BMC Genomics

RESEARCH

Open Access

Genome-wide enhancer-gene regulatory maps of liver reveal novel regulatory mechanisms underlying NAFLD pathogenesis

Ruofan Li[1,2†], Kaiyan Su[3†], Tianzhun Wu[4†], Li Xu[5†], Wenyu Song[6], Dandan Sun[7], Tao Zeng[1,2], Jinzhang Chen[3*], Haibei Xin[8*], Yuanfeng Li[6*], Mengya Zang[3*] and Minggen Hu[1,2*]

Abstract

Introduction Non-alcoholic fatty liver disease (NAFLD) represents the most widespread liver disease globally, ranging from non-alcoholic fatty liver (NAFL) and steatohepatitis (NASH) to fibrosis/cirrhosis, with potential progression to hepatocellular carcinoma (HCC). Genome-wide association studies (GWASs) have identified several single nucleotide polymorphisms (SNPs) associated with NAFLD. However, numerous GWAS signals associated with NAFLD locate in non-coding regions, posing a challenge for interpreting their functional annotation.

Results In this study, we utilized the Activity-by-Contact (ABC) model to construct the enhancer-gene maps of liver by integrating epigenomic data from 15 liver tissues and cell lines. We constructed the most comprehensive genome-wide regulatory maps of the liver, identifying 543,486 enhancer-gene connections, including 267,857 enhancers and 16,872 target genes. Enrichment analyses revealed that the ABC SNPs are significantly enriched in active chromatin regions and active chromatin state. By combining the ABC regulatory maps and NAFLD GWAS data, we systematically identified ABC SNPs associated with NAFLD risk. Through the functional annotations, such as pathway enrichment and drug-gene interaction analyses, we identified 6 genes (*GGT1*, *ACTG1*, *SPP1*, *EPHA2*, *PROZ* and *SHMT1*) as candidate NAFLD genes, with *SHMT1* previously reported. Among the SNPs connected to the candidate genes, the ABC SNP *rs2017869* (odds ratio [OR] for the C allele = 1.10, 95% CI = 1.04–1.16, $P = 5.97 \times 10^{-4}$) had the highest ABC score. According to the ABC maps, *rs2017869* links to *GGT1*, and several drugs targeting this gene, such as liothyronine,

Introduction

Non-alcoholic fatty liver disease (NAFLD), recently redefined as metabolic dysfunction-associated steatotic liver disease (MASLD) [1], is the most common liver disorder worldwide, affecting approximately 24% of the population [2]. Meanwhile, its incidence is rising at an alarming and concerning pace [3]. NAFLD encompasses a disease spectrum progressing from non-alcoholic fatty liver (NAFL) and non-alcoholic steatohepatitis (NASH) to fibrosis/cirrhosis, with the potential to ultimately develop into hepatocellular carcinoma (HCC) [2, 4]. Compelling evidence also establishes connections between NAFLD

†Ruofan Li, Kaiyan Su, Tianzhun Wu and Li Xu contributed equally to this work.

*Correspondence:
Jinzhang Chen
chenjinzhang@smu.edu.cn
Haibei Xin
xinhaibei@qq.com
Yuanfeng Li
liyf_snp@163.com
Mengya Zang
zangmy@163.com
Minggen Hu
hmg301@126.com

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

showed potential benefits to patients with NAFLD. Furthermore, we identified that another novel gene, *EPHA2*, may play a crucial role in NAFLD by regulating the GGT levels.

Conclusions Our study provides the most comprehensive ABC regulatory maps of the liver to date. This resource offers a valuable reference for identifying regulatory variants and prioritizing susceptibility genes of liver diseases, such as NAFLD.

Keywords Non-alcoholic fatty liver disease, Genome-wide association study, Single nucleotide polymorphism, *GGT1*

and several other chronic diseases, including cardiovascular disease (CVD) [5, 6], type 2 diabetes mellitus (T2DM) [7], dyslipidemia [8], and chronic kidney disease (CKD) [9]. Consequently, NAFLD not only shortens life expectancy [10], but also imposes a significant economic burden [11]. Despite its growing impact, effective treatments for NAFLD remain limited [12].

Identifying the risk factors for NAFLD is crucial for pinpointing patients who are most susceptible to increased morbidity and mortality, thereby guiding precision therapeutic strategies. Studies have shown that NAFLD arises from a combination of genetic predisposition and environmental factors [13]. Environmental contributors, such as obesity, type 2 diabetes, hypertension, dyslipidemia, and other emerging conditions, have been implicated in the development of NAFLD [14]. In addition, the significant role of genetic factors in NAFLD risk cannot be overlooked. Genome-wide association studies (GWASs) have identified several single nucleotide polymorphisms (SNPs) associated with NAFLD. Among them, four common SNPs are robustly associated with NAFLD development: *rs738409* (patatin-like phospholipase domain-containing 3, *PNPLA3*) [15, 16], *rs58542926* (transmembrane 6 superfamily member 2, *TM6SF2*) [17–19], *rs641738* (membrane-bound O-acyltransferase domain containing 7, *MBOAT7*) [20, 21], and *rs1260326* (glucokinase regulator, *GCKR*) [22, 23]. Additionally, GWASs have uncovered genetic variants implicated in NAFLD progression, such as *rs72613567* (hydroxysteroid 17-beta dehydrogenase 13, *HSD17B13*) [24] and *rs4374383* (MER proto-oncogene, tyrosine kinase, *MERTK*) [25]. Despite the abundance of GWAS findings, numerous NAFLD-associated SNPs in non-coding regions such as enhancers and promoters remain challenging to functionally annotate [26], potentially leaving some susceptibility genes undetected.

To gain a deeper understanding of the impacts of non-coding SNPs on gene expression and biological pathways, researchers have applied various molecular quantitative trait locus (xQTL) analyses, including expression QTL (eQTL) and splicing QTL (sQTL) studies. However, the existing xQTL datasets have explained only a small portion of the GWAS heritability associated with diseases [27, 28], indicating that additional and varied functional genomic data, extending beyond gene transcription, are required to comprehensively unravel the mechanisms underlying the disease.

Tremendous efforts have been made to link GWAS signals to different mechanisms of gene regulation. Approaches such as predicting enhancers utilizing histone chromatin immunoprecipitation sequencing (ChIP-seq) data enriched for H3K27 acetylation marks (H3K27ac) and estimating 3D genome interactions using high-throughput chromosome conformation capture (Hi-C) data have been widely adopted [29]. To improve predictions of enhancer-gene interactions, Joseph et al. developed the Activity-by-Contact (ABC) model, which identifies non-coding SNPs situated within ABC enhancers and their target genes [30]. This approach has been shown to outperform previous methods in predicting regulatory elements and their associated target genes [30, 31]. Despite the establishment of ABC enhancer-gene connections across various tissues and cell types, comprehensive enhancer-gene maps specific to the liver remain lacking. This gap poses a significant barrier to a more thorough exploration of NAFLD's regulatory mechanisms using this approach.

The objective of this study was to advance the understanding of the regulatory mechanisms underlying NAFLD pathogenesis through integrative analyses of genome-wide enhancer-gene maps, GWAS data, and eQTL data. Our study linked non-coding SNPs to NAFLD biological mechanisms and identified novel susceptibility genes. The most comprehensive enhancer-gene maps of the liver to date provide an essential resource for understanding gene regulation and the genetic basis of NAFLD and other liver diseases.

Methods

Epigenomic profiling of liver tissues and cell lines

To construct Activity-by-Contact (ABC) maps for liver tissues and cell lines, we curated published epigenomic data, including DNase I hypersensitive sites sequencing (DNase-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq), H3K27ac chromatin immunoprecipitation sequencing (H3K27ac ChIP-seq) and high-throughput chromosome conformation capture (Hi-C) data from ENCODE [32] and the Roadmap Epigenomics Project [33]. In total, we achieved data from 15 liver tissues and cell lines, and the sources of data for each

biosample are detailed in Table S1. First, we retrieved the BAM files for DNase-seq, ATAC-seq, H3K27ac ChIP-seq, and Hi-C from both ENCODE and the Roadmap Epigenomics Project (accessible at https://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/). For the BAM files sourced from ENCODE, we chose those that were aligned to the hg38 reference genome, marked as "released", and did not have flags indicating "unfiltered", "extremely low spot score", "extremely low read depth", "not compliant", or "insufficient read depth". These files were employed as the input data for the ABC model. Finally, considering previous research has demonstrated that averaged Hi-C data from multiple cell-type specific Hi-C matrices produces results that are comparable to those derived from cell-type specific promoter capture Hi-C data [30, 34], we downloaded the average Hi-C data (https://www.encodeproject.org/files/ENCFF134PUN/@@download/ENCFF134PUN.bed.gz) and used it in analyses for biosamples where cell-type specific Hi-C data was not available.

ABC model construction

To predict enhancer-gene connections in liver tissues and cell lines, we utilized the ABC model (https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction) by integrating data from chromatin accessibility assays (ATAC-seq or DNase-seq), histone modification profiles (H3K27ac ChIP-seq), and chromatin conformation capture (Hi-C) [30, 34].

Briefly, for each biosample, we constructed the ABC model by adhering to these steps using python (v3.10.13): (1) Peaks in the chromatin accessibility data were called using MACS2 with a lenient p-value cutoff of 0.1; (2) Chromatin accessibility reads were counted for each peak, and the top 150,000 peaks with the highest read counts were retained. Each peak was resized to 500 bp, centered on the peak summit. We also included 500 bp regions centered on all gene transcription start sites (TSSs) and excluded any peaks overlapping blacklisted regions. Overlapping peaks were merged, resulting in a final set of candidate regions; (3) Element activity was determined by counting the reads in each candidate region from both chromatin accessibility and H3K27ac ChIP-seq datasets, followed by computing the geometric mean of these two measurements; (4) The ABC score for each element-gene pair was computed as the normalized product of activity and contact, where normalization was achieved by dividing by the product of activity and contact for all other elements located within a 5 Mb region surrounding that gene [30]. To identify significant gene regulatory effects, the best threshold of ABC score was automatically chosen based on the input [34]. Element-gene pairs that surpassed this threshold were classified as "enhancer-gene connections", whereas elements that were predicted to regulate at least one gene were labeled as "ABC enhancers".

Genome-wide association study (GWAS) data collection

To investigate the associations between SNPs in ABC enhancers and NAFLD susceptibility, we searched for available GWAS data for NAFLD. We downloaded GWAS data from the FinnGen study, which included 3,006 cases and 450,727 controls (https://r11.finngen.fi/pheno/NAFLD). Based on the International Classification of Diseases Tenth Revision (ICD-10), participants assigned the code K76.0 for hospital discharge or cause of death were designated as NAFLD cases, and those without a NAFLD diagnosis were designated as controls. This GWAS data was used in the discovery stage in this study. We further obtained another large-scale GWAS statistics from the GWAS Catalog, comprising 4,761 cases and 373,227 controls (https://www.ebi.ac.uk/gwas/studies/GCST90054782). This study defined NAFLD as any hospital admission with an ICD-9 or ICD-10 code relating to NAFLD (571.5; K75.8, K76.0) or any primary care encounter with a Read code (clinical terminology system used in UK Primary Care settings) relating to NAFLD (C32y5, J6154, J61y1, J61y7, J61y8, J61y9). Controls were identified as individuals who did not have a diagnosis of NAFLD. This GWAS data was used in the replication stage. In total, this study included 7,767 cases and 823,954 controls. In this study, we retained the NAFLD nomenclature instead of the updated MASLD terminology to align with the ICD-9/10-based case definitions used in the GWAS data. All samples were of European descent.

Characteristics analyses of ABC SNPs

ABC SNPs were defined as those SNPs located in the predicted enhancer regions, as indicated by the ABC maps. By overlapping the SNP list from the Single Nucleotide Polymorphism Database (dbSNP, https://www.ncbi.nlm.nih.gov/snp/) (GRCh38.p7), which contains 660,146,174 SNPs, with the ABC maps, we identified 15,895,042 SNPs as ABC SNPs. For enrichment analyses, a set of control SNPs (non-ABC SNPs) was generated, matching the allele frequencies, linkage disequilibrium (LD) patterns, and genomic distribution of ABC SNPs using the web tool vSampler (http://www.mulinlab.org/vsampler/) [35].

For enrichment analyses of genomic distribution, we used SnpEff (v5.1) [36] to annotate both ABC and non-ABC SNPs. SNPs were categorized into the following genomic features: upstream gene, downstream gene, 5'UTR, 3'UTR, intron, intergenic region, and others. Enrichment analyses of these genomic annotations were conducted using a two-tailed Fisher's exact test, and the results were presented in a 2×2 contingency table. The table had columns representing ABC SNPs and non-ABC

SNPs, and rows distinguishing SNPs located within and outside the annotated genomic regions.

For enrichment analyses of ABC SNPs within functional annotations, we obtained ChIP-seq peak data for histone modifications, including H3K4 monomethylation marks (H3K4me1), H3K4 trimethylation marks (H3K4me3), H3K27 acetylation marks (H3K27ac), H3K9 acetylation marks (H3K9ac), H3K27 trimethylation marks (H3K27me3), H3K36 trimethylation marks (H3K36me3), and transcription factor binding sites (TFBSs) from the ENCODE portal (https://www.encodeproject.org). Using bedtools (v2.26.0), we identified overlaps between ABC or control SNPs and the peaks of regulatory elements. The core 15 chromatin state was downloaded from the Roadmap Epigenomics Project (https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state). Enrichment of ABC SNPs within these regulatory elements was assessed using a two-tailed Fisher's exact test, with a 2×2 contingency table. The table had columns representing ABC SNPs and non-ABC SNPs, and rows distinguishing SNPs located within and outside the regulatory elements.

For enrichment analyses of ABC SNPs among NAFLD-related GWAS loci, we procured summary statistics from the FinnGen study. Subsequently, GWAS loci were defined as genomic regions encompassing SNPs in LD with the index SNP, with an r^2 threshold of 0.2 or higher. Enrichment of ABC SNPs within these NAFLD-related GWAS loci was analyzed using two-tailed Fisher's exact test, presented in a 2×2 contingency table (columns: ABC SNPs and non-ABC SNPs; rows: SNPs within and not within the GWAS loci). Next, to estimate the heritability enrichment of ABC SNPs, we performed LD score regression (LDSC) using GWAS summary statistics. GWAS SNPs that also corresponded to ABC SNPs were extracted and used to generate a quantile-quantile (QQ) plot of the GWAS *P* values for those SNPs.

Pathway enrichment analyses

Pathway enrichment analyses were conducted utilizing Gene Ontology (GO) [37], Kyoto Encyclopedia of Genes and Genomes (KEGG) (http://www.genome.jp/kegg/) [38], Reactome Knowledgebase (https://reactome.org) [39], and WikiPathways (http://www.wikipathways.org) [40] via "g: Profiler" (https://biit.cs.ut.ee/gprofiler/) [41]. GO includes three complementary categories: biological process (BP), cellular component (CC), and molecular function (MF), providing curated and predicted gene annotations across multiple species. KEGG is a comprehensive repository that integrates genomic, chemical, and functional data. Reactome offers detailed molecular insights into a wide range of biological processes, both normal and disease-associated. WikiPathways serves as a resource for biological pathways, providing a platform for the publication and curation of biological knowledge.

Identification of novel candidate drugs for NAFLD

To uncover potential drugs that target the candidate genes, we employed the Drug-Gene Interaction Database (DGIdb) (https://dgidb.org) [42] to investigate drug-gene interactions. DGIdb integrates data from multiple sources to predict druggable genes. The relationships between the candidate gene targets and their associated drugs were visualized using Cytoscape 3.10.3.

Other methods

Expression quantitative trait locus (eQTL) analyses, colocalization analyses, and meta-analysis are described in Supplementary Methods.

Statistical analyses

Demographic characteristics between cases and controls were compared using the two-sided χ^2 test or Student's t-test, as appropriate. Enrichment analyses were conducted using a two-tailed Fisher's exact test with Bonferroni correction applied, and $P<0.05$ was considered statistically significant. The associations between genetic variants and NAFLD risk were assessed using unconditional multivariate logistic regression, with odds ratios (ORs) and 95% confidence intervals (CIs) calculated. $P<1.36\times10^{-5}$ (1/73,724 independent ABC SNPs in NAFLD GWAS data) was considered statistically significant. All statistical analyses were performed using R (v4.3.1) software.

Results

Landscape of the genome-wide enhancer-gene maps of liver

To construct genome-wide enhancer-gene maps across 15 liver tissues and cell lines, we computed ABC scores for each gene and chromatin accessible element, considering a window of 5 Mb around each gene. This was achieved by integrating data from enhancer accessibility assays (such as ATAC-seq or DNase-seq), enhancer activity marks (H3K27ac ChIP-seq), and normalized contact frequency data (Hi-C) (Fig. 1). The data sources are listed in Table S1. In total, we identified 543,486 enhancer-gene connections involving 267,857 ABC enhancers and 16,872 genes regulated by identified ABC enhancers (ABC genes) (Figs. 2A-C). The average number of enhancer-gene connections identified was 36,232, with a range varying from a minimum of 31,044 to a maximum of 51,237. On average, each ABC enhancer was estimated to regulate 3.0 genes, while each gene was regulated by 2.0 ABC enhancers (Fig. 2D and E). Additionally, the median genomic distance separating each enhancer-gene connection was found to be 33,629 bp (Fig. 2F). Of the

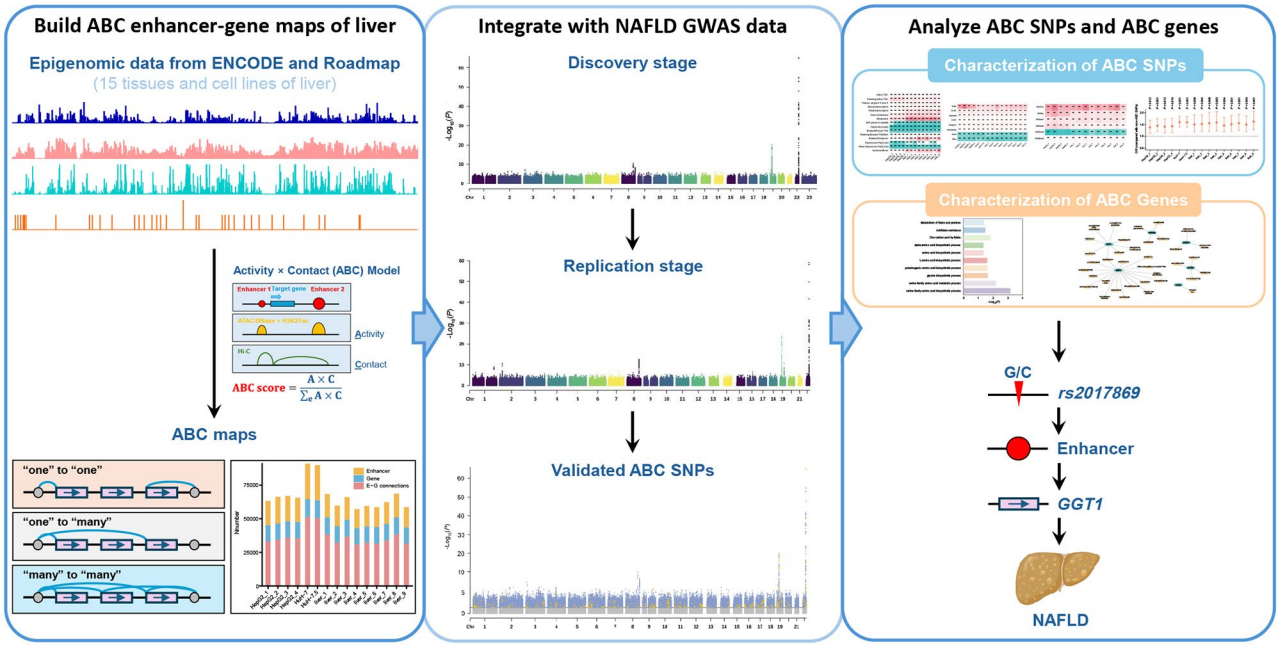

Fig. 1 Study overview. Left: Construction of the ABC enhancer-gene maps of liver. The epigenomic data, including ATAC-seq, DNase-seq, H3K27ac ChIP-seq and HiC-seq data were from 15 liver biosamples. The bar chart represents the characteristics of ABC maps. Middle: We integrated the constructed ABC regulatory maps of liver and the NAFLD GWASs. Manhattan plots show the genome-wide association statistics of the discovery and replication cohorts, and the validated ABC SNPs. Right: The characterization of ABC SNPs and target genes, with the graphic summary of connecting *rs2017869* within the non-coding region to NAFLD pathogenesis. Heatmaps show the characterization of ABC SNPs. Bar chart of pathway enrichment analyses and drug-gene interaction network represent the characterization of ABC genes. ABC, Activity-by-Contact; ATAC-seq, assay for transposase-accessible chromatin using sequencing; DNase-seq, DNase I hypersensitive sites sequencing; H3K27ac ChIP-seq, H3K27ac chromatin immunoprecipitation sequencing; Hi-C, high-throughput chromosome conformation capture; NAFLD, non-alcoholic fatty liver disease; GWAS, genome-wide association study; SNP, single nucleotide polymorphism; *GGT1*, gamma-glutamyltransferase 1

identified ABC genes, 690 (4.1%) were uniquely detected in their respective biosamples (Fig. 2G). Compared to constructing ABC maps from a single biosample (i.e., HepG2_1) [31], our multi-biosample analyses revealed an additional 607 ABC genes. In summary, we constructed the liver-specific regulatory maps based on multi-omics data from liver tissues and cell lines, providing a valuable resource for establishing connections between non-coding SNPs and their respective target genes.

Characterization of ABC SNPs

To describe the characteristics of SNPs located within ABC enhancers (ABC SNPs), we obtained the list of ABC SNPs by integrating the whole SNP list from dbSNP database, which comprises 660,146,174 SNPs, with the ABC maps. Compared to non-ABC SNPs, ABC SNPs were significantly enriched in TFBSs, 5'UTRs and upstream gene regions (Fig. 3A). We also examined whether ABC SNPs were enriched in genomic regions marked by histone modification to assess their potential regulatory function. As expected, ABC SNPs were significantly enriched in active chromatin regions, including H3K27 acetylation marks (H3K27ac), H3K9 acetylation marks (H3K9ac), H3K4 monomethylation marks (H3K4me1), H3K4 trimethylation marks (H3K4me3), and H3K36 trimethylation marks (H3K36me3), while being less enriched in repressive epigenetic marks like H3K27 trimethylation marks (H3K27me3) (Fig. 3B). Additionally, ABC SNPs were found to be enriched in active chromatin state, such as enhancers and active transcription start sites (TSSs) (Fig. 3C). Collectively, these findings offer compelling evidence that underscores the regulatory function of ABC SNPs.

We conducted an additional investigation to determine if ABC SNPs were enriched with susceptibility SNPs associated with NAFLD, utilizing GWAS summary statistics obtained from the FinnGen study, which included 3,006 NAFLD cases and 450,727 controls (Table S2). Our analyses revealed that ABC SNPs are significantly enriched in GWAS loci for NAFLD compared to non-ABC SNPs (Fig. 3D), suggesting that ABC SNPs may offer insights into NAFLD heritability. To quantify their contribution to NAFLD heritability, we used LD score regression (LDSC) and found that the heritability of NAFLD explained by ABC SNPs was 2.62% (SE = 0.0162), indicating a significant fraction (Fig. 3E). Furthermore, ABC SNPs showed stronger population-associated *P* values compared to the genome-wide SNPs (Fig. 3F). Together, these findings imply that ABC SNPs may have a notable role in the heritability of NAFLD.

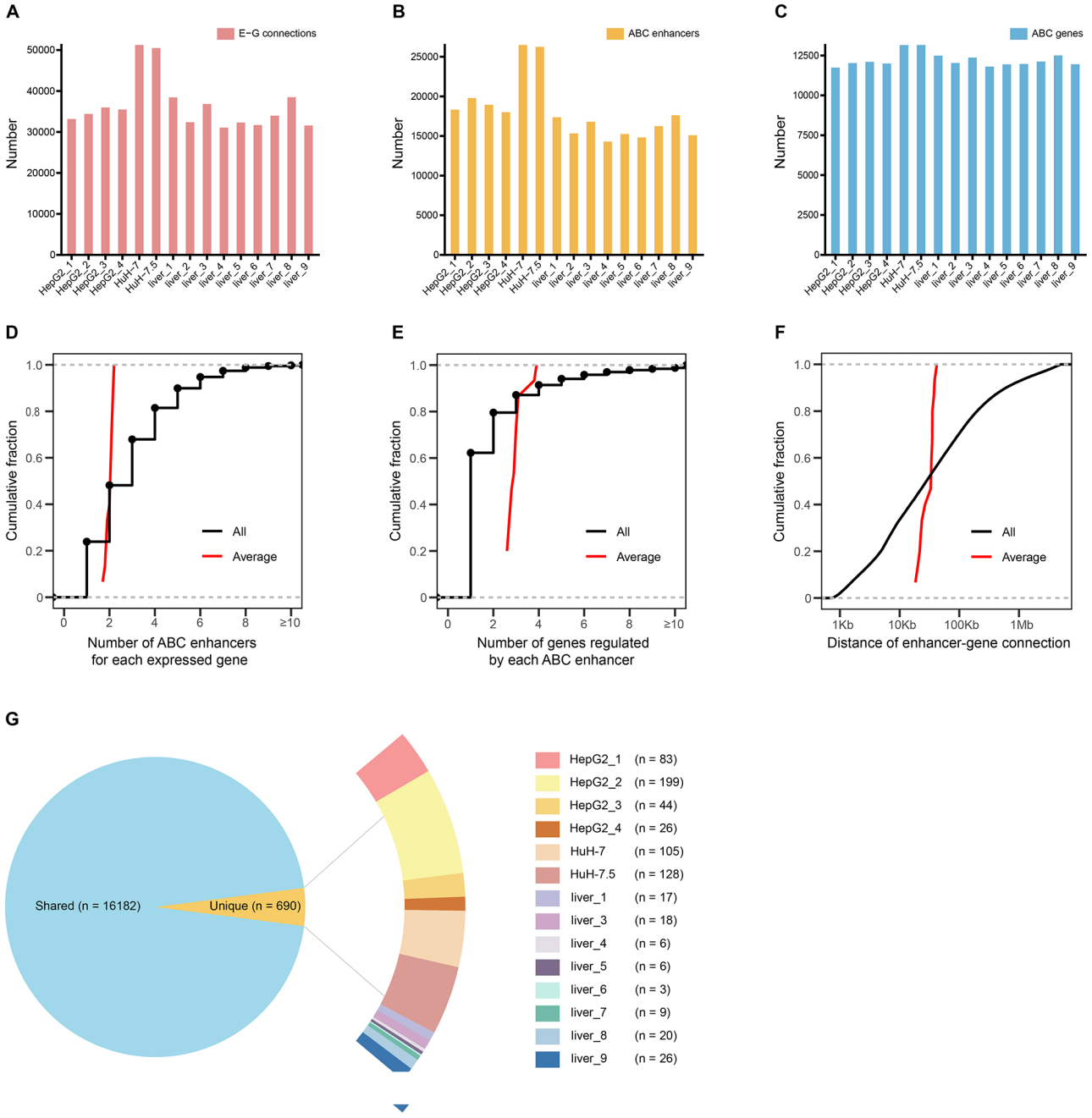

Fig. 2 Liver-specific genome-wide enhancer-gene maps landscape. (**A-C**) Bar charts represent the number of enhancer-gene connections (E-G connections) (**A**), ABC enhancers (**B**) and ABC genes (**C**) in each liver biosample. (**D**) Cumulative fractions of the number of enhancers predicted to regulate each gene in each liver biosample (black line; mean = 2.0) and the mean number of enhancers predicted to regulate each gene in each liver biosample (red line; median = 2.1). (**E**) Cumulative fractions of the number of genes regulated by each ABC enhancer in each liver biosample (black line; mean = 3.0) and the mean number of genes regulated by each ABC enhancer in each liver biosample (red line; median = 2.9). (**F**) Cumulative fractions of the genomic distances between the enhancer and the gene for each predicted enhancer-gene connection in each liver biosample (black line; median = 28,036 bp) and the median genomic distance between each enhancer-gene connection in each liver biosample (red line; median = 33,629 bp). (**G**) Among all identified ABC genes, 690 (4.1%) were uniquely detected in their respective biosamples. Compared to ABC maps constructed from a single liver biosample (e.g., HepG2_1), this analysis identified an additional 607 ABC genes. ABC, Activity-by-Contact

Characterization of ABC genes associated with NAFLD

To identify novel NAFLD-associated loci and link these loci to their target genes, we conducted a joint analysis of ABC SNPs with FinnGen GWAS data in the discovery stage, which included 3,006 cases and 450,727 controls (Table S2). We identified 25,691 ABC SNPs that demonstrated associations with NAFLD risk ($P < 0.05$). To confirm the associations between these ABC SNPs and the

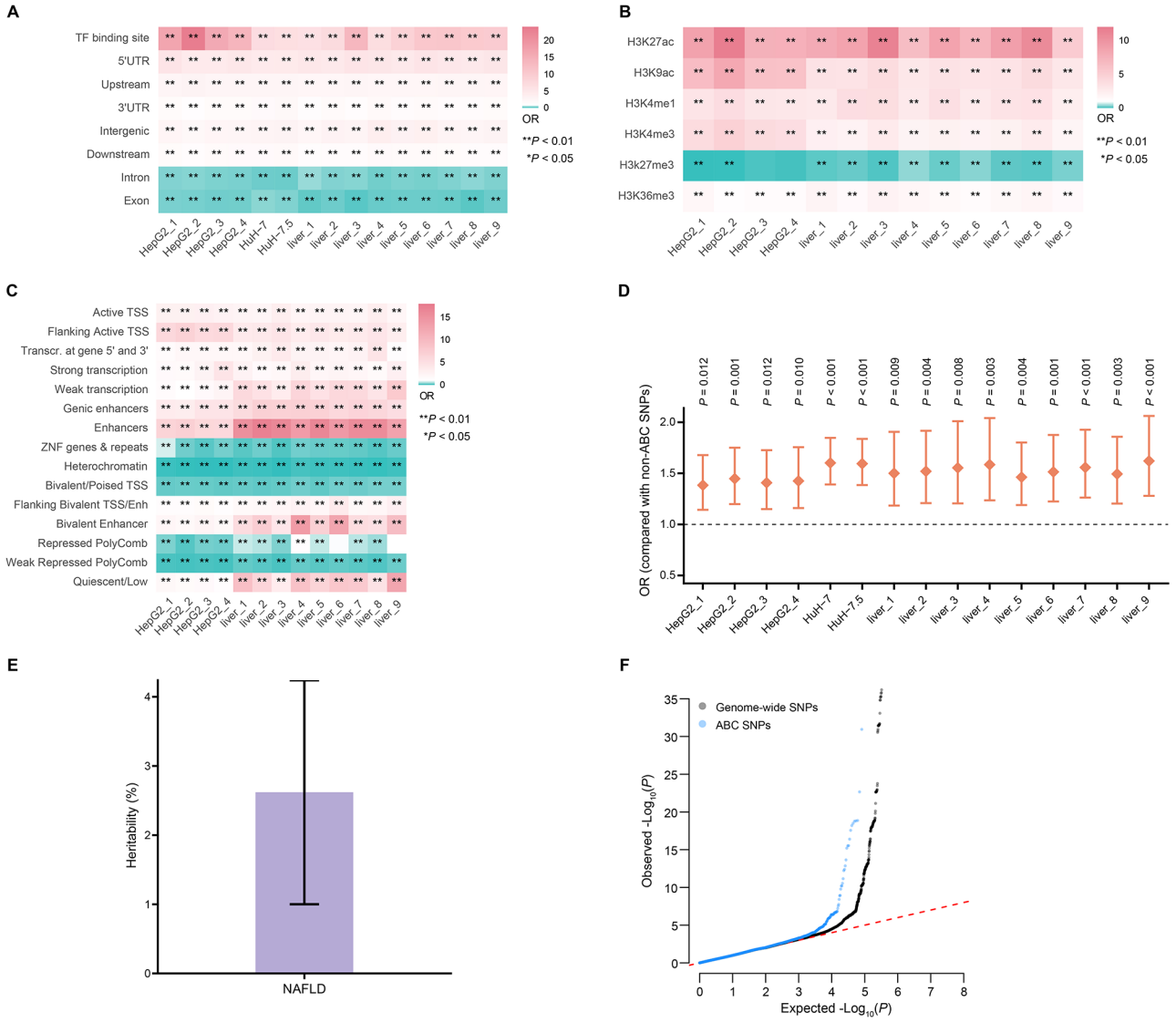

Fig. 3 Characterization of ABC SNPs. (**A**) Heatmap shows the genomic distribution of ABC SNPs compared with non-ABC SNPs. *P*-values were calculated by two-tailed Fisher's exact test. (**B**) Heatmap shows the histone modification enrichment of ABC SNPs in regulatory elements including H3K27ac, H3K9ac, H3K4me1, H3K4me3, H3K27me3, and H3K36me3, compared with non-ABC SNPs. *P*-values were calculated by two-tailed Fisher's exact test. (**C**) Heatmap shows the chromatin state enrichment of ABC SNPs compared with non-ABC SNPs. *P*-values were calculated by two-tailed Fisher's exact test. (**D**) Enrichment analyses of ABC SNPs in FinnGen NAFLD-related GWAS SNPs compared with non-ABC SNPs. *P*-values were calculated by two-tailed Fisher's exact test and bars indicate 95% CIs. (**E**) Proportion of GWAS heritability of NAFLD explained by ABC SNPs. The error bars represent standard error. (**F**) Quantile-quantile (QQ) plots of *P* values from GWAS of NAFLD. ABC SNPs were shown in comparison with genome-wide SNPs. ABC, Activity-by-Contact; SNP, single nucleotide polymorphism; H3K27ac, H3K27 acetylation marks; H3K9ac, H3K9 acetylation marks; H3K4me1, H3K4 monomethylation marks; H3K4me3, H3K4 trimethylation marks; H3K27me3, H3K27 trimethylation marks; H3K36me3, H3K36 trimethylation marks; NAFLD, non-alcoholic fatty liver disease; TSS, transcription start site; Transcr., transcription; Ehn, enhancers; GWAS, genome-wide association study; OR, odds ratio; CI, confidence interval

risk of NAFLD, we assessed their effects in an independent NAFLD cohort from European populations, comprising of 4,761 cases and 373,227 controls (Table S2). A total of 846 ABC SNPs were validated ($P < 0.05$ with the same direction of association in the discovery stage), which were connected to 1,181 ABC genes (Fig. 4A). Given that eQTL analysis is a well-established method for exploring the effects of genetic variants on gene expression [43], we further used the liver eQTL data from the Genotype-Tissue Expression Project (GTEx v8). Among the validated ABC SNPs, 81 SNPs were discovered to be eQTLs that are associated with expression of 27 genes (eGenes) (Table S3).

To investigate the possible role of these genes in the development of NAFLD, we examined their functional roles through pathway enrichment and drug-gene interaction analyses. Pathway enrichment analyses identified ten pathways significantly associated with NAFLD

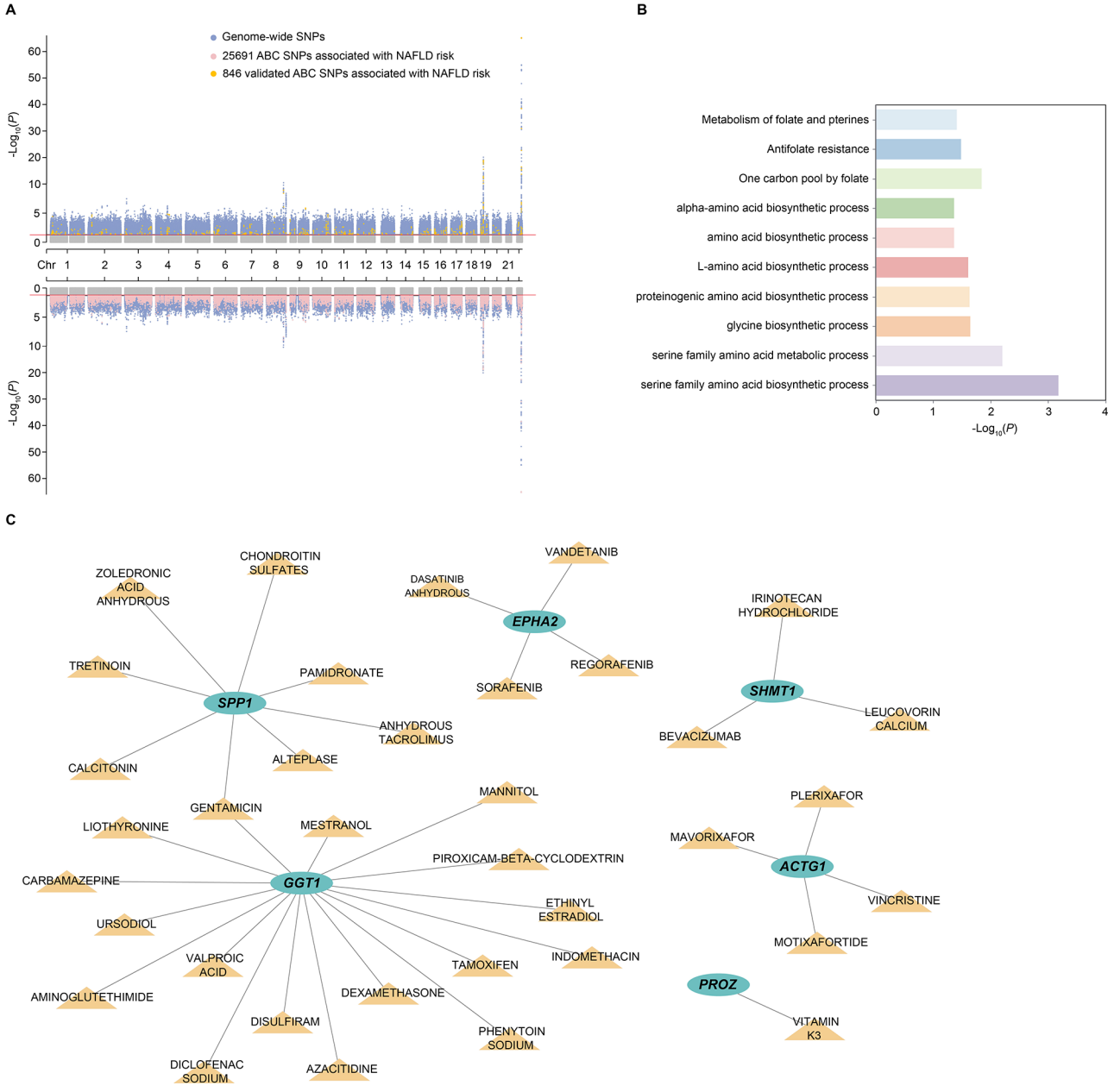

Fig. 4 Characterization of ABC genes associated with NAFLD. (**A**) Manhattan plots for the associations between ABC SNPs and NAFLD risk in NAFLD GWAS data. The grey dots represent the SNPs that are not associated with NAFLD risk ($P > 0.05$). The blue dots represent the genome-wide SNPs associated with NAFLD risk ($P < 0.05$). The pink dots represent the ABC SNPs associated with NAFLD risk ($P < 0.05$) in the discovery stage. The yellow dots represent the validated ABC SNPs in the replication stage ($P < 0.05$). The red line indicates the significance threshold of $P = 0.05$. The x-axis represents the genomic position (human genome assembly hg38), and the y-axis shows the $-\log_{10}(P)$. (**B**) Bar chart shows the results of pathways enrichment analyses. (**C**) Drug-gene interaction network. ABC, Activity-by-Contact; SNP, single nucleotide polymorphism; NAFLD, non-alcoholic fatty liver disease; GWAS, genome-wide association study

(adjusted $P < 0.05$; Fig. 4B). These pathways were mainly related to the metabolism of one-carbon units, amino acid and folate. Several of these pathways have been implicated in NAFLD. For example, the serine and glycine biosynthetic processes supply one-carbon units essential for one-carbon metabolism [44], which is strongly associated with NAFLD, with the liver serving as a primary site for one-carbon metabolism [45]. Additionally, the folate cycle, a key component of one-carbon metabolism, supports the synthesis of porphyrins, thymidine, purines, glutathione and S-adenosylmethionine (SAM) [46]. These findings indicate that the candidate genes in metabolism of one-carbon units may be involved in NAFLD.

Currently, no medications have been clinically approved for the treatment of NAFLD [47]. However, drugs that improve insulin resistance [14, 48], modulate disordered lipid metabolism [48], inhibit oxidative stress [49], provide anti-inflammatory and anti-fibrotic effects [50], and correct gut microbiota dysbiosis [51], may offer potential therapeutic benefits for NAFLD. To discover potential therapeutic drugs for NAFLD, we utilized the Drug-Gene Interaction Database (DGIdb) to investigate medications that target the 27 candidate genes implicated in NAFLD. These analyses identified 36 approved drugs targeting 6 genes: gamma-glutamyltransferase 1 (*GGT1*), actin gamma 1 (*ACTG1*), secreted phosphoprotein 1 (*SPP1*), EPH receptor A2 (*EPHA2*), protein Z vitamin K-dependent plasma glycoprotein (*PROZ*), and serine hydroxymethyltransferase 1 (*SHMT1*) (Fig. 4C; Table S4). Among these 6 genes, *SHMT1* was previously recognized as a promising therapeutic target for NAFLD [52]. The other five genes have not been reported as NAFLD targets. Notably, *GGT1*, a target of the drug liothyronine, arouses our interest (Table S4). Liothyronine, a thyroid hormone medication used to manage hypothyroidism, has been observed to be associated with a significantly higher incidence in patients with NAFLD compared to age-matched controls [53, 54]. Together, these findings suggest novel therapeutic targets for NAFLD and warrant further investigation into their potential clinical applications.

GGT1 at 22q11.23

Next, we focused on *GGT1* for further analyses. Among the regulatory SNPs in this candidate gene, *rs2017869* at 22q11.23 emerged as the most promising SNP, as this SNP showed the highest ABC score (0.219345; Table S5). This SNP showed significantly associated with NAFLD, which reached the statistical significance threshold (OR = 1.10, 95% CI = 1.04–1.16, $P = 5.97 \times 10^{-4}$ in the discovery stage; OR = 1.06, 95% CI = 1.02–1.11, $P = 4.15 \times 10^{-3}$ in the replication stage; $P_{meta} = 1.24 \times 10^{-5}$).

The eQTL and colocalization analyses facilitate prioritization of candidate causal genes. Thus, to further confirm that *GGT1* is the potentially causative gene at this locus, we checked the results of eQTL using several publicly available datasets (Supplementary Methods). According to the GTEx v8 database, the risk allele C of *rs2017869* exhibited a significant association with elevated expression levels of the *GGT1* gene in liver tissue ($\beta = 0.43$; $P = 4.57 \times 10^{-8}$; Table S6). This eQTL signal was further replicated in blood samples from the GTEx database, as well as in liver tissues and blood from both the QTLbase and eQTLGen databases (Table S6). These eQTL analyses indicated that *GGT1* stands out as a highly probable candidate gene at this locus. Furthermore, we conducted colocalization analyses. Although the posterior probability of hypothesis 4 (PP.H4) for *GGT1* (PP.H4 score = 0.3; data not shown) was higher than those of nearby genes, it failed to meet the predefined colocalization threshold of 0.8, precluding definitive conclusions on colocalization.

To further investigate the role of *GGT1* in NAFLD development, we examined the pathways associated with *GGT1*. Among the ten identified pathways, *GGT1* was enriched in pathways such as "serine family amino acid biosynthetic process", "serine family amino acid metabolic process", "proteinogenic amino acid biosynthetic process", "L-amino acid biosynthetic process", "amino acid biosynthetic process", and "alpha-amino acid biosynthetic process" (Fig. 5A). Previous studies have suggested that altered amino acid concentrations are frequently observed in NAFLD, with serine exhibiting a negative association with NAFLD [55, 56]. Additionally, serine synthesis, coupled one-carbon metabolism, produces glutathione, which is essential for maintaining redox balance. Glutathione is a primary substrate for gamma-glutamyl transferase (GGT), the enzyme encoded by *GGT1*. These findings highlight the crucial role of *GGT1* in NAFLD.

We next investigated potential drug targets for *GGT1*. Drug-gene interaction analyses from DGIdb revealed 17 approved drugs or compounds that interact with *GGT1* (Fig. 5B; Table S4), some of which have significant clinical relevance. Among these drugs, Liothyronine may provide benefits to patients with NAFLD through the supplementation of thyroid hormone [53, 54]. Diclofenac sodium, indomethacin and piroxicam-beta-cyclodextrin complex are non-steroidal anti-inflammatory drugs (NSAIDs), while dexamethasone is a glucocorticoid. Both NSAIDs and glucocorticoids exhibit anti-inflammatory effects, which are a key therapeutic strategy in the treatment of NAFLD [57, 58]. Ursodiol, known for its hepatoprotective properties in NAFLD [59], has been shown to significantly reduce GGT levels [60].Collectively, these findings underscore the crucial role that *GGT1* plays in the pathogenesis of NAFLD and its potential as a promising therapeutic target.

EPHA2 at 1p36.13

Besides the *rs2017869-GGT1*, the *rs1497406-EPHA2* pair seemed to be an interesting finding. The SNP *rs1497406*, located at 1p36.13, was previously found to be associated with GGT concentration [61]. However, the candidate gene for *rs1497406* was not identified in that study [61]. In contrast, using the ABC regulatory maps, we identified *EPHA2*, which was one of the five novel NAFLD targets, as the target gene of *rs1497406*. Therefore, we identified another SNP-gene pair, *rs1497406-EPHA2*, which was significantly associated with NAFLD (OR = 1.12, 95% CI = 1.06–1.19, $P = 3.81 \times 10^{-5}$ in the discovery

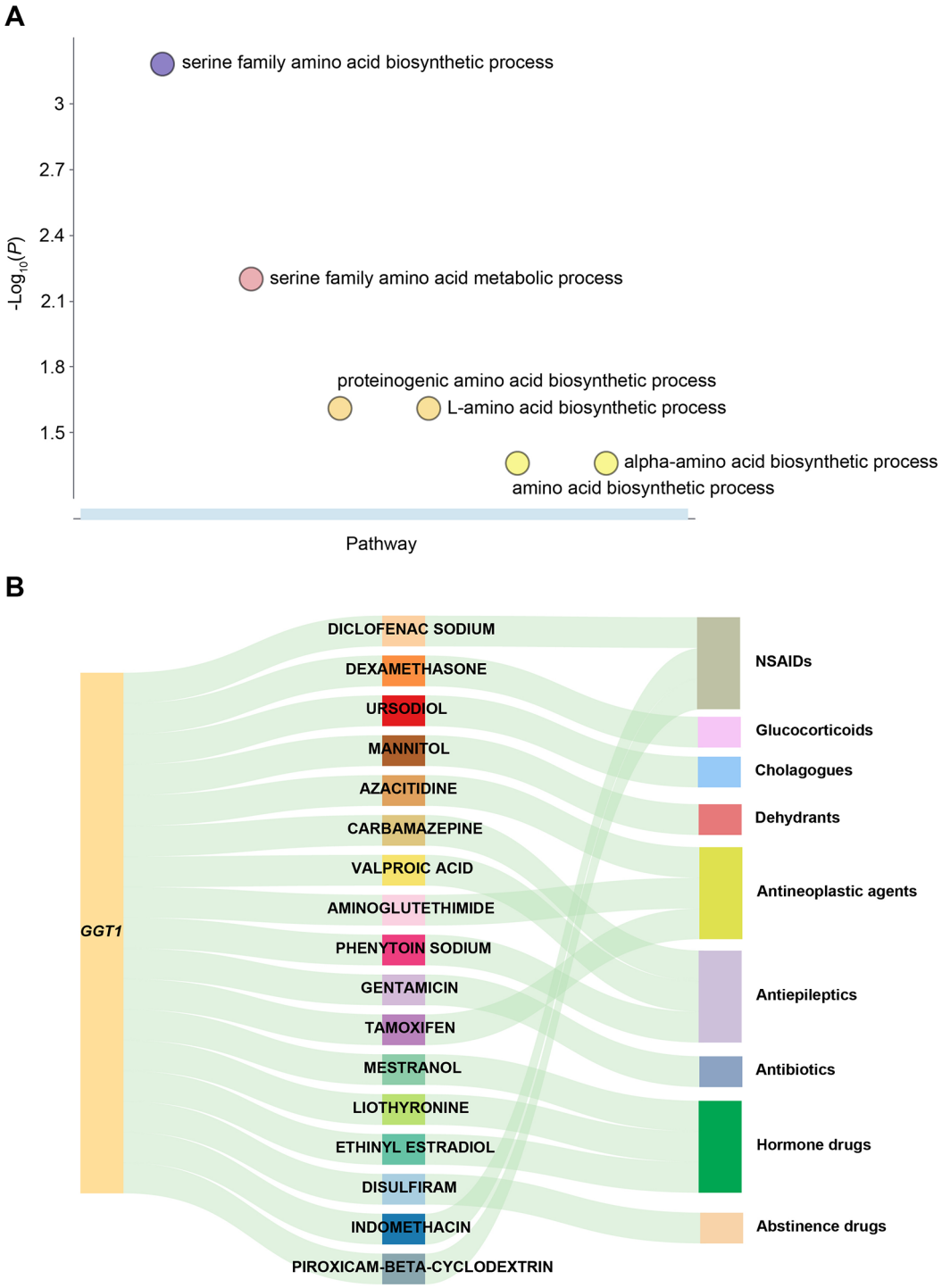

Fig. 5 Associated pathways and drugs of *GGT1* at 22q11.23. (**A**) Bubble plot shows the pathways associated with *GGT1*. The x-axis represents the enriched pathways, while the y-axis represents the -$\log_{10}(P)$. (**B**) Sankey diagram indicates the identified drugs of drug-gene interaction analyses from DGIdb targeting *GGT1*. The right blocks indicate the corresponding categories of these drugs. *GGT1*, gamma-glutamyltransferase 1; DGIdb, Drug-Gene Interaction Database

stage; OR = 1.07, 95% CI = 1.03–1.12, $P = 7.99 \times 10^{-4}$ in the replication stage; $P_{meta} = 2.14 \times 10^{-7}$). To validate this finding, we performed eQTL and colocalization analyses (Supplementary Methods), demonstrating that the NAFLD-associated SNP *rs1497406* colocalizes with eQTL signals for *EPHA2* (Fig. 6A; Table S7). Thus, these findings indicate that *EPHA2* plays a consistent and pivotal role in regulating GGT levels, thereby influencing

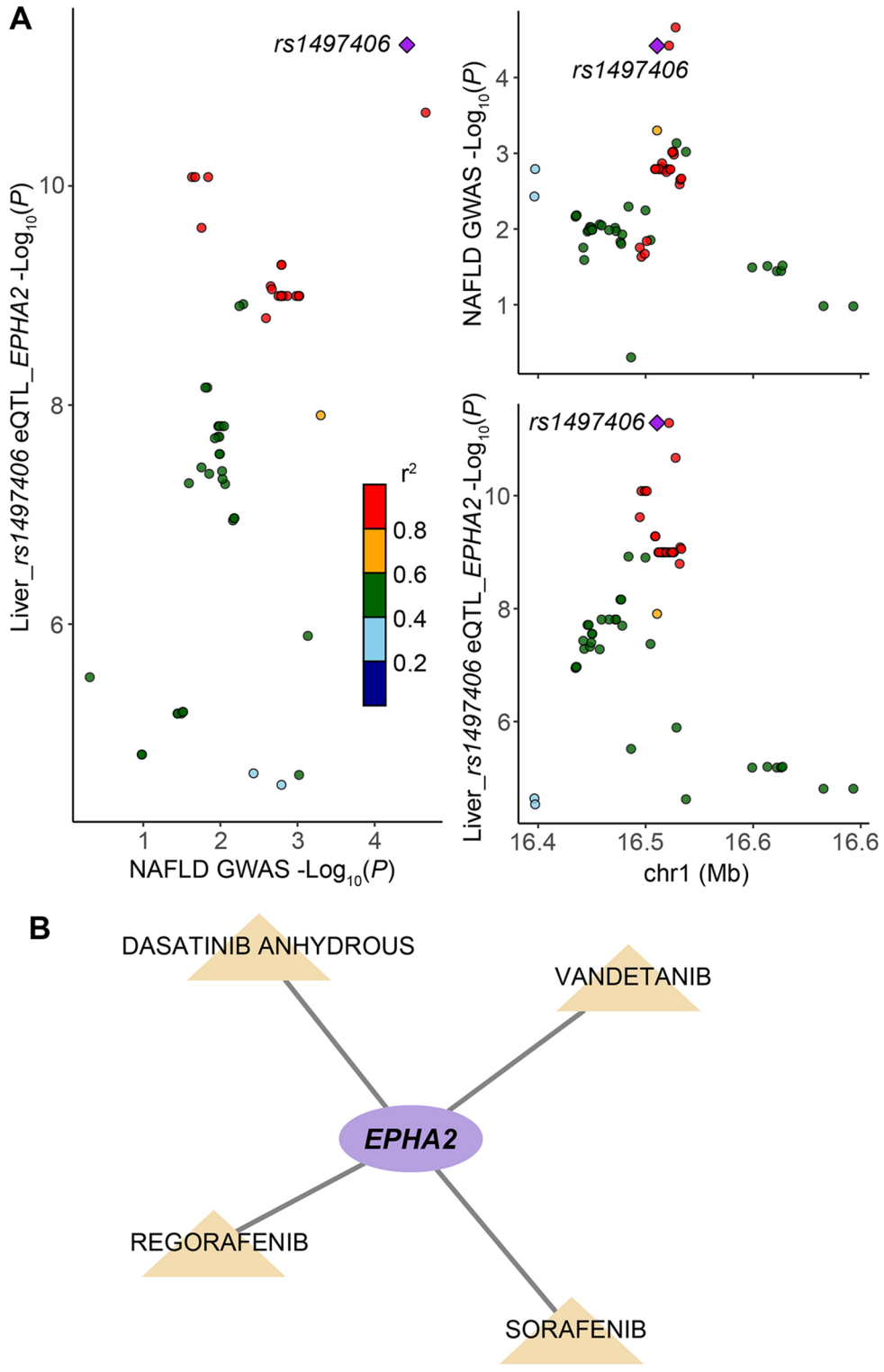

Fig. 6 *EPHA2* at 1p36.13. (**A**) The NAFLD GWAS summary statistics were obtained from the FinnGen study. The liver eQTL summary statistics for *EPHA2* were downloaded from GTEx v8. The LD values (r^2) between the SNP *rs1497406* and the other SNPs are based on European populations (from the 1,000 Genomes Project, Phase 3). The colocalization analyses were performed using the R package "coloc" (v5.2.3) and achieved a posterior probability of hypothesis 4 (PP.H4) score of 0.98, suggesting that the eQTLs and GWAS associations were highly likely to colocalize. (**B**) Drug-gene interaction network indicates the identified drugs of drug-gene interaction analyses from DGIdb targeting *EPHA2*. *EPHA2*, EPH receptor A2; NAFLD, non-alcoholic fatty liver disease; GWAS, genome-wide association study; eQTL, expression quantitative trait locus; GTEx, Genotype-Tissue Expression; SNP, single nucleotide polymorphism; LD, linkage disequilibrium; DGIdb, Drug-Gene Interaction Database

the risk of developing NAFLD, positioning it as another potential candidate gene for NAFLD.

The drug-gene interaction analyses identified four approved drugs targeting *EPHA2*, including regorafenib and sorafenib, which are clinically utilized as antitumor agents for the treatment of HCC (Fig. 6B). Regorafenib has been reported to induce hepatotoxicity by inhibiting *EPHA2* Ser897 phosphorylation, suggesting a potential role for *EPHA2* in liver damage [62]. Additionally, existing studies indicate that targeting *EPHA2* can effectively treat liver fibrosis [63], further supporting its significance in the development of NAFLD. Collectively, *EPHA2* emerged as a potential susceptibility gene influencing NAFLD risk in our analyses.

Discussion

Despite the identification of SNPs associated with NAFLD through large-scale GWASs, mapping SNPs located in non-coding regions, particularly those located within enhancers, to their target genes continues to pose a notable challenge. In this study, we integrated large-scale multi-omics data to construct the most comprehensive genome-wide regulatory maps of the liver utilizing the ABC model. We identified 543,486 enhancer-gene connections, involving 267,857 enhancers and 16,872 target genes. Enrichment analyses revealed that the ABC SNPs are significantly enriched in active chromatin regions and active chromatin state, providing strong evidence for their regulatory role. By combining the regulatory maps and GWAS data, we systematically identified ABC SNPs associated with NAFLD risk in European population. Among these identified SNPs, we found that the ABC SNP *rs2017869*, with the highest ABC score, is significantly associated with NAFLD risk. Therefore, *GGT1* was identified as a novel NAFLD susceptibility gene with *rs2017869* in its enhancer region. Additionally, our study also indicated that *EPHA2*, another novel susceptibility gene, may play a pivotal role in NAFLD by modulating GGT levels. These findings emphasize the significance of the ABC regulatory maps in linking non-coding SNPs to their respective target genes, providing valuable insights into the regulatory mechanisms of NAFLD development. Given the minimal clinical differences between NAFLD and MASLD [64], our findings also hold significant implications for research on MASLD.

Although previous studies have established ABC enhancer-gene connections in liver biosamples, these analyses typically relied on data from only one or two liver biosamples to construct the ABC maps [30, 31]. Given the considerable diversity within major liver cell populations [65], there is a need to expand the scope of ABC regulatory maps. To address this, we systematically explored available epigenomic data from the liver and integrated datasets from 15 liver tissues and cell lines to build more comprehensive ABC regulatory maps. Compared to the number of the ABC genes identified in a single liver biosample (HepG2_1) [31], our expanded ABC maps uncovered an additional 607 ABC genes. This provides a more extensive resource of regulatory maps for investigating the genetic risk factors associated with the development of liver diseases.

The ABC regulatory maps offer a distinct advantage in clearly delineating enhancer-gene relationships, facilitating the linkage of GWAS signals in enhancer regions to their target genes. Previous GWASs have identified several common SNPs in genes such as *PNPLA3*, *TM6SF2*, *MBOAT7*, *GCKR*, and *HSD17B13*, which are consistently and robustly associated with NAFLD [66]. However, most of these SNPs are located within coding sequences (Table S8), highlighting the challenges of connecting non-coding SNPs associated with NAFLD to their underlying target genes. By integrating the ABC maps with GWAS data, we reported for the first time that *rs2017869*, a non-coding GWAS signal located in the enhancer region of *GGT1* at 22q11.23, may influence NAFLD risk. To the best of our knowledge, no previous GWAS has linked risk SNPs mapped to the *GGT1* gene to NAFLD. Therefore, by using ABC regulatory maps, we successfully identified *GGT1* as a novel gene influencing NAFLD risk. Additionally, we examined whether NAFLD-associated SNPs were located within ABC enhancers of these well-established NAFLD susceptibility genes. Among the five NAFLD susceptibility genes analyzed, the integration of the ABC maps and GWAS summary data revealed that only *TM6SF2* expression may be regulated by NAFLD-associated SNPs in its ABC enhancer, thereby influencing NAFLD risk (Table S9).

The identification of *rs2017869-GGT1* and *rs1497406-EPHA2* pairs prompted us to investigate SNPs associated with gamma-glutamyl transferase (GGT) levels. Seo et al. identified that two SNPs within *GGT1* are strongly associated with GGT levels (*rs5751901*, $P = 6.44 \times 10^{-15}$; *rs2006092*, $P = 1.26 \times 10^{-15}$) [67]. Similarly, Yuan et al. found that *rs4820599* is associated with GGT levels, with a P value of 4.0×10^{-11} [68]. All these three SNPs are in strong LD with our identified SNP *rs2017869* ($r^2 > 0.6$). GGT, encoded by *GGT1*, is well established as a key biomarker for NAFLD and plays a critical role in regulating glutathione (GSH) levels. As the primary substrate of GGT, GSH has anti-oxidation as one of its most critical biological functions. Previous studies have demonstrated that oral administration of GSH has therapeutic effects on NAFLD patients [69], while reduced GGT activity could benefit these patients [70]. These findings indicate that *GGT1* contributes to NAFLD through GGT-mediated degradation of GSH, which impairs the suppression of oxidative stress. In addition, *rs1497406* was previously found to be associated with GGT concentration [61].

According to the ABC regulatory maps, its target gene is *EPHA2*, suggesting a link between *EPHA2* and GGT levels. These findings collectively suggest that the relationship between dysregulated GGT levels/activity and oxidative stress may be a key determinant in NAFLD. Our study provides evidence that enhancer SNPs (*rs2017869* and *rs1497406*) regulate the target genes (*GGT1* and *EPHA2*) and suggests that altered expression of these genes affects GGT levels, thereby influencing NAFLD through oxidative stress mechanisms. Further investigation should focus on the precise molecular mechanisms and biological functions of these two genes.

In our analyses, the effect size of *rs2017869* was relatively small (OR = 1.10 in the discovery stage). As studies grow larger in scale, the average effect size of the novel identified variants will decrease [71]. However, for insights into disease pathogenesis, validated small-effect loci can reveal novel causal mechanisms [72]. It should be noted that although the effect sizes of some SNPs are small, those on molecular phenotypes and therapeutic effects of gene targets can be large [73, 74]. For example, an identified enhancer SNP *rs4810856* associated with colorectal cancer (CRC) also exhibited small OR value (OR = 1.11, 95% CI = 1.04–1.16, $P = 4.02 \times 10^{-5}$). However, this SNP was validated as a CRC risk locus via functional studies [31]. Therefore, SNPs with small effect sizes may remain clinically significant.

Drugs targeting the identified candidate genes, particularly *GGT1*, could represent promising therapeutic agents for NAFLD, as their mechanisms align with established therapeutic strategies [48–50], such as modulating oxidative stress, regulating lipid metabolism, and exerting anti-inflammatory effects. Future research elucidating the impact of enhancer SNPs on drug responses could further advance precision medicine. Numerous studies have demonstrated that SNPs are associated with both disease susceptibility and drug responses [75, 76], establishing a theoretical and clinical foundation for individualized treatment. Notably, Sun et al. revealed that the enhancer SNP *rs2017869* of *GGT1* alters the outcomes of neoadjuvant chemotherapy in breast cancer by regulating GGT levels [77]. This finding provides a critical reference for future validation of *rs2017869* and *rs1497406* in modulating drug responses.

Our study has several limitations. First, due to the cell-type specificity of enhancer elements, the use of epigenomic data from single-cell technologies would provide a more detailed understanding of enhancer-gene connections. However, such resources are currently limited. Second, the ABC approach does not fully encompass the impacts of distal enhancers and may fail to account for other forms of transcriptional or post-transcriptional regulatory elements. Third, the ABC model depends on computational predictions to ascertain enhancer-gene connections. To validate putative gene-element interactions at transcriptional, functional, and spatial levels, experimental validation would be required in further studies. Luciferase reporter assays can assess the transcriptional regulatory activity of candidate DNA elements by linking enhancers to a luciferase reporter gene. CRISPR-Cas9-mediated genome editing can functionally disrupt specific regulatory elements to determine their necessity for target gene expression. Chromosome conformation capture (3 C) techniques such as Hi-C can map three-dimensional chromatin architecture to verify physical interactions between genes and regulatory elements. Fourth, the PP.H4 score of 0.3 for *GGT1* in colocalization analysis suggested no conclusive signal, weakening the causal claim for *rs2017869*. Thus, further investigation with experimental validation is required. Fifth, as our analyses were primarily based on individuals of European descent, the findings are likely more applicable to this population. Replication in multi-ethnic cohorts is critical to determine whether these results are generalizable beyond Europeans.

Conclusions

In conclusion, our study provides the most comprehensive ABC regulatory maps of the liver to date by integrating multi-omics data that reflects the activity of candidate regulatory elements and chromatin interaction frequencies. This resource serves as a valuable reference for identifying regulatory variants and prioritizing genes susceptible to liver diseases, including NAFLD.

Abbreviations

3 C	Chromosome conformation capture
ABC	Activity-by-Contact
ACTG1	Actin gamma 1
BP	Biological process
CC	Cellular component
CI	Confidence interval
CKD	Chronic kidney disease
CRC	Colorectal cancer
CVD	Cardiovascular disease
DGIdb	Drug-Gene Interaction Database
EPHA2	EPH receptor A2
GCKR	Glucokinase regulator
GGT	Gamma-glutamyl transferase
GGT1	Gamma-glutamyltransferase 1
GO	Gene Ontology
GSH	Glutathione
GTEx	Genotype-Tissue Expression
GWAS	Genome-wide association study
HCC	Hepatocellular carcinoma
HSD17B13	Hydroxysteroid 17-beta dehydrogenase 13
ICD	International Classification of Diseases
KEGG	Kyoto Encyclopedia of Genes and Genomes
LD	Linkage disequilibrium
LDSC	LD score regression
MASLD	Metabolic dysfunction-associated steatotic liver disease
MBOAT7	Membrane-bound O-acyltransferase domain containing 7
MF	Molecular function
NAFL	Non-alcoholic fatty liver
NAFLD	Non-alcoholic fatty liver disease
NASH	Non-alcoholic steatohepatitis

NSAID	Non-steroidal anti-inflammatory drug
OR	Odds ratio
PNPLA3	Patatin-like phospholipase domain-containing 3
PP.H4	Posterior probability of hypothesis 4
PROZ	Protein Z vitamin K-dependent plasma glycoprotein
QTL	Quantitative trait locus
SAM	S-adenosylmethionine
SHMT1	Serine hydroxymethyltransferase 1
SNP	Single nucleotide polymorphism
SPP1	Secreted phosphoprotein 1
T2DM	Type 2 diabetes mellitus
TFBS	Transcription factor binding site
TM6SF2	Transmembrane 6 superfamily member 2
TSS	Transcription start site

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11668-w.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
The authors would like to thank ENCODE, the Roadmap Epigenomics Project, FinnGen consortia, GWAS Catalog, GTEx, QTLbase and eQTLgen for contributing data.

Author contributions
M.H., M.Z., Y.L., H.X. and J.C. designed the study; R.L., K.S., and T.W. performed bioinformatics analyses; W.S., T.Z., and D.S. helped with the tables and figures; R.L., K.S., M.Z., Y.L., and M.H. periodically discussed and interpreted the data; R.L., K.S., and L.X. wrote the paper; M.H. and M.Z. supervised the study; All authors read and approved the final version of the manuscript.

Funding
This work was supported by grant from the special project of the key research and development tasks of Xinjiang Uygur autonomous region (No. 2022B03005-2).

Data availability
All data of this article is described in supplementary information and available from the corresponding author on request. The DNase-seq, ATAC-seq, H3K27ac ChIP-seq and Hi-C data were obtained from ENCODE and Roadmap databases, and the data sources are listed in Table S1. The summary GWAS statistics for NAFLD were downloaded from the FinnGen study (https://www.r11.finngen.fi/pheno/NAFLD) and GWAS Catalog (https://www.ebi.ac.uk/gwas/studies/GCST90054782). The eQTL datasets were obtained from GTEx v8 (https://www.gtexportal.org/home/downloads/adult-gtex/qtl), QTLbase (http://www.mulinlab.org/qtlbase), and eQTLgen (https://www.eqtlgen.org/).

Declarations

Ethics approval and consent to participate
Ethical approval was obtained from the original studies.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Medical School of Chinese People's Liberation Army (PLA), 28 Fuxing Road, 100853 Beijing, China
[2]Faculty of Hepato-Biliary-Pancreatic Surgery, The First Medical Center of Chinese People's Liberation Army (PLA) General Hospital, 28 Fuxing Road, Beijing 100853, China
[3]State Key Laboratory of Organ Failure Research, Guangdong Provincial Key Laboratory for Prevention and Control of Major Liver Diseases, Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, 1,838 North Guangzhou Ave, Guangzhou, Guangdong 510515, China
[4]Department of Digestive Oncology, Guangxi Medical University Cancer Hospital, Nanning 530021, China
[5]Department of Hepatopancreatobiliary Surgery, China-Japan Friendship Hospital, Beijing 100029, China
[6]State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences at Beijing, Beijing Institute of Radiation Medicine, 27 Taiping Road, Beijing 100850, China
[7]Department of Radiation Oncology, Harbin Medical University Cancer Hospital, Harbin 150081, China
[8]Department of Hepatobiliary Surgery, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai 200071, China

Received: 4 January 2025 / Accepted: 2 May 2025
Published online: 15 May 2025

References

1. Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, Romero D, Abdelmalek MF, Anstee QM, Arab JP, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. Hepatology. 2023;78(6):1966–86.
2. Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, George J, Bugianesi E. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol. 2018;15(1):11–20.
3. Riazi K, Azhari H, Charette JH, Underwood FE, King JA, Afshar EE, Swain MG, Congly SE, Kaplan GG, Shaheen AA. The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol. 2022;7(9):851–61.
4. Pais R, Barritt ASt, Calmus Y, Scatton O, Runge T, Lebray P, Poynard T, Ratziu V, Conti F. NAFLD and liver transplantation: current burden and expected challenges. J Hepatol. 2016;65(6):1245–57.
5. Brouwers M, Simons N, Stehouwer CDA, Isaacs A. Non-alcoholic fatty liver disease and cardiovascular disease: assessing the evidence for causality. Diabetologia. 2020;63(2):253–60.
6. Targher G, Day CP, Bonora E. Risk of cardiovascular disease in patients with nonalcoholic fatty liver disease. N Engl J Med. 2010;363(14):1341–50.
7. Targher G, Corey KE, Byrne CD, Roden M. The complex link between NAFLD and type 2 diabetes mellitus - mechanisms and treatments. Nat Rev Gastroenterol Hepatol. 2021;18(9):599–612.
8. Deprince A, Haas JT, Staels B. Dysregulated lipid metabolism links NAFLD to cardiovascular disease. Mol Metab. 2020;42:101092.
9. Musso G, Gambino R, Tabibian JH, Ekstedt M, Kechagias S, Hamaguchi M, Hultcrantz R, Hagström H, Yoon SK, Charatcharoenwitthaya P, et al. Association of non-alcoholic fatty liver disease with chronic kidney disease: a systematic review and meta-analysis. PLoS Med. 2014;11(7):e1001680.
10. Byrne CD, Targher G. NAFLD: a multisystem disease. J Hepatol. 2015;62(1 Suppl):S47–64.
11. Younossi ZM, Blissett D, Blissett R, Henry L, Stepanova M, Younossi Y, Racila A, Hunt S, Beckerman R. The economic and clinical burden of nonalcoholic fatty liver disease in the united States and Europe. Hepatology. 2016;64(5):1577–86.
12. Dufour JF, Caussy C, Loomba R. Combination therapy for non-alcoholic steatohepatitis: rationale, opportunities and challenges. Gut. 2020;69(10):1877–84.
13. Rinella ME. Nonalcoholic fatty liver disease: A systematic review. JAMA. 2015;313(22):2263–73.
14. Chalasani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, Harrison SA, Brunt EM, Sanyal AJ. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American association for the study of liver diseases. Hepatology. 2018;67(1):328–57.
15. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, Boerwinkle E, Cohen JC, Hobbs HH. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet. 2008;40(12):1461–5.
16. Singal AG, Manjunath H, Yopp AC, Beg MS, Marrero JA, Gopal P, Waljee AK. The effect of PNPLA3 on fibrosis progression and development

of hepatocellular carcinoma: a meta-analysis. Am J Gastroenterol. 2014;109(3):325–34.
17. Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjærg-Hansen A, Vogt TF, Hobbs HH, Cohen JC. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet. 2014;46(4):352–6.
18. Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, Guo Y, Zhang J, Langhammer A, Løchen ML, et al. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. Nat Genet. 2014;46(4):345–51.
19. Liu YL, Reeves HL, Burt AD, Tiniakos D, McPherson S, Leathart JB, Allison ME, Alexander GJ, Piguet AC, Anty R, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nat Commun. 2014;5:4309.
20. Mancina RM, Dongiovanni P, Petta S, Pingitore P, Meroni M, Rametta R, Borén J, Montalcini T, Pujia A, Wiklund O, et al. The MBOAT7-TMC4 variant rs641738 increases risk of nonalcoholic fatty liver disease in individuals of European descent. Gastroenterology. 2016;150(5):1219–e12301216.
21. Luukkonen PK, Zhou Y, Hyötyläinen T, Leivonen M, Arola J, Orho-Melander M, Orešič M, Yki-Järvinen H. The MBOAT7 variant rs641738 alters hepatic phosphatidylinositols and increases severity of non-alcoholic fatty liver disease in humans. J Hepatol. 2016;65(6):1263–5.
22. Santoro N, Zhang CK, Zhao H, Pakstis AJ, Kim G, Kursawe R, Dykas DJ, Bale AE, Giannini C, Pierpont B, et al. Variant in the glucokinase regulatory protein (GCKR) gene is associated with fatty liver in obese children and adolescents. Hepatology. 2012;55(3):781–9.
23. Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, Gudnason V, Eiriksdottir G, Garcia ME, Launer LJ, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet. 2011;7(3):e1001324.
24. Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, Liu Y, Kozlitina J, Stender S, Wood GC, et al. A Protein-Truncating HSD17B13 variant and protection from chronic liver disease. N Engl J Med. 2018;378(12):1096–106.
25. Petta S, Valenti L, Marra F, Grimaudo S, Tripodo C, Bugianesi E, Cammà C, Cappon A, Di Marco V, Di Maira G, et al. MERTK rs4374383 polymorphism affects the severity of fibrosis in non-alcoholic fatty liver disease. J Hepatol. 2016;64(3):682–90.
26. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190–5.
27. Yao DW, O'Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. Nat Genet. 2020;52(6):626–33.
28. Chen H, Wang Z, Gong L, Wang Q, Chen W, Wang J, Ma X, Ding R, Li X, Zou X, et al. A distinct class of pan-cancer susceptibility genes revealed by an alternative polyadenylation transcriptome-wide association study. Nat Commun. 2024;15(1):1729.
29. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, Shoresh N, Whitton H, Ryan RJ, Shishkin AA, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518(7539):337–43.
30. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, Jones TR, Nguyen TH, Ulirsch JC, Lekschas F, et al. Genome-wide enhancer maps link risk variants to disease genes. Nature. 2021;593(7858):238–43.
31. Ying P, Chen C, Lu Z, Chen S, Zhang M, Cai Y, Zhang F, Huang J, Fan L, Ning C, et al. Genome-wide enhancer-gene regulatory maps link causal variants to target genes underlying human cancer risk. Nat Commun. 2023;14(1):5958.
32. An integrated encyclopedia. Of DNA elements in the human genome. Nature. 2012;489(7414):57–74.
33. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–30.
34. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. Nat Genet. 2019;51(12):1664–9.
35. Huang D, Wang Z, Zhou Y, Liang Q, Sham PC, Yao H, Li MJ. vSampler: fast and annotation-based matched variant sampling tool. Bioinformatics. 2021;37(13):1915–7.
36. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92.
37. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25(1):25–9.
38. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353–61.
39. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, Haw R, Jassal B, Matthews L, May B, et al. The reactome pathway knowledgebase 2024. Nucleic Acids Res. 2024;52(D1):D672–8.
40. Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, Pico AR. WikiPathways: Building research communities on biological pathways. Nucleic Acids Res. 2012;40(Database issue):D1301–1307.
41. Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res. 2023;51(W1):W207–12.
42. Cannon M, Stevenson J, Stahl K, Basu R, Coffman A, Kiwala S, McMichael JF, Kuzma K, Morrissey D, Cotto K, et al. DGIdb 5.0: rebuilding the drug-gene interaction database for precision medicine and drug discovery platforms. Nucleic Acids Res. 2024;52(D1):D1227–35.
43. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010;464(7289):768–72.
44. Sowers ML, Herring J, Zhang W, Tang H, Ou Y, Gu W, Zhang K. Analysis of glucose-derived amino acids involved in one-carbon and cancer metabolism by stable-isotope tracing gas chromatography mass spectrometry. Anal Biochem. 2019;566:1–9.
45. Stead LM, Brosnan JT, Brosnan ME, Vance DE, Jacobs RL. Is it time to Reevaluate Methyl balance in humans? Am J Clin Nutr. 2006;83(1):5–10.
46. Newman AC, Maddocks ODK. One-carbon metabolism in cancer. Br J Cancer. 2017;116(12):1499–504.
47. Roeb E, Geier A. Nonalcoholic steatohepatitis (NASH) - current treatment recommendations and future developments. Z Gastroenterol. 2019;57(4):508–17.
48. Harrison SA, Bashir MR, Guy CD, Zhou R, Moylan CA, Frias JP, Alkhouri N, Bansal MB, Baum S, Neuschwander-Tetri BA, et al. Resmetirom (MGL-3196) for the treatment of non-alcoholic steatohepatitis: a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. Lancet. 2019;394(10213):2012–24.
49. Sanyal AJ, Chalasani N, Kowdley KV, McCullough A, Diehl AM, Bass NM, Neuschwander-Tetri BA, Lavine JE, Tonascia J, Unalp A, et al. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. N Engl J Med. 2010;362(18):1675–85.
50. Ratziu V, de Ledinghen V, Oberti F, Mathurin P, Wartelle-Bladou C, Renou C, Sogni P, Maynard M, Larrey D, Serfaty L, et al. A randomized controlled trial of high-dose ursodesoxycholic acid for nonalcoholic steatohepatitis. J Hepatol. 2011;54(5):1011–9.
51. Fan JG, Xu ZJ, Wang GL. Effect of lactulose on establishment of a rat nonalcoholic steatohepatitis model. World J Gastroenterol. 2005;11(32):5053–6.
52. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J. Genome-scale metabolic modelling of hepatocytes reveals Serine deficiency in patients with non-alcoholic fatty liver disease. Nat Commun. 2014;5:3083.
53. Sinha RA, Bruinstroop E, Singh BK, Yen PM. Nonalcoholic fatty liver disease and hypercholesterolemia: roles of thyroid hormones, metabolites, and agonists. Thyroid. 2019;29(9):1173–91.
54. Bohinc BN, Michelotti G, Xie G, Pang H, Suzuki A, Guy CD, Piercy D, Kruger L, Swiderska-Syn M, Machado M, et al. Repair-related activation of Hedgehog signaling in stromal cells promotes intrahepatic hypothyroidism. Endocrinology. 2014;155(11):4591–601.
55. Pietzner M, Budde K, Homuth G, Kastenmüller G, Henning AK, Artati A, Krumsiek J, Völzke H, Adamski J, Lerch MM, et al. Hepatic steatosis is associated with adverse molecular signatures in subjects without diabetes. J Clin Endocrinol Metab. 2018;103(10):3856–68.
56. Hasegawa T, Iino C, Endo T, Mikami K, Kimura M, Sawada N, Nakaji S, Fukuda S. Changed amino acids in NAFLD and liver fibrosis: A large Cross-Sectional study without influence of insulin resistance. Nutrients. 2020;12(5).
57. Markowska J, Kasprzak-Drozd K, Niziński P, Dragan M, Kondracka A, Gondek E, Oniszczuk T, Oniszczuk A. Quercetin: A Promising Candidate for the Management of Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD). Molecules. 2024;29(22).
58. Wang Z, Gao P, Gao J, Liang B, Ma Q, Sun Q, Hu Y, Wang Y, Peng Y, Liu H, et al. Daphnetin ameliorates hepatic steatosis by suppressing peroxisome

proliferator-activated receptor gamma (PPARG) in Ob/ob mice. Biochem Pharmacol. 2024;230(Pt 3):116610.
59. Patel VS, Mahmood SF, Bhatt KH, Khemkar RM, Jariwala DR, Harris B, George MM, Kurudamannil RA, Anyagwa OE, Tak RS, et al. Ursodeoxycholic acid's effectiveness in the management of nonalcoholic fatty liver disease: A systematic review and Meta-analysis. Euroasian J Hepatogastroenterol. 2024;14(1):92–8.
60. Kogiso T, Ogasawara Y, Taniai M, Shimada E, Inai K, Tokushige K, Nakai Y. Impact of ursodeoxycholic acid treatment on Fontan-associated liver disease. J Gastroenterol 2024.
61. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, Holm H, Sanna S, Kavousi M, Baumeister SE, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat Genet. 2011;43(11):1131–8.
62. Yan H, Wu W, Hu Y, Li J, Xu J, Chen X, Xu Z, Yang X, Yang B, He Q, et al. Regorafenib inhibits EphA2 phosphorylation and leads to liver damage via the ERK/MDM2/p53 axis. Nat Commun. 2023;14(1):2756.
63. Zhang Q, Ran T, Li S, Han L, Chen S, Lin G, Wu H, Wu H, Feng S, Chen J, et al. Catalpol ameliorates liver fibrosis via inhibiting aerobic Glycolysis by EphA2/FAK/Src signaling pathway. Phytomedicine. 2024;135:156047.
64. Song SJ, Lai JC, Wong GL, Wong VW, Yip TC. Can we use old NAFLD data under the new MASLD definition? J Hepatol. 2024;80(2):e54–6.
65. Aizarani N, Saviano A, Sagar, Mailly L, Durand S, Herman JS, Pessaux P, Baumert TF. Grün D: A human liver cell atlas reveals heterogeneity and epithelial progenitors. Nature. 2019;572(7768):199–204.
66. Trépo E, Valenti L. Update on NAFLD genetics: from new variants to the clinic. J Hepatol. 2020;72(6):1196–209.
67. Seo JY, Lee JE, Chung GE, Shin E, Kwak MS, Yang JI, Yim JY. A genome-wide association study on liver enzymes in Korean population. PLoS ONE. 2020;15(2):e0229374.
68. Yuan X, Waterworth D, Perry JR, Lim N, Song K, Chambers JC, Zhang W, Vollenweider P, Stirnadel H, Johnson T, et al. Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. Am J Hum Genet. 2008;83(4):520–8.
69. Honda Y, Kessoku T, Sumida Y, Kobayashi T, Kato T, Ogawa Y, Tomeno W, Imajo K, Fujita K, Yoneda M, et al. Efficacy of glutathione for the treatment of nonalcoholic fatty liver disease: an open-label, single-arm, multicenter, pilot study. BMC Gastroenterol. 2017;17(1):96.
70. Šmíd V, Dvořák K, Šedivý P, Kosek V, Leníček M, Dezortová M, Hajšlová J, Hájek M, Vítek L, Bechyňská K, et al. Effect of Omega-3 polyunsaturated fatty acids on lipid metabolism in patients with metabolic syndrome and NAFLD. Hepatol Commun. 2022;6(6):1336–49.
71. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 10 Years of GWAS discovery: biology, function, and translation. Am J Hum Genet. 2017;101(1):5–22.
72. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356–69.
73. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, Jansen R, de Geus EJ, Boomsma DI, Wright FA, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48(3):245–52.
74. Nelson MR, Johnson T, Warren L, Hughes AR, Chissoe SL, Xu CF, Waterworth DM. The genetics of drug efficacy: opportunities and challenges. Nat Rev Genet. 2016;17(4):197–206.
75. Pérez-Ramírez C, Cañadas-Garre M, Molina M, Cabeza Barrera J, Faus-Dáder MJ. Impact of single nucleotide polymorphisms on the efficacy and toxicity of EGFR tyrosine kinase inhibitors in advanced non-small cell lung cancer patients. Mutat Res Rev Mutat Res. 2019;781:63–70.
76. Gummadi AC, Guddati AK. Genetic polymorphisms in pharmaceuticals and chemotherapy. World J Oncol. 2021;12(5):149–54.
77. Sun L, Wu Z, Lin Y, Xu S, Ye Y, Yin W, Zhou L, Lu J. Genetic polymorphisms of GGT1 gene (rs8135987, rs5751901 and rs2017869) are associated with neoadjuvant chemotherapy efficacy and toxicities in breast cancer patients. BMC Med Genomics. 2023;16(1):267.

Publisher's note