has identified a significant relationship between the liver-derived protein *S100A6* and the progression of NAFLD.[26] Elevated serum levels of *S100A6* were observed in both human patients with NAFLD and in a high-fat diet-induced mouse model, correlating negatively with β-cell insulin secretory capacity. Depletion of hepatic *S100A6* in mice improved glycemia, suggesting a contributory role of *S100A6* in the pathophysiology of diabetes associated with NAFLD. Additionally, a review by Delangre et al highlighted that the aberrant activity of S100 isoforms, including S100A6, contributes to the dysregulation of lipid metabolism leading to hepatic steatosis and insulin resistance (IR), which are hallmarks of NAFLD.[27] While the exact mechanisms are not fully elucidated, it was suggested that S100 proteins may influence cell proliferation, apoptosis, migration, and inflammation, which are all relevant to the pathophysiology of NAFLD. In our study, we discovered that higher levels of *S100A6* might be associated with an increased risk of developing NAFLD, possibly by the dysregulation of lipid metabolism and promotion of hepatic steatosis. Furthermore, our findings propose a novel avenue for therapeutic intervention, where modulating *S100A6* expression or its regulatory pathways could be explored as a strategy to slow or halt disease progression in NAFLD patients. Additional research is required to fully understand the complex role of *S100A6* in hepatic health and disease, and to determine whether diminishing its effects could offer a viable treatment approach for those at risk of NAFLD.

In addition to *S100A6, ENDOG* was demonstrated to be a protective factor for NAFLD. *ENDOG* is a gene that encodes the mitochondrial protein Endonuclease G, a crucial enzyme involved in various cellular processes, particularly apoptosis and DNA metabolism. In the context of NAFLD, research has uncovered that *ENDOG* promotes NAFLD development via regulating the expression of lipid synthesis-associated genes like ACC1, ACC2, and FAS.[28] Loss of *ENDOG* was found to repress high-fat diet-induced liver lipid accumulation.[28] Therefore, targeting *ENDOG* could be a potential therapeutic approach for NAFLD. However, our study proposed the opposite, in which *ENODG* expression was negatively associated with NAFLD incidence. The controversy between *ENDOG* and NAFLD could be due to the multifactorial and dynamic nature of *ENDOG* in NAFLD pathogenesis. Additionally, the role of *ENDOG* might be context-dependent, with its expression and activity influenced by various environmental and genetic factors that could alter its function from protective to pathogenic, underscoring the complexity of its involvement in NAFLD.

*TP53I3*, also known as tumor protein p53 inducible protein 3, functions as a quinone oxidoreductase, which is involved in cellular redox reactions. Due to its role in apoptosis and stress responses, *TP53I3* has been implicated in cancer research.[29] However, no direct evidence about *TP53I3* in NAFLD has been presented. In this study, we demonstrated that *TP53I3* expression was negatively associated with the incidence of NAFLD, suggesting it as a potential protective factor. We could postulate that *TP53I3* is involved in the generation of ROS and participates in p53-mediated cell death pathways associated with NAFLD progression.

By integrating multi-omics analysis of mQTL and eQTL, we uncovered a potential regulatory axis in NAFLD pathogenesis: DNA methylation at specific loci suppresses *S100A6* gene expression, reducing S100A6 protein levels and decreasing the susceptibility to NAFLD. This opens up new avenues for therapeutic intervention in NAFLD, such as targeting this regulatory axis to modulate gene expression. Potential interventions might include the use of methylating agents or therapies to reduce *S100A6* expression. Additionally, the *S100A6* methylation-*S100A6* axis could serve as

a biomarker for early detection, prognosis, and monitoring of therapeutic responses in NAFLD patients, thereby enhancing personalized clinical care.

This study represents the first evaluation of the associations between senescence-related genes and NAFLD using SMR and colocalization. The main strength of this study is its use of SMR, allowing simultaneous assessment of the associations between methylation, expression, and protein abundance of senescence-related genes and NAFLD in independent European populations. Additionally, colocalization approaches effectively eliminate potential bias caused by linkage disequilibrium. Additionally, GWAS datasets with large sample sizes increased the statistical power of our study. Nonetheless, some limitations have to be addressed. First, due to the limited number of senescence-related proteins in the pQTL dataset, the current study did not fully explore the causal relationship between senescence protein abundance and the risk of NAFLD. Second, the exclusive use of *cis*-QTLs in SMR analysis may limit the comprehensiveness of the identified genetic associations and overlook long-range regulatory effects relevant to NAFLD pathogenesis. Third, SMR also has limited ability to exclude horizontal pleiotropy, where a gene affects disease through pathways independent of expression. Fourth, the tissue-specific nature of eQTL/mQTL associations means that the relevance of the selected QTL tissues to the disease-affected tissues directly impacts the reliability of the findings. Fifth, conclusions should be treated with caution when extending to other populations, as this study was based solely on European ancestry. Lastly, the findings from SMR analysis, while valuable for identifying potential causal associations, may not fully reflect clinical observations. SMR relies on genetic data and statistical models, which may not capture the full complexity of biological pathways or the influence of environmental factors on NAFLD. Additionally, SMR reflects the lifelong exposure effects associated with genetic variants, which may differ from the short-term effects of interventions or environmental exposures. Therefore, the results need to be contextualized with observational or clinical studies to better understand their relevance and applicability in clinical settings.

## Conclusions

Our findings suggest potential causal relationships between senescence-related gene methylation, expression, and protein abundance and NAFLD, with *S100A6, ENDOG* and *TP53I3* emerging as notable candidates in NAFLD pathogenesis. These findings provide a foundation for future research endeavors and clinical applications, but further investigations are needed to confirm these associations and their therapeutic implications.

## Abbreviations

GWAS, genome-wide association study; HEIDI, heterogeneity independent instruments; HEIDI, heterogeneity in the dependent instrument; HL-7702, Human Liver-7702; HE, hematoxylin and eosin; IVs, instrumental variables; MR, Mendelian randomization; NAFLD, Nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; PPH4, posterior probability of H4; QTL, quantitative trait locus; qRT-PCR, Quantitative reverse transcription-polymerase chain reaction; SMR, summary-data Mendelian randomization; SASP, senescence-associated secretory phenotype.

## Data Sharing Statement

The GWAS summary statistics for NAFLD can be accessed via the FinnGen and GWAS Catalog under the search term of GCST90275041 and GCST008469. The QTLs data for senescence-related genes can be obtained via CellAge.

## Ethics Approval and Consent to Participate

According to Item 1 and 2 of Article 32 of "the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects", this study is exempt from ethical review and approval, as it utilized summary statistics from public GWAS studies. All animal experiments received approval from the Institutional Animal Care and Use Committee of Guilin Medical University (GLMC-IACUC-20241090). All animal experiments strictly adhered to the National Standards for Laboratory Animal Welfare issued by the Chinese government (GB/T 35892-2018) and the Guide for the Care and Use of Laboratory Animals (National Research Council, 8th Edition, 2011).

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

## Disclosure

The authors declare that they have no competing interests.

## References

1. Riazi K, Azhari H, Charette JH, et al. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol*. 2022;7(9):851–861. doi:10.1016/S2468-1253(22)00165-0
2. Fan X, Song Y, Zhao J. Evolving liver disease insights from NAFLD to MASLD. *Trend Endocrinol Metabol*. 2024;35(8):683–686. doi:10.1016/j.tem.2024.02.012
3. Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease. *Lancet*. 2021;397(10290):2212–2224. doi:10.1016/S0140-6736(20)32511-3
4. Yuan S, Chen J, Li X, et al. Lifestyle and metabolic factors for nonalcoholic fatty liver disease: Mendelian randomization study. *Eur J Epidemiol*. 2022;37(7):723–733. doi:10.1007/s10654-022-00868-3
5. Ogrodnik M, Miwa S, Tchkonia T, et al. Cellular senescence drives age-dependent hepatic steatosis. *Nat Commun*. 2017;8:15691. doi:10.1038/ncomms15691
6. Serrano M, Barzilai N. Targeting senescence. *Nature Med*. 2018;24(8):1092–1094. doi:10.1038/s41591-018-0141-4
7. Giannakoulis VG, Dubovan P, Papoutsi E, Kataki A, Koskinas J. Senescence in HBV-, HCV- and NAFLD- mediated hepatocellular carcinoma and senotherapeutics: current evidence and future perspective. *Cancers*. 2021;13(18):4732. doi:10.3390/cancers13184732
8. Meadows V, Baiocchi L, Kundu D, et al. Biliary epithelial senescence in liver disease: there will be SASP. *Front Mol Biosci*. 2021;8:803098. doi:10.3389/fmolb.2021.803098
9. Nguyen PT, Kanno K, Pham QT, et al. Senescent hepatic stellate cells caused by deoxycholic acid modulates malignant behavior of hepatocellular carcinoma. *J Cancer Res Clin Oncol*. 2020;146(12):3255–3268. doi:10.1007/s00432-020-03374-9
10. Irvine KM, Skoien R, Bokil NJ, et al. Senescent human hepatocytes express a unique secretory phenotype and promote macrophage migration. *World J Gastroenterol*. 2014;20(47):17851–17862. doi:10.3748/wjg.v20.i47.17851
11. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. *Nat Rev Method Primers*. 2022;2. doi:10.1038/s43586-021-00092-5
12. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018;362:k601. doi:10.1136/bmj.k601
13. Verduijn M, Siegerink B, Jager KJ, Zoccali C, Dekker FW. Mendelian randomization: use of genetics to enable causal inference in observational studies. *Nephrology Dialysis Trans*. 2010;25(5):1394–1398. doi:10.1093/ndt/gfq098
14. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481–487. doi:10.1038/ng.3538
15. Wu Y, Zeng J, Zhang F, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun*. 2018;9(1):918. doi:10.1038/s41467-018-03371-0
16. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ*. 2021;375:n2233. doi:10.1136/bmj.n2233
17. Sun Z, Pan X, Tian A, et al. Genetic variants in HFE are associated with non-alcoholic fatty liver disease in lean individuals. *JHEP Reports*. 2023;5(7):100744. doi:10.1016/j.jhepr.2023.100744
18. Namjou B, Lingren T, Huang Y, et al. GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE Network. *BMC Med*. 2019;17(1):135. doi:10.1186/s12916-019-1364-z
19. Liu Y, Li B, Ma Y, Huang Y, Ouyang F, Liu Q. Mendelian randomization integrating GWAS, eQTL, and mQTL data identified genes pleiotropically associated with atrial fibrillation. *Front Cardiovascular Med*. 2021;8:745757. doi:10.3389/fcvm.2021.745757
20. Giambartolomei C, Zhenli Liu J, Zhang W, et al. A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics*. 2018;34(15):2538–2545. doi:10.1093/bioinformatics/bty147
21. Chen BY, Bone WP, Lorenz K, Levin M, Ritchie MD, Voight BF. ColocQuiaL: a QTL-GWAS colocalization pipeline. *Bioinformatics*. 2022;38(18):4409–4411. doi:10.1093/bioinformatics/btac512
22. Hormozdiari F, van de Bunt M, Segrè AV, et al. Colocalization of GWAS and eQTL signals detects target genes. *Am J Hum Genet*. 2016;99(6):1245–1260. doi:10.1016/j.ajhg.2016.10.003
23. Li X, Liang Z. Causal effect of gut microbiota on pancreatic cancer: a Mendelian randomization and colocalization study. *J Cell & Mol Med*. 2024;28(8):e18255. doi:10.1111/jcmm.18255

24. Zimmermann S, Biniossek ML, Maurer C, et al. Proteomic profiling in distinct cellular compartments of tumor cells reveals p53-dependent upregulation of S100A6 upon induction of telomere dysfunction. *Proteomics*. 2009;9(22):5175–5187. doi:10.1002/pmic.200900232

25. Słomnicki LP, Leśniak W. S100A6 (calcyclin) deficiency induces senescence-like changes in cell cycle, morphology and functional characteristics of mouse NIH 3T3 fibroblasts. *J Cell Biochem*. 2010;109(3):576–584. doi:10.1002/jcb.22434

26. Dogra S, Das D, Maity SK, et al. Liver-derived S100A6 propels β-Cell dysfunction in NAFLD. *Diabetes*. 2022;71(11):2284–2296. doi:10.2337/db22-0056

27. Delangre E, Oppliger E, Berkcan S, Gjorgjieva M, Correia de Sousa M, Foti M. S100 proteins in fatty liver disease and hepatocellular carcinoma. *Int J Mol Sci*. 2022;23(19):11030. doi:10.3390/ijms231911030

28. Wang W, Tan J, Liu X, et al. Cytoplasmic Endonuclease G promotes nonalcoholic fatty liver disease via mTORC2-AKT-ACLY and endoplasmic reticulum stress. *Nat Commun*. 2023;14(1):6201. doi:10.1038/s41467-023-41757-x

29. Chaudhry SR, Lopes J, Levin NK, Kalpage H, Tainsky MA. Germline mutations in apoptosis pathway genes in ovarian cancer; the functional role of a TP53I3 (PIG3) variant in ROS production and DNA repair. *Cell Death Discovery*. 2021;7(1):62. doi:10.1038/s41420-021-00442-y